Systematics of Andean *Lupinus* L. and the origin of *L. mutabilis* Sweet

Ruth Jamina Eastwood

Thesis submitted to the Mathematical, Physical and Life Sciences Division,
University of Oxford for the degree of Doctor of Philosophy

Trinity Term 2006

St Cross College

Abstract

Systematics of Andean *Lupinus* L. and the origin of *L. mutabilis* Sweet

Thesis submitted for the degree of Doctor of Philosophy, University of Oxford, Trinity Term 2006

Ruth J. Eastwood
St Cross College

This thesis contributes to the central goals of systematics - the inventory of species, building the tree of life and understanding evolutionary processes - through a descriptive taxonomic study, molecular phylogenetic analysis and morphological survey of the genus *Lupinus* L.

Lupinus comprises ca. 290 annual and perennial species placed in the Genistoid clade of the Papilionoid legumes. In this study three new DNA sequence datasets are used to reconstruct species relationships. The new phylogeny is used to investigate character evolution, species diversification, domestication and as the basis for a new infrageneric classification.

Approaches to plant species-level phylogeny reconstruction and criteria for selecting DNA sequence loci are evaluated from a theoretical standpoint, and experimental work to develop hypervariable nuclear DNA sequence loci is presented.

The comparative morphology of *Lupinus* species is investigated and presented. This includes studies of seed coats, cotyledons, leaves, and chromosomes. Optimisation of these data onto the 3-locus phylogeny reveals congruence between chromosome number and geography and shows that non-digitate leaves have evolved twice independently within *Lupinus*.

Evolutionary rates analysis is used to estimate divergence times for clades and to investigate the tempo of species diversification. Notably, a strongly supported clade representing ca. 81 Andean species is estimated to be 1.18-1.76 Myr, implying a diversification rate of 2.49-3.72 species per Myr. This exceeds previous estimates for plants providing the most spectacular example of explosive plant species diversification documented to date.

Lupin species have been independently domesticated as grain legume crops in the Old and New Worlds. The origins of the Andean domesticate, *L. mutabilis*, are investigated. Lack of resolution amongst the Andean species, attributable to exceptionally rapid species diversification, meant that the progenitor(s) of *L. mutabilis* could not be firmly established. However, four Andean species are tentatively identified as closely related to *L. mutabilis* based on morphology. A taxonomic account of these species is presented.

Finally, a new infrageneric classification based on morphological diagnosability and monophyly is proposed.

Acknowledgements

My first thanks go to my supervisor **Colin Hughes** for his endless time, patience, support and encouragement. The Monty Python lines *"He seeks them here ... he seeks them there ... he seeks those lupins everywhere."* are highly appropriate to Colin whose extensive collecting trips have provided collections not only invaluable to my own work but whose quality will be valued by many legume taxonomists for years to come.

I thank **Julie Hawkins**, my second supervisor, for providing a fresh view on many topics and provoking discussion.

In Oxford I extend my gratitude to **Denis Filer** for BRAHMS assistance, **Serena Marner** and **Alison Strugnell** for their curatorial expertise, **Anne Sing**, **Rosemary Wise** for her exquisite botanical illustrations, **John Wood** for discussion, encouragement and tea, and The Systematics Discussion Group including **Robert Scotland** and **Jill Harrison**. For lab assistance and encouragement I thank **Donovan Bailey**, **Stephen Harris**, **Darren Obbard**, **Sarah Rendell** and **Ashley Robertson**. Additional thanks go to fellow Church Lab members, **Alex Wortley** and **Tim Waters,** for their support and encouragement. I would also like to thank the library staff for their invaluable assistance tracking down inter-library loans. I am grateful to **Timothy Walker** and the staff at Oxford Botanic Garden for allowing me to utilise their facilities and for tending my seedlings. I must thank **Barrie Juniper** and **Mark Knight** for allowing me to study in Oxford and fostering my botanical interest.

A special thanks goes to **Donovan Bailey** for the invitation to spend time in New Mexico. Both Jo and Donovan made me feel very welcome and I thank them for trips to White Sands National Monument. I thank **Norma Oropeza** for her company and discussion.

I thank the legume community for their discussion particularly **Gwil Lewis** for stimulating legume sorts. I am grateful to **Dawn Edwards** for her discussion. I am extremely grateful to those who specifically collected or donated *Lupinus* material for this project, particularly **Chris Drummond**, **Greg Kenicer**, **Silvia Miotto, Maria Pinheiro**, **Richard Spellenberg** and **John Wood**.

I extend my thanks to the many people who assisted me in my fieldwork. **Ancieto Daza's** great sense of humour and hard work whilst in the field were invaluable. Without stopping to look at birds for **Tim Budden** I would not have found so many lupins. **Ruth Clark** worked very hard. The staff at MOL and USM herbaria were great in their assistance with fieldwork. Special thanks go to **Homero Vargas**, **Louis Vargas** and their family for allowing me to have an insight into community life in Ecuador.

I thank my family and friends for their continued support throughout my D.Phil. and for keeping me going when things were tough. Their visits always provided a firm distraction from work. **Mum** and **Dad** provided a hectic retreat when Oxford got too much. I thank **Lauren Gardner, Hilary Ketchum, Alex Lord, Tim Waters, Barbara Willett,** and particularly **Effie Wiseman**, for sharing D.Phil. disasters, useful advice, shopping trips and always being on the end of a phone. I thank my housemates **Yuki Yasimura, Amande Pauls, Lisa Heather and Yvette Sturdy** for putting up with me and all my plants. I also thank **Stephanie Leung, Michael Kalantidis, Craig Stanton**, for cheering me up, eating cakes and drinking tea. Special thanks go to **Richard Black** for his long-distance support, an un-erring belief my thesis would be completed and latterly for ensuring I took breaks from work and giving me so much to look forward to.

I am indebted to the Biotechnology and Biological Sciences Research Council for my funding. I also thank The Systematics Association and The Genetics Society who provided me with conference and fieldwork funding.

Table of Contents

1 Introduction 1
1.1 Molecular systematics 1
1.2 *Lupinus* 3
1.3 Origin of the Andes and the Andean Flora 4
1.4 Domestication 7
1.5 Objectives and development of thesis research 8
1.6 Outline of thesis 10
Chapter 1 Figures 11

2 Selecting DNA data for plant species-level phylogeny reconstruction 12
2.1 Considerations for plant species-level phylogeny reconstruction 12
2.1.1 Species trees versus gene trees 12
2.1.2 Low molecular diversity 14
2.2 Sources of DNA data 17
2.2.1 Molecular markers 18
2.2.2 Chloroplast DNA sequence data 20
2.2.3 Nuclear ribosomal DNA sequence data 21
2.2.4 Nuclear DNA sequence data 24
2.2.5 Approaches to locate hyper-variable nDNA loci 27
2.3 Conclusion 33
Chapter 2 Figures 34

3 Phylogeny and Biogeography of *Lupinus* 38
3.1 Previous phylogenetic studies 38
3.2 Age of *Lupinus* 40
3.3 Methods 41
3.3.1 Taxon sampling 41
3.3.2 Outgroup selection 43
3.3.3 DNA Extraction 43
3.3.4 Amplification of ITS 43
3.3.5 Development and amplification of GPAT1 44
3.3.6 Development of SCAR-based DNA sequence regions 46
3.3.7 DNA sequencing 47
3.3.8 Sequence verification and alignment 48
3.3.9 Data set compatibility and combination 48
3.3.10 Parsimony analyses and support values 49
3.3.11 Bayesian analyses 49
3.3.12 Dating 51
3.3.13 Species diversification rates 51
3.3.14 Biogeography 51
3.4 Results and discussion 52
3.4.1 Phylogenetic utility of ITS for *Lupinus* species-level phylogenetic reconstruction 52
3.4.2 Phylogenetic utility of GPAT for species-level phylogenetic reconstruction 54

3.4.3 Phylogenetic utility of LEGCYC1A for species-level phylogenetic reconstruction 57
3.4.4 Phylogenetic utility of multiple loci and combined analysis for species-level phylogenetic reconstruction 58
3.4.5 Utility of Bayesian analyses 61
3.4.6 Phylogeny and biogeography of *Lupinus* 63
3.4.7 Radiation of Andean *Lupinus* 66
3.4.8. Utility of the SCAR approach and candidate loci 70
3.5 Conclusion 72
Chapter 3 Figures 74

4 Domestication of *L. mutabilis* 94
4.1 Domestication 94
4.2 Approaches to understanding domestication 94
4.3 Morphological change associated with pulse domestication 97
4.4 South American prehistory 100
4.5 Domestication in the Andes 102
4.6 Archaeological evidence for Tarwi, *L. mutabilis* domestication 104
4.7 Previous hypotheses of the origin of Tarwi, *L. mutabilis* 110
4.8 Methods 112
4.9 Results and discussion 113
4.9.1 Seed size, shape and colour 113
4.9.2 Domestication and origin of *L. mutabilis* 115
4.10 Conclusion 119
Chapter 4 Figures 120

5 Comparative morphology of *Lupinus* 133
5.1 Morphology versus molecules 133
5.2 Homology Assessment 134
5.3 Characters, character states and coding 136
5.4 Morphological studies within *Lupinus* 138
5.5 Methods 139
5.5.1 Herbarium material, field observations and greenhouse material 139
5.5.2 Scanning electron microscopy 140
5.5.3 Chromosome data 140
5.5.4 Characters and character states 140
5.6 Results and discussion 141
5.6.1 General 141
5.6.2 Metaphylls 141
5.6.3 Eophylls 146
5.6.4 Seed coat surface pattern 147
5.6.5 Cotyledons 153
5.6.6 Chromosome number 154
5.7 Conclusion 155
Chapter 5 Figures 156

6 Character evolution within *Lupinus* 171
6.1 Using phylogeny to assess character evolution 171

6.2 Leaf evolution 172
6.3 *Lupinus* cytological data and interpretations 175
6.4 Character optimisation methods 177
6.5 Results and discussion 178
6.5.1 Leaf evolution 178
6.5.3 Cotyledon evolution 182
6.5.4 Chromosome evolution 183
6.6 Conclusion 185
Chapter 6 Figures 186

7 A new infrageneric classification of *Lupinus* 197
7.1 Taxonomic history 197
7.2 Data and characters used in previous infrageneric classifications 203
7.3 Criteria for classification 205
7.3.1 Natural versus artificial classification 205
7.3.2 Diagnosability 206
7.3.3 Monophyly 206
7.3.4 Information content 208
7.3.5 Stability 208
7.3.6 Rank 209
7.3.7 Linnaean classification versus phylogenetic nomenclature 210
7.4 Methods 212
7.4.1 Criteria 212
7.4.2 Phylogeny data 212
7.4.3 Characters 213
7.5 Results 213
7.5.1 Infrageneric classification 213
7.6 Discussion 214
7.6.1 Diagnosability of *Lupinus* clades 214
7.6.2 Monophyly of infrageneric groups within *Lupinus* 214
7.6.3 Stability of the infrageneric classification 215
7.6.4 Rank 215
7.6.5 Diagnostic characters utilised in the infrageneric classification of *Lupinus*. 216
7.6.6 Discussion of new system relative to previous systems 217
7.6.7 Limitations of the new system 219
7.6.8 Description and discussion of Sections and Series 220
7.7 Conclusion 222
Chapter 7 Figures 223

8 Conclusions 226
8.1 Achievements and broader implications 226
8.2 Challenges 230
8.3 Further work 231

References 233

Abbreviations 274

Appendix 1 Taxonomic account of *L. mutabilis* and allies 275
Introduction 275
Distribution and Habitat 276
Key characters 276
Key to *L. mutabilis* and allies 278
Abbreviations 279
1 *Lupinus mutabilis* 280
Lupinus mutabilis Figures 290
2 *Lupinus piurensis* 297
2a *L. piurensis* var. *piurensis* 298
2b *L. piurensis* var. *cotumazicus* 299
Lupinus piurensis Figures 301
3 *Lupinus semperflorens* 304
3a *Lupinus semperflorens* var. *semperflorens* 306
3b *Lupinus semperflorens* var. *amazonus* 307
Lupinus semperflorens Figures 309
4 *Lupinus ellsworthianus* 312
Lupinus ellsworthianus Figures 316
5 *Lupinus praestabilis* 319
Lupinus praestabilis Figures 322
References for taxonomic account 325

Appendix 2 From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction (Hughes, Eastwood and Bailey, 2006) 328

Appendix 3 Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes (Hughes and Eastwood, 2006) 343

Appendix 4 Studies assessed to investigate the status of species-level phylogeny reconstruction 349

Appendix 5 Sequence data utilised, vouchers and Genbank numbers 351

Appendix 6 Protocols and Solutions 360

Appendix 7 Primer sequences and PCR protocols 368

Appendix 8 Bayes and r8s Blocks 369

Appendix 9 Age estimates 374

Appendix 10 ISSR primers 380

Appendix 11 ISSR primer combinations 381

Appendix 12 Comparison of ISSR-based SCAR loci 382

Appendix 13 Seed data 383

Appendix 14 Observations of articulation and pulvinule/pulvinules in non-digitate species 391

Appendix 15 *Lupinus* seed coat surface pattern character states and representative species 392

Appendix 16 Chromosome numbers reported in *Lupinus* 395

Appendix 17 Morphological matrix 406

Appendix 18 Infrageneric classification of *Lupinus* 411

List of Figures

Figure 1.1 Higher-level phylogeny of Leguminosae showing overall relationships of *Lupinus*, model legumes and genera of major agricultural importance. 11

Figure 2.1 Species trees versus gene trees 34
Figure 2.2 Resolution obtained in recent species-level phylogenetic studies. 35
Figure 2.3 Loci used in species-level phylogenetic studies. 36
Figure 2.4 Potential informativeness different types of DNA sequence loci in plant species-level phylogenetic analyses. 37

Figure 3.1 *Lupinus* strict consensus ITS gene tree redrawn from Aïnouche and Bayer (1999) 74
Figure 3.2 Diagram of the full-length GPAT open reading frame 75
Figure 3.3 Agar gel showing two PCR GPAT products in some Genisteae 76
Figure 3.4 Fig. 3.4 SCAR method. 77
Figure 3.5 ITS parsimony gene tree 78
Figure 3.6 GPAT parsimony gene tree 79
Figure 3.7 GPAT1 parsimony gene tree 80
Figure 3.8 LEGCYC1A parsimony gene tree 81
Figure 3.9 3-locus (with no missing data) parsimony gene tree 82
Figure 3.10 3-locus (with missing data) parsimony gene tree 83
Figure 3.11 Statistical summary of parsimony analyses 84
Figure 3.12 ITS Bayesian gene tree 85
Figure 3.13 GPAT1 Bayesian gene tree 86
Figure 3.14 ITS and LEGCYC1A Bayesian gene tree 87
Figure 3.15 ITS and LEGCYC1A Bayesian gene tree showing branch lengths 88
Figure 3.16 3-locus (with missing data) Bayesian gene tree 89
Figure 3.17 Statistical summary of Bayesian analyses 90
Figure 3.18 Resolution in the Andean clade 90
Figure 3.19 Biogeography of *Lupinus* 91
Figure 3.20 Dispersal and vicariance analysis results 92
Figure 3.21 Phylogeny, distribution and diversity of the Andean *Lupinus* radiation 93

Figure 4.1 Archaeological sites referred to in text 120
Figure 4.2 Andean timeline showing *Lupinus* archaeological evidence and events 121
Figure 4.3 Proposed prehistoric depictions of Tarwi, *L. mutabilis*, on ceramic artifacts 122
Figure 4.4 Seed size variation within *Lupinus* 123
Figure 4.5 Rim aril 123
Figure 4.6 Rough and smooth seeds 123
Figure 4.7 Intraspecific seed colour and pattern variation 123
Figure 4.8 Variation in *Lupinus* seed shape. 124
Figure 4.9 *Lupinus* seed coat colour and pattern variation 125
Figure 4.10 Variation in seed width within *Lupinus* 126
Figure 4.11 Variation in seed length within *Lupinus* 127
Figure 4.12 Variation in seed thickness within *Lupinus* 128
Figure 4.13 Variation in seed weight within *Lupinus* 129
Figure 4.14 *Lupinus mutabilis* and putative progenitors 130

Figure 4.15 Seed weight of 32 Andean species 131
Figure 4.16 Morphological change through domestication 132

Figure 5.1 Simple versus foliolate *Lupinus* metaphylls 156
Figure 5.2 Variation in *Lupinus* metaphyll morphology 157
Figure 5.3 Variation in digitate metaphyll morphology 158
Figure 5.4 Variation in leaflet morphology of digitate metaphylls 159
Figure 5.5 Indications of articulation 160
Figure 5.6 Variation in unifoliolate metaphyll morphology 161
Figure 5.7 Leaf character states 162
Figure 5.8 Leaf and leaflet measurements for *L. mutabilis* and *L. semperflorens* 163
Figure 5.9 Stipule measurements for *L. sp2, L. semperflorens* and *L. mutabilis* 164
Figure 5.10 Eophyll character states 165
Figure 5.11 Seed coat surface patterns at low magnification 166
Figure 5.12 Seed coat surface pattern character states 167
Figure 5.13 Variation within *Lupinus* seed coat surface pattern character states 168
Figure 5.14 Variation within *Lupinus* seed coat surface pattern character states 169
Figure 5.15 Cotyledon character states 170

Figure 6.1 ACCTRAN versus DELTRAN optimisation. 186
Figure 6.2 Evolution of metaphylls in *Lupinus* 187
Figure 6.3 Evolution of eophylls in *Lupinus* 188
Figure 6.4 Leaf evolution within clade G (Section Orientalis). 189
Figure 6.5 Evolution of seed coat surface pattern in *Lupinus* 190
Figure 6.6 Evolution of seed coat surface pattern in *Lupinus* 191
Figure 6.7 Evolution of cotyledons in *Lupinus* 192
Figure 6.8 Evolution of chromosome number in *Lupinus* 193
Figure 6.9 Evolution of chromosome number in *Lupinus* 194
Figure 6.10 Evolution of chromosome number in *Lupinus* 195
Figure 6.11 Possible chromosome evolution within *Lupinus* 196

Figure 7.1 Comparison genus-wide classifications of *Lupinus* 223
Figure 7.2 Infrageneric classification and the *Lupinus* phylogeny 224
Figure 7.3 Distribution map of *Lupinus* Sections/Series 225

1 Introduction

1.1 Molecular systematics

Systematics is fundamental to research across biology (Eldredge, 1992; Forey *et al.*, 1994; Nelson & Platnick, 1981; Wiley, 1981). It provides tools for identification e.g. floras, monographs, and keys, and frameworks for comparative biology e.g. classifications and phylogenies. Species-level phylogenies underpin studies of domestication, hybridisation, character evolution, infrageneric classification, biogeography and the evolutionary processes underlying patterns of diversity (Appendix 2, Hughes *et al.*, 2006). The pivotal role of phylogenies across the whole of comparative biology means that reconstructing the tree of life is a major aim of the systematic community. Reconstruction of phylogenies forms the fundamental core of this work and has blossomed with the formulation of systematic techniques and the advent of molecular data (Donoghue & Cracraft, 2004; Hughes *et al.*, 2006). This advance has been concurrent with improvements in methods of analysis and generation of explicit models of sequence evolution (Appendix 2, Hughes *et al.*, 2006). Progress has not been without controversy surrounding both methods and data sources (e.g. morphological vs. molecular data, parsimony vs. likelihood analysis, combined vs. individual analyses). Nevertheless, together these have prompted spectacular progress in assembling the Tree of Life (Donoghue & Cracraft, 2004; Palmer *et al.*, 2004).

However, while considerable progress has been made in the construction of higher level phylogenies, building robust species-level phylogenies, at least for plants, has been less successful (Appendix 2, Hughes *et al.*, 2006). This is hindering investigations of many fundamental questions in plant evolutionary biology since phylogenies are most useful when well-sampled, well-resolved and strongly-supported (Hawkins, 2006). Both DNA marker data and sequence loci have been widely used in species-level phylogenies with the latter more commonly utilised. The most popular approach to improve phylogeny quality has been the application of low-copy nuclear genes as a source for DNA sequence data. However, this has had limited success because levels of variation among nuclear intron

sequences across groups of closely related species are extremely variable and generally lower than conventionally used loci (Appendix 2, Hughes *et al.*, 2006), and because no universally useful low-copy nuclear DNA sequence loci have been developed (Sang, 2002). This suggests that solutions will, for the most part, be lineage-specific, prompting a move away from 'universal-loci' thinking for species-level phylogenetics. Methods to generate lineage-specific loci are proliferating, but problems with outgroup amplification, copy number and low numbers of parsimony informative characters are commonly encountered by all methods. These challenges indicate there will not be a quick, easy and universal solution for locating hyper-variable sequence loci for species-level phylogeny reconstruction.

Phylogenies have potential to provide information in addition to hypotheses of relationships. Phylogenies are commonly utilised to investigate character evolution (e.g. floral (Guzman & Vargas, 2005), life history strategy (Wilson & Calvin, 2006), pollination (Kay *et al.*, 2005), chromosome number (Ellison *et al.*, 2006)). Rates analysis allows dating of nodes on a tree and subsequent estimates of diversification rate (Appendix 3, Hughes & Eastwood, 2006; Lavin *et al.*, 2005). Tree-based approaches are also applied to studies of domestication (Emshwiller & Doyle, 1998; Sanjur *et al.*, 2002), hybridisation (Hughes *et al.*, 2002) and polyploidy (Straub *et al.*, 2006). A more controversial area is if and how phylogenetic data should be incorporated into classifications (Brummitt, 2002; Pfeil & Crisp, 2005). This has led to the suggestion of an alterative classification system, the Phylocode (Cantino & de Queiroz, 2006), built solely on phylogenetic data.

I have investigated the challenge of selecting and developing appropriate loci for species-level phylogeny reconstruction using the genus *Lupinus* L.. I have also considered leaf evolution, the origins of the Andean flora and the use of phylogenetic data in classification by studying this genus.

1.2 *Lupinus*

Lupinus is a member of the Leguminosae. It is in the tribe Genisteae, which is nested in the large informal Genistoid clade within the subfamily Papilionoideae (Fig 1.1) (Crisp *et al.*, 2000; Lewis *et al.*, 2005a; Wojciechowski *et al.*, 2004). *Lupinus* is a large genus of around 290 species distributed in the Old and New Worlds (Fig 7.3). Thirteen species are distributed around the Mediterranean and North Africa (Ainouche *et al.*, 2004; Gladstones, 1974, 1998). However, the two main centres of species diversity are in western North America and the Andes in South America. Naturalized lupins are also found in South Africa, northern Europe, Australia and New Zealand, where, in some cases, large invasive populations have formed which threaten native ecosystems (Williams *et al.*, 2002).

The Leguminosae is one of the most economically important families (Doyle & Luckow, 2003; Lewis *et al.*, 2005a) and additionally contain many of the earliest domesticated crops (Zohary & Hopf, 1973). Lupin species are utilised in agriculture and horticulture. Domestication of lupins, as a pulse crop for human and animal consumption, occurred in parallel in the Old World (*L. albus* L. and *L. luteus* L.*)* and the New World (*L. mutabilis* Sweet) (Gladstones *et al.*, 1998; Gross, 1982, 1986; Hondelmann, 1984; Plitmann & Heyn, 1984). Lupins are important in modern agriculture and it is estimated that over 700,000 tonnes of seed from the domesticated Old World species were harvested in Western Australia in 2004 (Harries, 2005). Although little known worldwide, *L. mutabilis* is a major subsistence crop in the Andes (Gross, 1982).

Taxonomic research on *Lupinus* over the last fifty years has been fragmentary, largely focused geographically, and lacking monographic synthesis and consistency in terms of species delimitation. Detailed studies have been made almost entirely on a regional basis; e.g. Old World (Gladstones, 1974; 1998), Intermountain region of North America (Barneby, 1990), California (Riggins & Sholars, 1993), Canada and Alaska (Dunn & Gillett, 1966), Argentina (Planchuelo & Dunn, 1984), and North and South American countries (Smith, 1938-1953) or within morphologically defined groups; e.g. Brazilian

unifoliolates (Monteiro & Gibbs, 1986a) and various North American species complexes (Dunn, 1955, 1957, 1959; Sholars, 1996; Vaughn & Dunn, 1977). A very large number of species names have been published partly because investigations have been confined within political boundaries but also as a result of the work of Charles Piper Smith who alone described 675 new species and infraspecific taxa between 1917 and 1953. It is widely acknowledged (Barneby, 1990; Standley & Steyermark, 1946) that a large proportion of Smith's names do not represent distinct species, and indeed that at least latterly he suffered from senility (Thomas, 1974). The burden caused by Smith's surfeit of names presents a significant hurdle to taxonomic studies of *Lupinus*, especially in the Andes where Smith carried out much of his work. Although widely acknowledged as a centre of diversity for *Lupinus* (Planchuelo, 1996a), the Andes have received little recent taxonomic attention following Smith's daunting legacy. Re-assessment of the species delimitation considering morphological variation across the Andes as a whole has up to now been lacking.

Elucidation of the phylogeny of *Lupinus* has been hampered by low taxon and character sampling and poor resolution in gene trees. Taxon sampling has been biased towards the wild and cultivated Old World species although some New World, predominately North American, species have been included. Only one, out of the ca. 85, Andean species was included in previous phylogeny estimates (Ainouche & Bayer, 1999; Ainouche *et al.*, 2004; Kass & Wink, 1997a).

New work on the taxonomy of Andean *Lupinus* was started by Colin Hughes in 2000 with the aims of delimiting species and assembling a new taxonomic account. My work forms part of this new initiative, focusing on delimitation of a group of Andean species related to *L. mutabilis* and producing detailed species accounts for these.

1.3 Origin of the Andes and the Andean Flora

The Andes are the longest North-South mountain chain in the world, spanning 7000 miles

from Venezuela to Argentina and Chile, reaching over 6000 m in height and dropping below 2500 m only in northern Peru at the Huancabamba Depression (Brako & Zarucchi, 1993; Jørgensen & Lâeon-Yâanez, 1999; Young *et al.*, 2002). They are also exceptionally dissected and heterogeneous topographically comprising of high peaks, deep valleys and high plains. Due to the elevated height the climatic conditions differ dramatically from lowland areas at the same latitude. Differences include high diurnal temperature fluctuations, lower humidity and critically for plants, increased levels of solar radiation. Additionally, the eastern slopes of the Andes receive considerably greater precipitation than the semi-arid western slopes which fringe the hyper-arid Atacama desert to the west (Van der Hammen & Cleef, 1986).

The Andes as a whole are a biodiversity hotspot for plants and animals (Brummitt & Lughadham, 2003; Lamoreux *et al.*, 2006; Mittermeier *et al.*, 1998; Myers *et al.*, 2000). They contain an estimated 45,000 plant species, 44% of which are endemic (Myers *et al.*, 2000). Additionally, the high elevation flora of the northern Andes is by far the most species-rich of any tropical mountain massif (Smith & Cleef, 1988). Endemism is estimated to be even higher, up to 60%, in the high altitude páramos (Luteyn, 1999). There is and has been for over 200 years great interest in understanding the origin of this rich flora (Jørgensen & Lâeon-Yâanez, 1999; von Humboldt & Bonpland, 1807). For this the origin of the Andes themselves must be taken into consideration.

Geological evidence shows that the Andes are comparatively young and moreover their orogeny was extremely rapid (Ghosh *et al.*, 2006). However, the Andes are not a single continuous mountain chain and each separate part has an individual uplift history (Gregory-Wodzicki, 2000). The earliest uplift is thought to have begun ca. 25 Mya in the South and subsequent uplift rates are thought to be variable (Kroonenberg *et al.*, 1990). Studies estimate the northern Andes were half their current height only 10 Mya and reached their present altitudes 6-4 Mya (Kroonenberg *et al.*, 1990). Although most uplift rate estimates are less than 0.2 mm per year, a recent estimate from the Bolivian Andes suggests uplift as

fast as one millimetre a year (Ghosh *et al.*, 2006).

Elucidating the origin of the Andean flora requires both knowledge of the present-day flora and investigations into how Andean habitats were colonised. Although South America was long thought to have been isolated after its spilt from Africa ca. 90 Mya until the formation of the Panama Land Bridge ca. 3 Mya (Simpson, 1980), current evidence highlights the possibility of several land or island chain connections allowing plant dispersal during this period (Pennington & Dick, 2004). The importance of the Panama Land Bridge for dispersal between the Americas has been demonstrated for animals (Stehli & Webb, 1985) and has been assumed likewise to be influential for plants despite their different dispersal abilities. Pollen data support the arrival of some genera (e.g. *Alnus* L., *Hedysarum* L. and *Quercus* L.) in South America after the formation of the Panama Land Bridge and Mexican forest floras contain many elements derived from South America (Burnham & Graham, 1999; Graham, 1995). It has however, been suggested that plant dispersal across the Panama Land Bridge was limited to species adapted to tropical climates, at least until the late Pliocene (Graham, 1992; Simpson & Neff, 1985). The Andean páramo flora is composed of both temperate and Neotropical genera in near equal quantities (Van der Hammen & Cleef, 1986). The temperate contribution is greatest in herbaceous genera (Hooghiemstra & Cleef, 1995). This is reflected in the domination by temperate elements in the superpáramo, above the treeline, and is consistent with the idea that pre-adapted species have a greater chance of successful colonisation (Daehler, 2005).

It is clear that the Andes are a biodiversity hotspot and are also very recent in age (Gregory-Wodzicki, 2000) leading to the suggestion that the orogeny of the Andes was a major cause of the high levels of species diversity observed today (Gentry, 1982). A further factor suggested to have had an important influence on diversification is Pleistocene climate fluctuations (Prance, 1982; Simpson, 1975). The huge number of extant species in the Andes presents practical and logistical problems for taxonomic work. For instance cataloguing species alone is a considerable task. The first checklists and catalogues for Ecuador, the Pàramos and Peru

were only published within the last fifteen years (Brako & Zarucchi, 1993; Jørgensen & Lâeon-Yâanez, 1999; Luteyn, 1999) and many large countries e.g. Bolivia have no complete checklists.

Part of my work aims to investigate plant colonisation and species diversification in the high Andes using the genus *Lupinus*.

1.4 Domestication

In addition to being a biodiversity hotspot the Andes have also been a major centre for domestication (Cook, 1925; Diamond, 1997, 2002). They are one of the eight or nine geographic regions where domestication of crops and livestock occurred independently 13,000 – 8,000 years ago (Diamond, 1997, 2002; Piperno & Pearsall, 1998; Smith, 1998, 2001). Like plant species-diversity more generally, Andean crops are poorly understood and often referred to as the "Lost Crops of the Incas" (National Research Council, 1989b) despite the fact they remain a major component of the modern Andean diet.

Domestication of plants and animals and the advent of agriculture have been one of the most influential factors in the development of human society (Cook, 1925; Diamond, 1997, 2002; Harlan, 1975; Hawkes, 1983; Weiss *et al.*, 2004). Growing and distributing food sufficient to feed the world's population is a global problem (Food and Agriculture Organization of the United Nations 1996). Crop improvements through selection and breeding will continue to be vital to human survival. In order to do this quickly and efficiently it is important that the biology, evolutionary history and wild relatives of domesticates are fully understood. It has been suggested that studies of lesser known domesticates may facilitate crop improvement just as much as the study of major crops, because their domestication history may be less confounded by modification from intensive breeding (Emshwiller, 2006).

There are a large number of different legume crops whose uses include; human and animal food sources, oil, firewood, timber, fibre, flavourings, medicine, green manure, insecticides, ornamentals and resins (Duke, 1981; Gill *et al.*, 1980; Lewis *et al.*, 2005a; Smartt, 1990; Smartt & Simmonds, 1995). The most commonly used part is the seed pod although in some cases the whole plant, wood, seeds alone or tubers are utilised (Isely, 1982). Detailed knowledge of legume crops will facilitate maintenance and improvement of vitally important domesticates.

Domesticated lupins are found in both the Old and New Worlds. In contrast to the intensively studied Old World species (Ainouche & Bayer, 1999, 2000; Ainouche *et al.*, 2004; Gladstones, 1974; Gladstones *et al.*, 1998; Kass & Wink, 1997b) the Andean domesticate Tarwi, *L. mutabilis*, is relatively poorly known. This crop is on trial in Australia where the other domesticated lupin species are grown in large quantities (Clements *et al.*, 2005b).

My work aims to investigate the origin of a widespread Andean crop, *L. mutabilis*, with a view to understanding domestication in the Andes more widely and pinpointing genetic resources which may be useful in future breeding schemes. My research has involved comparative morphological studies and molecular phylogeny reconstruction. I have also reviewed archaeological reports of *Lupinus* in order to compile a timeline for the domestication of *L. mutabilis*.

1.5 Objectives and development of thesis research

The original objectives of my research were:

- To construct a well-resolved species-level molecular phylogeny for Andean *Lupinus* densely sampling species closely related to *L. mutabilis*.
- To delimit *L. mutabilis* and closely related species.
- To identify the close relatives of *L. mutabilis*.

- To write a taxonomic account for *L. mutabilis* and closely related species.
- To quantify phenotypic change resulting from domestication of *L. mutabilis* to gain insights into early agriculture and crop domestication processes.

Although all these objectives have not been fully achieved, all have been attempted. Preliminary phylogeny reconstruction with the nuclear ribosomal DNA (nrDNA) sequence locus ITS encountered extremely low levels of resolution between the majority of the New World species. Reconstruction of the *Lupinus* phylogeny utilising the nuclear (nDNA) locus LEGCYC1A although producing a greater level of resolution and resolving the majority of Andean species in a clade with high support, was also uninformative about relationships between them. The lack of resolution in both loci stems directly from very low levels of sequence divergence amongst the Andean species. This discovery altered the focus of my project in a number of ways. Firstly, it simulated an assessment of the status of species-level plant phylogenetics in the form of a survey of DNA sequence loci currently in use for species-level phylogeny work and approaches to develop and screen suitable loci (Appendix 2, Hughes *et al.*, 2006). Secondly, it prompted the development of an additional nDNA locus for *Lupinus,* glyceraldehyde phosphate acetyl transferase copy one (GPAT1), as well as an attempt to locate faster evolving loci when GPAT1 was also found to be insufficiently variable to resolve relationships between the Andean taxa. Also, it prompted me to investigate the reasons for the low level of sequence variation among the Andean taxa. This involved evolutionary rates analysis to estimate divergence times for key nodes on the best phylogenetic estimate. As my research progressed and more data were generated, it emerged that the Andean lupins comprise the most spectacular example of rapid plant diversification documented to date. Given the challenges associated with obtaining a resolved tree for the Andean species and hence shedding light onto the origins of *L. mutabilis*, I decided to pursue new lines of research to investigate the comparative morphology, biogeography, character evolution and infrageneric classification of *Lupinus*.

1.6 Outline of thesis

Following this introductory chapter my thesis contains one theoretical/review chapter (Chapter 2), five empirical chapters (Chapters 3-7) each presenting methods, results and discussions and a brief concluding chapter (Chapter 8). Each chapter addresses a specific area of research. Figures follow each chapter. Additional material is placed in Appendices including: two papers (Appendices 1 and 11) which include elements of work carried out as part of my D.Phil project and are presented in full in Chapters 2 and 3, and a full taxonomic account including illustrations, distribution maps and specimen citation lists for five species (Appendix 1). Electronic appendices are included (see attached CD ROM) containing alignments of molecular data matrices.

Chapter 2 discusses the issues, both theoretical and empirical, influencing selection of loci for plant species-level phylogeny reconstruction. This paints the background for Chapter 3 which presents a phylogenetic analysis of *Lupinus* utilising three nuclear DNA sequence loci and also discusses the biogeography of the genus. Additionally, results from preliminary work to discover and isolate hyper-variable DNA sequence loci are presented. Chapter 4 focuses on the Andean domesticate *L. mutabilis* summarising archaeological data, presenting a new hypothesis for it's origin and assessing morphological change in comparison with wild species including the putative progenitor. Comparative morphological data for a set of key characters in *Lupinus* are presented in Chapter 5 and utilised in combination with the new phylogeny in Chapter 6 to assess character evolution within the genus. Finally, Chapter 7 combines morphological and phylogenetic data to present a new infrageneric classification for *Lupinus*.

Taxonomic work on the Andean lupins is ongoing and accurate names can not yet be assigned to all accessions utilised because the nomenclature (480 names) has not been fully resolved for all species. For this reason un-named accessions utilised in the phylogenetic analyses are referred to throughout as *L. sp1 – 9*. Un-named species utilised elsewhere remain as *L. spA-X* pending further taxonomic work. The collector name and number and voucher data for all samples utilised are listed in the Appendices such that in the future they can be matched with correct names

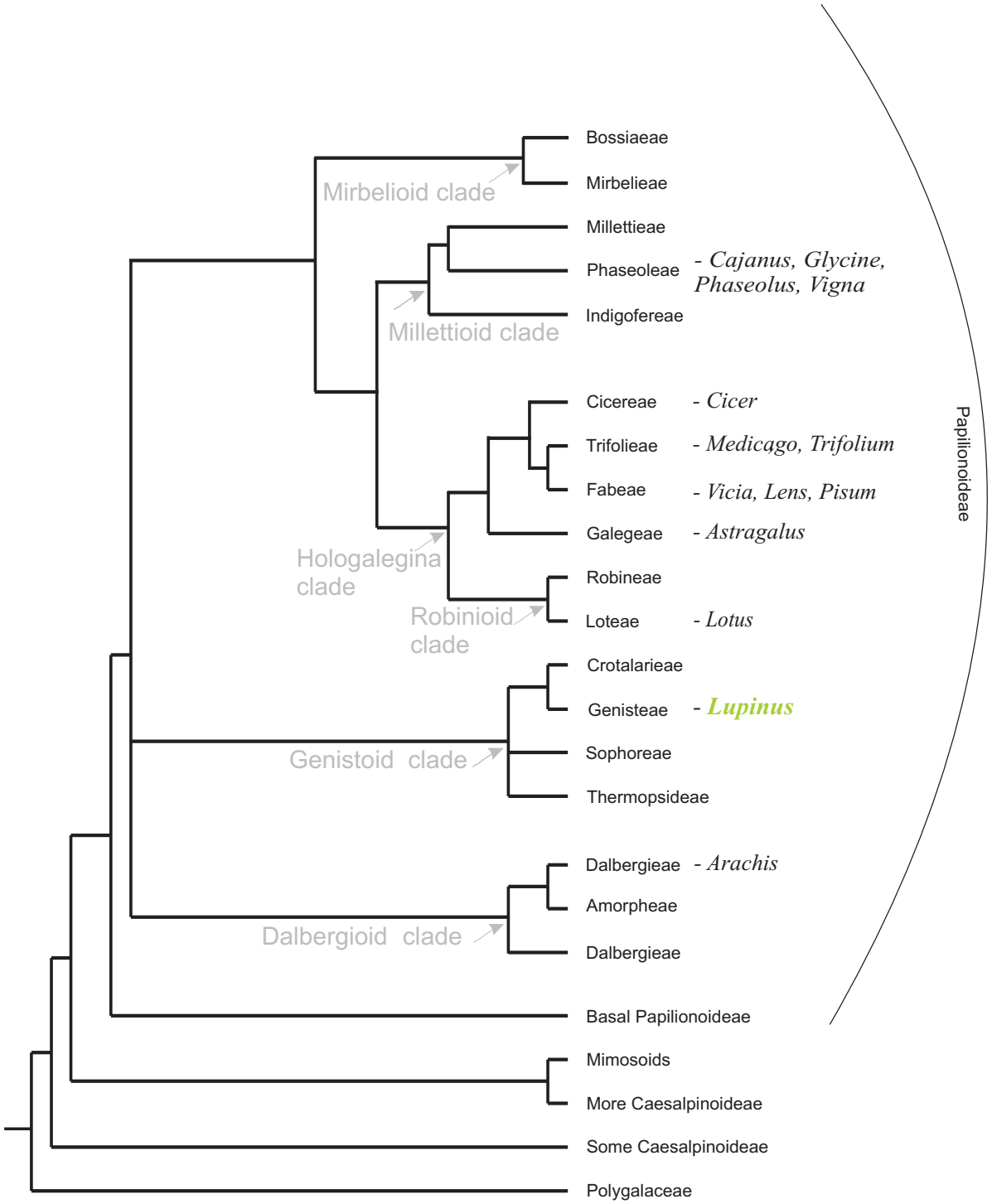


Fig. 1.1. Higher-level phylogeny of Leguminosae showing overall relationships of *Lupinus*, model legumes and genera of major agricultural importance. Note that *Lupinus* is not closely related to a model legume or a group of primary economic importance thus the genetic resources associated with these can not be directly utilised. This tree is a summary of Crisp *et al.* (2000), Doyle *et al*. (2000), Kass & Wink (1997b), Lavin *et al*. (2001a, 2003) and Wojciechowski *et al*. (2004).

2 Selecting DNA data for plant species-level phylogeny reconstruction

2.1 Considerations for plant species-level phylogeny reconstruction

2.1.1 Species trees versus gene trees

Conclusions drawn from molecular phylogenetic studies rely heavily on the assumption that the topology of a gene tree reflects the true species tree. However, it is widely recognised that this assumption may be broken due to the effects of incomplete lineage sorting, gene duplication, inter-specific hybridisation and recombination (Fig. 2.1) (Comes & Abbott, 2001; Doyle, 1992; Maddison, 1997; Smouse, 2000). In fact there is no general expectation that species-level gene trees will be equivalent to the true species tree (Nichols, 2001; Sanderson & Shaffer, 2002). Assessment of gene-tree species-tree conflicts is complicated by the fact that the true species tree is rarely known and conflicting topologies may be derived from different datasets with equally high support (e.g. Martin & Burg, 2002; Oh & Potter, 2003; Semerikov *et al.*, 2003).

Incomplete lineage sorting can occur when polymorphic ancestral alleles persist through one or more speciation events. If these are differentially retained or sampled in terminal taxa then the gene tree will not reflect the true species tree (Fig. 2.1a). This is a purely stochastic process (Slade *et al.*, 1994) which is more pronounced in lineages with large effective populations and those in which time since divergence is short (Maddison, 1997; Pamilo & Nei, 1988), i.e. rapid radiations.

Gene duplication or loss can produce a similar outcome to lineage sorting when duplicate copies are differentially lost or sampled amongst taxa (Fig. 2.1b). Phylogenetic analysis may reconstruct the true gene history, but this may not match species history (Fig. 2.1b). The dynamics of gene duplication and loss (Lynch & Conery, 2000) are different from allele sorting since duplicate gene copies are not competing for the same locus and may both be retained with independent fates (Maddison, 1997). For this reason population size is

unlikely to influence whether this issue complicates species-level phylogeny reconstruction. Short divergence times and recency increase the likelihood of gene duplication issues as paralagous copies will be harder to distinguish and amplify separately due to little genetic divergence. Since most genes belong to gene families, identification and amplification of orthologous genes from a set of taxa can be challenging, yet this homology underpins the phylogenetic value of any dataset.

One of the central assumptions of cladistic analysis is strictly hierarchical evolutionary history. However, interspecific hybridization is extremely frequent and widespread in plants (Andreasen & Baldwin, 2003; Gross *et al.*, 2003; Hegarty & Hiscock, 2005; Hughes *et al.*, 2002; Lihova *et al.*, 2006; Linder & Rieseberg, 2004; Poke *et al.*, 2006; Rieseberg & Carney, 1998). Reconstructing relationships among potentially reticulating species is challenging due to introgression, potential chimeric sequences and the need for comparisons between multiple loci from different genomes. Chloroplast DNA (cpDNA) and nrDNA loci individually are of reduced utility in this situation (Linder & Rieseberg, 2004). However, uniparental inheritance of plastids makes cpDNA essential for teasing apart maternal versus paternal contributions to putative hybrid species when compared to other data. Similarly, nrDNA multi-copy arrays that have undergone complete concerted evolution do not retain evidence of biparental inheritance (Baldwin *et al.*, 1995; Wendel *et al.*, 1995). The unpredictable extent and direction of concerted evolution among nrDNA repeats, which can even vary among individuals of the same species (Doyle *et al.*, 2004), can create situations where nrDNA data provide partial and potentially confusing information about reticulation. These limitations mean that multiple independent bi-parentally inherited loci are essential for species-level phylogenetic studies where reticulation is a possibility (Ferguson & Sang, 2001; Hegarty & Hiscock, 2005; Linder & Rieseberg, 2004; Raymond *et al.*, 2002), and these can only come from the nuclear genome.

Potential gene tree/species tree conflicts need to be considered when deciding how many and which loci are sampled in species-level phylogenetic analyses. To recognize and address

impacts of incomplete lineage sorting, gene duplication, hybridization and recombination multiple independent loci from different genomes should be utilised (Doyle, 1992; Maddison, 1997; Pamilo & Nei, 1988). Congruence among multiple independent gene trees is one of the best indicators of phylogenetic accuracy (Doyle, 1992; Miyamoto & Fitch, 1995; Pamilo & Nei, 1988).

2.1.2 Low molecular diversity

Different sequence loci can be utilised at particular taxonomic levels because rates of molecular evolution are not uniform across genomes, loci and coding/non-coding regions (Pelser *et al.*, 2003; Small *et al.*, 2004; Wolfe *et al.*, 1987). A particular challenge for reconstruction of relationships between closely related species is low sequence divergence, and lack of parsimony informative characters (PIC). This is because there is less genetic divergence between closely related species compared to a higher level study sampling distantly related taxa. The lack of PICs is most critical in radiations (periods of rapid speciation) and recent speciation events where few DNA characters have accumulated to distinguish between species (Albertson *et al.*, 1999; Klak *et al.*, 2004; Kocher, 2003; Richardson *et al.*, 2001a; Rokas *et al.*, 2005). Simulation studies have shown that phylogenetic accuracy increases with number of characters (Hillis, 1996; Hillis *et al.*, 1994; Wortley *et al.*, 2005). The consequences of few characters are thus possible inaccuracy, lack of resolution, short branch lengths and weak node support (Hawkins, 2006).

A review of 345 species-level gene trees published in four journals in 2003-2004 (See Appendix 4 for details) revealed that lack of resolution is a widespread problem (Fig. 2.2) (Appendix 2, Hughes *et al.*, 2006). Resolution was assessed using the consensus fork index (CFI) (Colless, 1980). The CFI is calculated by dividing the number of resolved nodes on a strict consensus tree by the number of possible resolved nodes, n-2, where n is the number of terminals. Only 11.5% of published gene trees were fully resolved with a modal CFI of 0.51-0.6 (Fig. 2.2a) (Appendix 2, Hughes *et al.*, 2006). Lack of resolution is thus a serious

problem since unresolved phylogenies are less informative (Hawkins, 2006).

Polytomies (more than two branches arising from the same node) can be regarded as hard, due to extremely rapid or simultaneous speciation, or soft where bifurcating speciation events occurred but have not been resolved due to insufficient data or character conflict. Distinguishing between these is difficult but most polytomies are assumed to be soft (Braun & Kimball, 2001; Jackman *et al.*, 1999; Knowles, 2000; Megans *et al.*, 2004; Walsh & Friesen, 2001; Walsh *et al.*, 1999). There are many possible underlying reasons for polytomies in gene trees which can result from low levels of, conflicting, or too much DNA sequence variation. This can lead to insufficient PICs or those with conflicting signal. Conflicting signals can produce alternative equally parsimonious trees with weakly supported tree topologies and unresolved consensus trees. A number of scenarios can explain each dataset pattern.

Lack of sequence variation could be caused by:

- Rapid speciation
- Simultaneous speciation – i.e. hard polytomy
- A slowing in the rate of molecular evolution

Conflicting patterns can be caused by:

- Hybridisation and/or incomplete lineage sorting
- Gene duplication/extinction
- Missing data
- Chimeric sequences
- Increased rate of molecular evolution

The latter could also produce too much variation so that the phylogenetic signal is masked. Usually, lack of resolution is directly attributable to insufficient variation (Bailey *et al.*, 2004; Shaw *et al.*, 2005a; Shaw *et al.*, 2005b; Small *et al.*, 1998), particularly as a result of rapid or simultaneous speciation (Liu *et al.*, 2006; e.g. Richardson *et al.*, 2001a) rather than incongruence. Recency compounds the low levels of variation observed due to rapid/simultaneous speciation or a slowing in the rate of molecular evolution (Knowles, 2000).

Since the majority of speciation events involving extant taxa are recent this provides a further explanation for low levels of molecular diversity at the species level.

Little is known about the genetic basis of morphological diversity (Ehrenreich & Purugganan, 2006) yet it has been suggested that the genetic basis for major morphological changes can involve small sequence differences at only a few loci (Doebley, 2006; Lukens & Doebley, 2001). These loci may be predominately those involved with expression patterns i.e. transcription factors (Cong *et al.*, 2002; Doebley, 2006; Doebley & Lukens, 1998; Selinger & Chandler, 1999; Wang *et al.*, 1999). Well-characterized evidence for this hypothesis comes from studies of teosinte branched1, a transcription factor which distinguishes plant architecture in maize from its putative progenitor, teosinte (Lukens & Doebley, 2001). A further example is fruit weight2.2 in tomato, a quantitative trait locus (QTL) correlated with fruit weight (Cong *et al.*, 2002; Doganlar *et al.*, 2000; Nesbitt & Tanksley, 2002). Phylogenetic species (Eldredge & Cracraft, 1980; Nelson & Platnick, 1981; Nixon & Wheeler, 1990), are recognised on the basis of one or more morphological characters that show fixed differences between populations. If, as implied by studies of morphological changes associated with domestication, relatively small changes at few loci can determine these sort of fixed morphological characters, genetic differences between species may be very small. Given the low levels of sequence variation observed in some rapidly evolving DNA sequence loci across morphologically divergent species, it is clear that sequence evolution can be similarly low in unconstrained loci. Overall this implies that locating sufficient characters to reconstruct species-level phylogenies will be a significant challenge for many plant groups.

Molecular diversity although most commonly assessed in terms of base pair substitutions, can also take the form of structural mutations such as numbers of sequence repeats or presence/absence of insertions or deletions (indels). Both sequence repeats and indels have been implicated as equally or more important than substitions in determining morphological change and species differences (Britten *et al.*, 2003; Fondon & Garner, 2004) and are thus a

potentially useful source of characters for species-level phylogeny reconstruction. However, although alignment of repeat sequences is possible, confidence in homology is low since tracking losses and gains of identical sequence repeats is not possible without much denser population-level sampling or at least multiple accessions within species.

2.2 Sources of DNA data

DNA has several advantages over other molecular markers e.g. allozymes (Hillis *et al.*, 1996). It allows genotype rather than phenotype to be assayed. Practical considerations also favour use of DNA as it can be readily stored and potentially sampled from museum/herbarium specimens. In comparison to protein assays very little tissue is necessary for DNA analyses. A further advantage is the possibility to derive independent data from each of the three plant genomes mitochondrial, plastid and nuclear. These have different properties that make them suitable to address diverse questions (Pelser *et al.*, 2003). A potential further application for DNA sequence data is DNA barcoding (Chase *et al.*, 2005; DeSalle *et al.*, 2005; Savolainen *et al.*, 2005). Mitochondrial DNA is rarely (Fig. 2.3) (but see Sanjur *et al.*, 2002) used in species-level plant phylogenetic investigations because its rate of evolution is slow (Wolfe *et al.*, 1987), such that levels of sequence variation are generally too low to be informative at lower taxonomic levels. Additionally, the genome shows a high level of genomic rearrangement, complicating the scoring of molecular markers (Palmer, 1992). In contrast, data derived from mtDNA loci are universally applied to studies of animal phylogeny (e.g. Giuffra *et al.*, 2000) because in animals the mitochondrial genome is maternally inherited and has a rapid rate of sequence evolution.

Both molecular markers and sequence data have been used by systematists for species-level phylogeny reconstruction (Table 2.1 adapted from Gepts, 1993). Each system offers advantages and disadvantages. Sequence data are often preferred over other sources of DNA data for species-level phylogeny reconstruction. One obvious reason for this is that the genome of origin is usually known for DNA sequence data whereas many other molecular

markers produce a mixture of biparentally and maternally inherited markers complicating interpretation of trees (Harris, 1999). Furthermore, sophisticated models of sequence evolution provide additional information both for tree building and other analyses such as rates analysis to estimate divergence times (Felsenstein, 2004a).

Table 2.1 Comparison of properties of different sources of DNA data

DNA data source	DNA Molecular Markers					Sequence
	Random Amplified Polymorphic DNA (RAPD)	Simple Sequence Repeats (SSR) Microsatellites	Inter Simple Sequence Repeats (ISSR)	Restriction Fragment Length Polymorphism (RFLP)	Amplified Fragment Length Polymorphism (AFLP)	
Amount of variation	Low - High	Low - High	Low - High	Medium - High	High -Very High	Low - High
Number of loci	Medium- High	Medium - High	Medium - High	High	High	Low - High
Reproduci-bility	Low	Medium - High	Medium - High	High	High	High
Method	Quick	Medium	Medium	Slow	Medium - Fast	Medium - Fast
Cost	Cheap - Medium	Medium	Cheap - Medium	Expensive	Medium	Medium - Expensive
Amount of DNA needed	Low	Low - Medium	Low - Medium	High	High	Low
Type	Dominant	Co-dominant	Co-dominant	Co-dominant	Dominant	Co-dominant
Homology assessment	Very low reliability	Very low reliability	Low-medium reliability	Low-medium reliability	Medium reliability	Medium-high reliability

2.2.1 Molecular markers

Molecular markers allow sequence variation to be assayed indirectly from the number, size and conformation of DNA fragments on electrophoretic gels or detected through fluorescent tagging (Hillis *et al.*, 1996). Marker systems are designed to highlight polymorphisms in size or conformation representing nucleic acid changes at restriction or priming sites. They provide relatively cheap systems which can be automated for high through-put screening (Bussell *et al.*, 2005). Molecular markers are more commonly applied to population studies (Cubas *et al.*, 2005; El-Rabey *et al.*, 2002; He *et al.*, 2003) than species-level phylogenetic reconstruction but are increasingly utilised to build phylogenies for groups for which sequence data has so far failed to produce well-supported resolution (Koopman, 2005; Després *et al.*, 2003; Pelser *et al.*, 2003; Rossetto *et al.*, 2002b). Molecular marker data are usually analysed by phenetic or cladistic methods and normally are highly congruent with sequence based gene trees (Spooner *et al.*, 2005) but worryingly in some cases (Powell *et al.*, 1996) little agreement has been found between different marker systems.

A range of molecular markers have been utilised in species-level investigations (Table 2.1). Trees constructed from marker data are often more resolved than those built from sequence data (e.g. Ajibade *et al.*, 2000; Bullitta & Hayward, 1996; Friesen *et al.*, 1999; Gepts, 1996; Gimenes *et al.*, 2002; Koopman *et al.*, 2001; Lambertini *et al.*, 2006; Le Thierry d'Ennequin *et al.*, 2000; Liu & Wendel, 2001; Marhold *et al.*, 2004; Muellner *et al.*, 2005; Roman *et al.*, 2003; Rossetto *et al.*, 2002b; Vos *et al.*, 1995). Microsatellites and AFLPs are the most commonly used markers as they are more repeatable than RAPDs (Felsenstein, 2004b; Jones *et al.*, 1997).

There are four main concerns with the use of molecular markers for phylogenetic analysis; band homology, reproducibility, use of negative evidence and analysis methods. Multiple studies have highlighted non-homology of co-migrating bands (See Section 3.5.8) (Bailey *et al.*, 2004; Rieseberg, 1996; Smith *et al.*, 1994; Thormann *et al.*, 1994) suggesting that such data are not reliable for phylogenetic analysis. Secondly, it is known that repeatability of markers varies with some e.g. RAPDs being notoriously poor (Jones *et al.*, 1997; Penner *et al.*, 1993; Weising *et al.*, 1995). Also, it has recently come to light that marker profiles of nDNA from different tissues may vary (Chakrabarti *et al.*, 2006) suggesting that choice of source DNA may affect results. Thirdly, scoring absence of a band as a character is unsatisfactory as failure to amplify can be due to technical reasons as well as true absence of primer sites. In terms of analysis methods, although computer algorithms such as BIOSYS-1 (Swofford & Selander, 1981) are available, marker evolution is poorly understood so that such models are rudimentary (Felsenstein, 2004b). The suitability of microsatellite data for phylogeny reconstruction is also questionable because of the rapid evolution of repeats which may mask relationships and mean some changes are untrackable. For example, if an addition and a loss of a repeat occur at the same time then base pair variation will not be detected (Wolfe *et al.*, 1987). In order to have confidence in alignment and fixed differences, sampling requirements are far greater than those currently applied in most species-level phylogenetic analyses. Additionally, the neutrality of microsatellites has been questioned and variation in repeat number has been proposed to contribute to quantitative traits (Fondon & Garner, 2004; Kashi *et al.*, 1996).

2.2.2 Chloroplast DNA sequence data

Sequence data derived from cpDNA account for ca. 60% of published data matrices for plant species-level phylogeny reconstruction (Fig. 2.3) (Appendix 2, Hughes *et al.*, 2006). Chloroplast loci are favoured because they are often variable at the species-level and are easy to amplify with universal primers (Shaw *et al.*, 2005b; Taberlet *et al.*, 1991). This can be attributed to the small genome size and high copy number (Baldwin *et al.*, 1995). However, cpDNA loci have several unfavourable characteristics including non-independence, generally maternal inheritance pattern, low intraspecific sequence divergence (Small *et al.*, 2004; 1998) and haplotypes that are commonly shared between species (McKinnon *et al.*, 2004). Chloroplast data alone, especially when multiple loci are combined, can in some cases produce relatively well-resolved species-level phylogenies (Pardo *et al.*, 2004; Rouhan *et al.*, 2004) but are more commonly utilised in combination with nDNA or nrDNA. However, chloroplast trees on average show lower levels of resolution than those built from nDNA or nrDNA (Fig. 2.2b) (Clarkson *et al.*, 2004; Appendix 2, Hughes *et al.*, 2006; Shaw & Small, 2005).

There are a large number of potential non-coding cpDNA loci (Kress *et al.*, 2005; reviewed by Shaw *et al.*, 2005b; Small *et al.*, 1998) but sequencing efforts have been dominated by trnL-trnF and trnK-matK (Olmstead & Palmer, 1994; Shaw *et al.*, 2005b). This may be due in part to limited information surrounding the phylogenetic utility of different loci. Shaw *et al.* (2005b) addressed this by sampling three species across ten groups for 21 non-coding cpDNA loci. They found that different loci evolve at different rates in different groups but overall trends in rate of evolution were visible. TrnL-trnF was amongst the five loci with the highest number of potentially informative characters (including indels) whereas trnK-matK performed poorly. This suggests that locus utility is not reliably predictable and that carrying out a preliminary study to screen candidate loci is advisable (Shaw *et al.*, 2005b).

Chloroplast DNA sequences are technically easier to work with than low- or single-copy nuclear gene regions but tend to show less sequence variation and fewer parsimony

informative characters (Fig. 2.4) (but see Cronn *et al.*, 2002; Appendix 2, Hughes *et al.*, 2006). An extreme example is shown by Small *et al.* (1998) who sequenced 7 kb from *Gossypium* L. cpDNA and found only four informative characters. In contrast, a 1.65 kb region of the nuclear gene alcohol dehydrogenase C showed 25 parsimony-informative characters which were sufficient to build a resolved tree for eight taxa (Small *et al.*, 1998).

Chloroplast loci can provide sufficient PICs to build well-resolved phylogenies for many genera (e.g. Cieslak *et al.*, 2005; Clarkson *et al.*, 2004; Drummond, unpublished manuscript; Kelchner & Clark, 1997; Knox & Palmer, 1995; Shaw & Small, 2005; Soltis *et al.*, 2001). However, in some cases, large amounts of sequence were necessary to achieve this (e.g. 4656 bp Clarkson *et al.*, 2004). In the survey of species-level gene trees, 15% of studies utilising single cpDNA loci alone produced fully resolved trees, while 70% of cpDNA gene trees had a consensus fork index of 0.7 or less (Fig. 2.2b) (Appendix 2, Hughes *et al.*, 2006). Thus, chloroplast loci are good for constructing a framework but are often insufficient for fine scale resolution. Therefore they are good for initial broad-scale studies.

Although not the most variable regions available, cpDNA loci are invaluable as independent data sets. For this reason many studies compare and/or combine loci from the chloroplast and nuclear genomes (e.g. Archibald *et al.*, 2005; Berry *et al.*, 2004; Bolmgren & Oxelman, 2004; Chat *et al.*, 2004; Comes & Abbott, 2001; Feng *et al.*, 2006; Guzman & Vargas, 2005; Hayashi *et al.*, 2001; Kim *et al.*, 1998; Lee *et al.*, 2004; Liu *et al.*, 2002; Milne, 2004; Oh & Potter, 2005; Pardo *et al.*, 2004; Pelser *et al.*, 2002; Rossetto *et al.*, 2002a; Von Hagen & Kadereit, 2001; Wilkin *et al.*, 2006; Yi *et al.*, 2004; Zhou *et al.*, 2006), particularly for the identification of hybrids using the difference in cpDNA and nDNA inheritance patterns (Chat *et al.*, 2004; Ellison *et al.*, 2006; Gross *et al.*, 2003; Hughes *et al.*, 2002).

2.2.3 Nuclear ribosomal DNA sequence data

Nuclear ribosomal DNA is found in one or more nucleolar organization regions. In plants

each nrDNA copy has the structure ETS - 18S - ITS1 - 5.8S - ITS2 – 26S – ETS, a transcriptional unit, found in thousands of consecutive repeats (Baldwin, 1992; 1995). The coding regions 18S, 5.8S and 26S are highly conserved but the internal transcribed spacers (ITS1 and ITS2, collectively known as ITS) are non-coding with higher rates of sequence evolution. The popularity of ITS as a source of sequence data is only equalled by cpDNA loci (Alvarez & Wendel, 2003; Shaw *et al.*, 2005b). A survey of 345 matrices generated for species-level phylogeny reconstruction found that the single most commonly utilised locus was ITS (94 matrices) (Fig. 2.3.) (Appendix 2, Hughes *et al.*, 2006).

The widespread use of ITS (Alvarez & Wendel, 2003) can be attributed to a combination of characteristics which make ITS an effective tool for phylogenetic work (Baldwin, 1992; 1995). Reliable universal polymerase chain reaction (PCR) primers have been developed in the highly conserved flanking regions and both ITS1 and ITS2 are less than 300 bp in length which facilitates the speed and ease of sequencing.

A high level of ITS sequence divergence is often seen between species such that ITS is informative at lower taxonomic levels. Although the number and percentage of PICs exhibited within congeneric ITS matrices vary widely (Fig. 2.4) (Appendix 2, Hughes *et al.*, 2006) on average they are higher than cpDNA and somewhat surprisingly also greater than nDNA introns (see discussion below). In fact ITS (Fritsch, 2001; Miller *et al.*, 1999; Mort *et al.*, 2004) exhibits the highest percentage of PICs so far recorded in species-level datasets (Fig. 2.4) (Appendix 2, Hughes *et al.*, 2006).

Alignment of ITS sequences between closely related taxa is usually unproblematic, eased by the clarity of the coding region boundaries (Baldwin, 1992). This, in addition to the high copy number, which increases the success of PCR amplification, makes ITS an easy, quick and cheap region to utilise. Complete or near complete concerted evolution is important because it homogenises the paralogues such that it does not matter which copy is amplified and compared (Baldwin, 1992; 1995). Few other genome regions have as many advantages for use in phylogenetic studies.

However, as the use of ITS has become more widespread, problems and concerns have been encountered (reviewed in Alvarez & Wendel, 2003; Bailey *et al.*, 2003; Hughes *et al.*, 2002). The precise mechanisms of concerted evolution are not known and the occurrence of individuals which are polymorphic, including potentially non-functional pseudogene copies, (Bailey *et al.*, 2003; Mayol & Rosselló, 2001; Samuel *et al.*, 1998) or chimeric ITS sequences suggest incomplete concerted evolution may be much more common than previously thought (Barkman & Simpson, 2002). If ITS sequences are not fully homogenised then care needs to be taken when building gene trees to ensure that orthologous copies are used. Conversely, if concerted evolution is complete, then ITS is less useful for identifying hybrids (Alvarez & Wendel, 2003; Wendel *et al.*, 1995).

The universality which makes ITS so attractive to phylogeneticists also contributes to one of its drawbacks, that of contamination (Camacho *et al.*, 1997; Zhang *et al.*, 1997). The highly conserved ITS primer sites mean that DNA from non-target species may amplify readily under the same conditions. The increasing availability of larger amounts of sequence data should allow contamination to be readily detected. Alvarez and Wendel (2003) also emphasised the high levels of homoplasy found in ITS datasets compared to other regions. Although some consider high homoplasy a disadvantage (Givnish & Sytsma, 1997) highly homoplastic datasets can still contain phylogenetic signal (Källersjö *et al.*, 1999; Von Hagen & Kadereit, 2002).

The success of ITS is demonstrated by its routine application to species-level phylogenetic studies often providing large numbers of PICs (Fig. 2.4) and medium to high levels of resolution (Fig. 2.2b) (Appendix 2, Hughes *et al.*, 2006; Marshall *et al.*, 2001; Miller *et al.*, 1999). However, an increasing number of studies find insufficient variation in ITS datasets to build resolved and robustly supported trees (Bayer, 1996; Emshwiller & Doyle, 1998; Hodges & Arnold, 1994; Miller *et al.*, 1999; Richardson *et al.*, 2001a; Samuel *et al.*, 1998; Schilling *et al.*, 1998; Sonnante *et al.*, 2003; Vargas *et al.*, 1998; Wink *et al.*, 1999). Poorly resolved ITS gene trees are seen particularly in studies of domestication and putative rapid

radiations (Emshwiller & Doyle, 1998; Harris *et al.*, 2000; Richardson *et al.*, 2001a; Wink *et al.*, 1999; Wojciechowski *et al.*, 1999; Xu & Sun, 2001). As discussed earlier there are several possible reasons for lack of resolution. The most common hypothesis to explain the lack of resolution within trees is that periods of rapid speciation have occurred (Small *et al.*, 1998) where accelerated morphological evolution has not been accompanied by significant molecular evolution because the rate of ITS evolution has remained constant (Bateman, 1999).

2.2.4 Nuclear DNA sequence data

The most common approach to resolve relationships among closely related species where cpDNA and ITS loci fail has been to use sequence data from low copy nuclear genes (LCNG) or specific members of multi-gene families and specially their introns (Cronn *et al.*, 2002; Doyle & Doyle, 1999; Mort & Crawford, 2004; Sang, 2002; Small *et al.*, 2004). Initial studies using nuclear sequences were restricted by the handful of LCNGs (e.g. floral development genes) under investigation in related taxa in other fields (Cronn *et al.*, 2002; Small *et al.*, 1998; Strand *et al.*, 1997). This approach has succeeded in locating sporadic highly variable loci for specific plant groups (e.g. Histone H3-D in *Glycine* L., Doyle *et al.*, 1996; and Rosaceae, Evans *et al.*, 2000; pgiC1 and pgiC2 in *Clarkia* Pursh., Ford & Gottlieb, 2003; PRK and RPB2 in *Areca* L., Loo *et al.*, 2006; GBSSI/waxy in grasses, Mason-Gamer *et al.*, 1998; cycloidea in *Lupinus*, Ree *et al.*, 2004; GPAT in *Paeonia* L., Tank & Sang, 2001). However, LCNG loci still comprise less than 10% of published species-level matrices (Fig. 2.3) (Appendix 2, Hughes *et al.*, 2006) and for the vast majority of plant genera, choice of LCNG loci remains limited and access to alternative loci a significant hurdle. Published studies utilising nDNA are on average more resolved than nrDNA or cpDNA gene trees (Fig. 2.2b) (Appendix 2, Hughes *et al.*, 2006) despite lower average PICs (Fig. 2.4a). This is likely to be due to a few extremely variable nDNA loci increasing the average value (Fig. 2.4a) (Appendix 2, Hughes *et al.*, 2006).

Nuclear loci are often technically challenging to work with. This is because they occur in low copy number, often show allelic variation, may have secondary structure and complex evolutionary dynamics e.g. recombination, concerted evolution and selection (Small *et al.*, 2004). Recombination resulting in chimeric sequences can hamper phylogeny reconstruction and increase the likelihood of reconstructing an incorrect tree (Poke *et al.*, 2006; Posada & Crandall, 2002).

In addition to problems of data collection, single or low-copy nuclear regions may also be subject to paralogy problems due to the presence of multiple copies of a gene following gene duplication. If paralogy is not recognised, an incorrect species tree may result (Martin & Burg, 2002). In theory paralogy should be a greater threat at higher taxonomic levels where there has been more opportunity for gene duplications and deletions (Sang, 2002). The copy number of genes can vary considerably among taxa (Zhou & Kleinhofs, 1996) and even within genera (Scherson *et al.*, 2005). The putative single copy genes, g3pdh used by Olsen & Schaal (1999), pistillata used in the phylogeny reconstruction of *Sphaerocardamum* S. Schauer (Bailey & Doyle, 1999) and waxy have all been found to be multi-copy in other taxa. Multiple copies can be detected by Southern blotting, the presence of multiple bands on a gel, double peaks in sequence traces, cloning, in situ hybridisation studies and assessing tree topology (Hughes *et al.*, 2002; Scherson *et al.*, 2005).

It is also increasingly clear that nuclear loci vary widely in nucleotide substitution rates among closely related species. For example Cronn *et al.* (2002) found a five-fold range in substitution rates among 12 nuclear encoded loci developed for *Gossypium*. In combined analysis of these nuclear loci plus ITS, 53% of the PICs were recovered from just two loci (ITS and Fad2-1). An even wider seven-fold difference in substitution rates was documented in a survey of 36 nuclear genes for the same genus (Senchina *et al.*, 2003). There are clear indications of similar variability among rates of evolution among intron sequences of closely related species from other genera (Doyle *et al.*, 2003a; Doyle *et al.*, 2004 for *Glycine*; Syring *et al.*, 2005 for *Pinus* L.). A wider survey of nuclear intron data sets confirms this variability

which ranges from 0.21% to 16% PICs across groups of congeneric species (Fig. 2.4b) (Appendix 2, Hughes *et al.*, 2006). Of the 24 LCNG intron data sets surveyed by Hughes *et al.* (2006, Appendix 2) which have ITS data for the same taxa, only seven provided more PICs than ITS (Fig. 2.4a). While this variability has been recognized for some time (Small *et al.*, 2004), what has not been widely acknowledged is that the majority of nuclear introns sequenced across groups of species so far, are less variable (and often much less variable) and provide fewer PICs than ITS. Only a small proportion exhibit similar or higher levels of variation (Fig. 2.4a and b) (Appendix 2, Hughes *et al.*, 2006). This is perhaps not surprising considering that some introns have been shown to influence patterns of gene expression, acting as cis-regulatory elements (e.g. Sieburth & Meyerowitz, 1997). A similarly wide span of levels of informative variation has also been found for LCNG loci used in phylogenetic analyses of animal species where many LCNG sequences used so far are less variable than mtDNA loci (e.g. Helbig *et al.*, 2005; Peters *et al.*, 2005).

It is notable that no universally useful LCNG sequence loci have been developed during the last ten years (Sang, 2002). Low copy nuclear loci have had to be developed specifically for each taxonomic group of interest. In fact, this has been the primary goal of many of the studies that have also discussed the potential broader utility of specific genes (e.g. Bailey & Doyle, 1999; Doyle *et al.*, 1996; Emshwiller & Doyle, 1999; Galloway *et al.*, 1998; Mason-Gamer *et al.*, 1998; Ree *et al.*, 2004; Small *et al.*, 1998), a potential that has not so far been realised. Difficulties of extrapolating loci that have provided informative variation in one group can be due to failure of primers to amplify, variation in copy number, differences in intron presence or length, or differences in nucleotide substitution rates in different plant groups. Thus, up to now it seems that development of LCNG sequence loci has been somewhat of a lottery spawning the occasional lucky jackpot against a backdrop of generally disappointing results, failures and considerable investment in developing and optimizing loci that often fail to produce much data (Fig. 2.4) (Appendix 2, Hughes *et al.*, 2006). This is reflected in the experiences of several researchers who have acknowledged an element of good fortune in developing highly informative LCNG sequence loci, when they

come to try to find similarly variable additional loci for the same group of plants (Doyle *et al.*, 2003a; Small *et al.*, 2004).

Nuclear loci are particularly suitable for phylogeny studies investigating hybrid and polyploid origins (Ford & Gottlieb, 1999; Lihova *et al.*, 2006; Sang & Zhang, 1999; Small *et al.*, 1998). This is because unlike cpDNA, nDNA loci are biparentally inherited and the signatures of both parents remain. Furthermore, it is thought that they suffer little from concerted evolution (Small *et al.*, 2004).

Due to the presence of multiple chromosomes and recombination mechanisms, the nuclear genome can provide many independent loci. The benefits of this are three-fold. Firstly, it allows comparison of multiple gene trees for congruence to assess the gene tree species tree problem. Secondly, hypotheses of relationships are strengthened by support from multiple independent lines of evidence (Miyamoto & Fitch, 1995; Small *et al.*, 2004). Lastly, it is common for sequence data from one region to be insufficient to build a resolved phylogeny and the nuclear genome has the potential to provide a large number of independent loci for combined analysis (Small *et al.*, 2004).

2.2.5 Approaches to locate hyper-variable nDNA loci

In the absence of universally accessible and informative LCNG loci, a number of alternative approaches to locate more variable nuclear DNA sequence loci for taxon-specific problems have been developed. These can be divided into three main types; sequence characterized amplified regions (SCARs) (Bailey *et al.*, 2004; Lockhart *et al.*, 2001; McLenachan *et al.*, 2000; Rameau *et al.*, 1998; Shaw *et al.*, 2003; Yang *et al.*, 2001), comparative anchor tagged sequence (CATS) approaches (Chandappa *et al.*, 2005; Choi *et al.*, 2006; Fulton *et al.*, 2002; Lyons *et al.*, 1997; Scherson *et al.*, 2005) and the 3'-untranslated region (3'-UTR) anchored primer method (Whittall *et al.*, 2006).

SCARs are generated through amplification and sequencing of molecular marker bands (Melotto *et al.*, 1996) that can be screened as potentially useful sequence loci prior to development of specific primers (Bailey *et al.*, 2004; McLenachan *et al.*, 2000; Shaw *et al.*, 2003). Under this method, PCR products of equal length amplified across a subset of species using combinations of random primers are excised, cloned, sequenced and aligned to evaluate variability (Fig 3.4, Chapter 3) (Bailey *et al.*, 2004). Specific SCAR primers are then designed for regions that show high levels of variability compared to conventional loci such as ITS. The SCAR-based approach has been used successfully to increase resolution and support for groups of closely related species of legumes (Bailey *et al.*, 2004) and mosses (Shaw *et al.*, 2003, 2005a). The same approach has been applied to restriction fragments (Choi *et al.*, 2006).

There are several potential benefits of using random genomic regions rather than previously characterized genic regions. First, no prior sequence information is required to develop many loci. Second, random amplification of regions across the entire genome eliminates the restriction of working within known genic exon/intron regions (and potentially adjacent 3' UTRs and 5' promoter regions), which are likely to represent the less variable half of the nuclear genome. Given that 53-57% of the *Arabidopsis* L. genome has been classified as non-coding DNA (A. T. G. Initiative, 2000), it presents significant untapped potential as a source of DNA characters (Bailey *et al.*, 2004; Shaw *et al.*, 2003, 2005a). Third, unlike the CATS-based approaches (below), the initial screening process facilitates selection of loci with levels of variation needed to address species-level problems prior to development of specific primers. Finally, by initially selecting for regions with limited length variation this reduces the chances of encountering highly length-variable regions that may be difficult to align (e.g. Syring *et al.*, 2005). This last feature may of course eliminate many potentially useful but length variable DNA sequence loci from consideration (Britten *et al.*, 2003; Fondon & Garner, 2004).

A number of potential drawbacks associated with SCAR-based loci are immediately apparent. First, there is no *a priori* indication that selected loci will be bi-parentally inherited which is critical for studies of reticulation. However, BLAST (Basic Local Alignment Search Tool) searches can be used to check for significant similarity to complete angiosperm cpDNA or mtDNA sequences available in GenBank eliminating known uniparentally inherited loci. Screening via artificial or other known hybrids provides a more certain route to ascertain this. A second potential limitation is the possibility of amplifying regions of junk repeat DNA with consequent paralogy problems. While this might suggest greater probability of encountering paralogy problems with SCAR-based loci, in practice all loci are subject to most, if not all of the same considerations. Finally, there are clearly tradeoffs to be weighed up when looking for SCAR-based loci in terms of the span of taxa to be included in the initial selection of amplified fragments. Inclusion of bands that amplify in just two or three closely related species runs the risk that more divergent taxa will not be amplified by primers developed from the initial sequence data, thereby potentially forfeiting inclusion of outgroup taxa (Bailey *et al.*, 2004). Conversely, selecting fragments that amplify across a wider range of taxa than the specific clade of interest may mitigate against finding the hyper-variable regions that are needed to resolve relationships within that clade. On the plus side, this is under the control of the investigator; different options can be tested and a wider-survey of loci size variation within a representative sample of legumes found little within species variation (Choi *et al.*, 2006).

Comparative anchor tagged sequence approaches involve comparisons of expressed sequence tags (EST) and/or complete genome sequences between model or crop organisms to identify evolutionarily conserved regions, termed comparative anchor tagged sequences (Lyons *et al.*, 1997; Syring *et al.*, 2005) or conserved ortholog set (COS) markers (Fulton *et al.*, 2002). While the primary goal of many of these studies has been comparative genome mapping, CATS markers can be used to develop sets of primers circumscribing potentially amplifiable sequence regions for phylogenetic analyses. For example, comparison of the tomato EST database with the *Arabidopsis* genome sequence identified a set of >1,000 COS

markers (Fulton *et al.*, 2002). Similarly, screening of the *Medicago truncatula* Gaertn. EST database against the *Arabidopsis* genome sequence, and other available legume sequence, identified 274 loci that show strong sequence similarity. A subset of these are being screened as potential DNA sequence loci for species-level analysis of the large legume genus *Astragalus* L. (Scherson *et al.*, 2005). This approach has also been used directly to identify conserved DNA oligomers that can be employed as primers to amplify orthologous DNA sequence loci for species-level phylogenetics by comparing the complete *Arabidopsis* and rice genome sequences (Padolina *et al.*, 2004; Xu *et al.*, 2004). Choi *et al.* (2006) compared the ESTs from *Medicago trunculata* and *Glycine max* Merr. and identified 196 EST-derived primer pairs. Initially, amplification success and product size were assessed using six Papilionoid legumes. Wider taxon sample was utilized for promising loci. This approach identified a large number of loci many of which amplified in more than four species and which the authors suggest may be useful even at the subspecies-level. The majority of the loci were intron-spanning and "most primer pairs amplified sequences with insufficient informative characters to produce well-resolved gene trees" (p 66, Choi *et al.*, 2006). As previously discussed, rate of sequence evolution in most introns is lower on average than ITS (Fig. 2.4a) so there is no guarantee or data to suggest that this method is destined to locate any hypervariable loci (if they exist). The legume primers designed by Choi *et al.* (2006) were published right at the end of the study so were not available for screening on *Lupinus*.

CATS-based approaches as currently implemented, generally compare EST/genomic sequences between widely divergent taxa (e.g. rice and *Arabidopsis*) (Appendix 2, Hughes *et al.*, 2006). While this may be advantageous for locating widely conserved markers (Fulton *et al.*, 2002) thereby facilitating design of potentially 'universal' primers, it also means that primers typically span relatively more conserved coding regions. The simple fact that alignments can be constructed between such divergent taxa suggests that such regions are less likely to provide highly variable loci between closely related taxa. However, this is presumably a temporary limitation of our current data bases that include few taxa with many loci, while many taxa have very few (or no) loci (e.g. Driskell *et al.*, 2004). As the taxon

x loci matrix becomes more densely populated, future CATS-based comparisons among more closely related taxa will become increasingly powerful for locating DNA sequence loci for specific groups of interest. For example, mapped conifer anchor loci developed from two species of pine (Brown *et al.*, 2001) are being used to screen and select potential DNA sequence loci to resolve incongruent and incompletely resolved cpDNA and nrDNA phylogenetic trees for *Pinus* (Syring *et al.*, 2005). Other examples of comparisons across narrower taxonomic spans illustrate this potential (e.g. in Brassicaceae, Kuittinen *et al.*, 2002; Lukens *et al.*, 2003).

The 3'-UTR method aims to amplify orthologous products in a range of taxa by creating specific primer pairs (Whittall *et al.*, 2006). Specificity is introduced through priming sites in the 3'-UTR which is thought to evolve rapidly. The 3'-UTR anchored primer method requires a cDNA library for one taxon. These are then sequenced and BLAST searched against GenBank records. This enables nuclear coding sequences to be selected. For each potential locus, primer pairs are designed, one in the putative exon and one in the 3'-UTR region. PCR reactions of genomic DNA can be used to investigate intron presence and length. One advantage of this method is starting with sequence from the taxa of interest which significantly increases the chances of good primer match and amplification success.

None of the methods discussed above have succeeding in locating hyper-variable nDNA loci. They have however found loci equivalent to ITS or currently known nDNA loci. These methods open up a new source of multiple anonymous DNA which may be sufficiently variable in some groups to built resolved trees.

One attribute these methods have in common is the ability to produce large numbers of potential loci (Appendix 2, Hughes *et al.*, 2006). This presents the necessity for an efficient screening process and allows rigorous criteria to be implemented. Criteria for rejection implemented in screening include: band size variation, low levels of sequence variation, multiple PCR products, double peaks in chromatograms and high numbers of single

nucleotide polymorphisms (SNPs) (Scherson *et al.*, 2005; Whittall *et al.*, 2006). Assessing homology from agar gels is inaccurate as bands of the same size may represent different products (Bailey *et al.*, 2004; Harris, 1999). Size variation can be an indication of indels which frequently represent a variable locus however, it can also indicate that the bands are not homologous. Quantitative knowledge of sequence variation within existing data sets is vital as these can be used as benchmarks for assessing new loci. However, multiple PCR products as indicated through double bands on agar gels are an indication of the presence of multiple copies of a locus or heterozygotes with divergent alleles. Multiple products may be cryptic, i.e. very similar in size, and only be recognisable through the presence of double peaks on chromatograms and high number of SNPs or in some cases through tree building methods. These regions are to be avoided as they are technically challenging to work with and it is difficult to ensure that the same copy is being amplified in each accession. In studies up to now, it has been found that adhering to these criteria results in the rejection of the majority of loci (Bailey *et al.*, 2004; Scherson *et al.*, 2005; Whittall *et al.*, 2006).

Taxon sampling used in the preliminary development of a locus may have an impact on the applicability of the locus to a wider range of taxa. Bailey *et al.* (2004) and Whittall *et al.* (2006) found that there is a compromise to be made regarding the sampling breath and the taxonomic scope of amplification. Bands which also amplify in the outgroup are less likely to be variable enough to resolve relationships within the ingroup, but lack of outgroup amplification presents rooting problems. Shaw *et al.*(2005b) concluded that a sample of three taxa is sufficient to assess the properties of a potential locus. However, this assumption depends on whether the taxa sampled represent all major lineages in the group of interest, and I believe that a larger sample is preferable.

2.3 Conclusion

Although both PCR-based marker and DNA sequence data have been applied to species-level phylogeny reconstruction the latter is far more widely implemented. ITS and cpDNA loci have made, and will continue to make, a substantial contribution to species-level phylogeny reconstruction. For most taxa gene trees built from these loci form at least a good backbone which can direct further character and taxon sampling. However, it is becoming increasingly clear that lineage-specific solutions will be required to produce highly resolved and strongly supported species-level trees. The hunt for more variable loci has focused on nDNA introns but the majority of those sequenced so far are less variable than ITS. Although various methods are now available to locate new nDNA loci, hyper-variable loci have generally not been forthcoming. Until they are, comparing and combining data from multiple independent loci presents the most satisfactory approach to species-level phylogeny reconstruction.

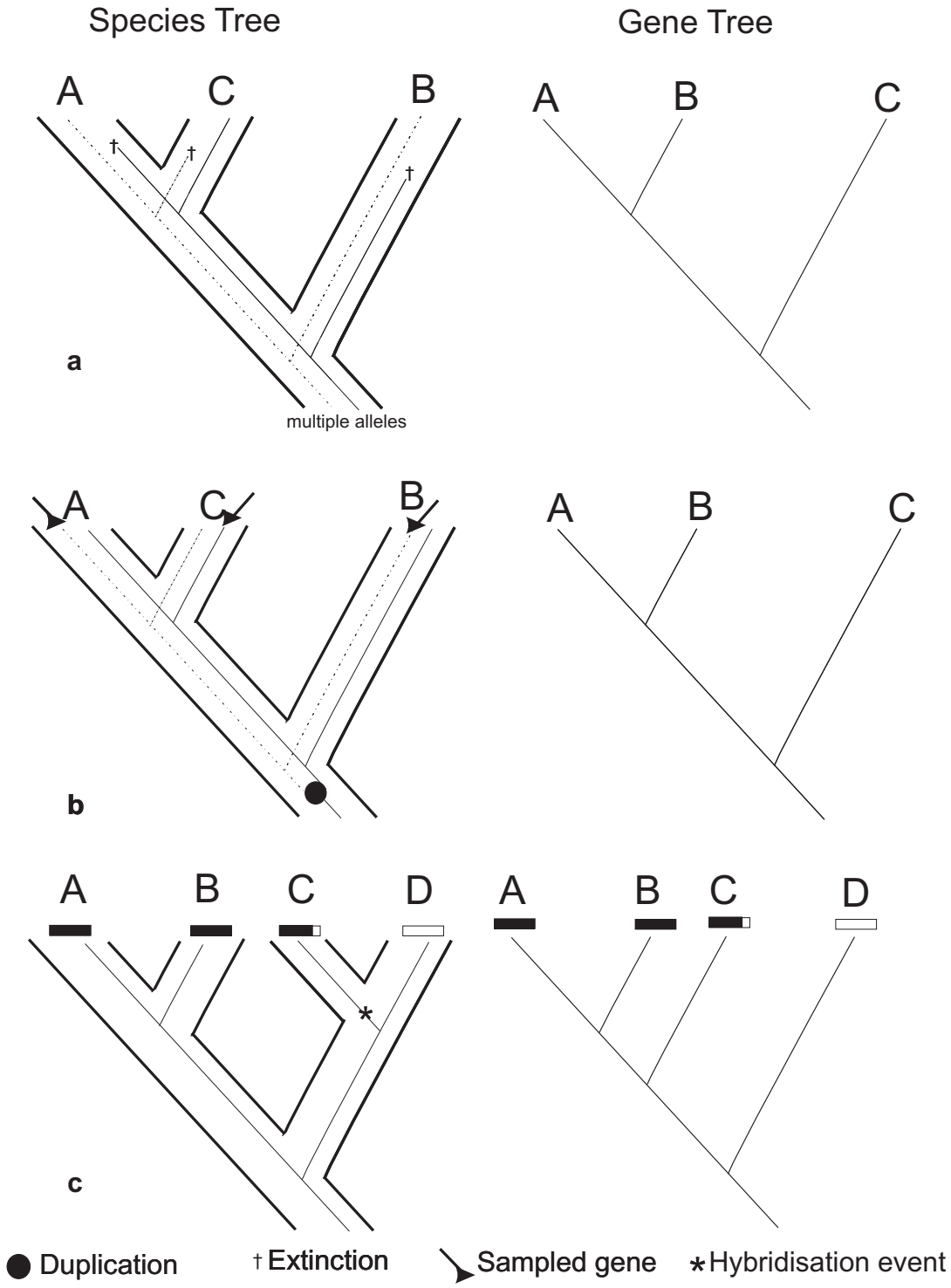
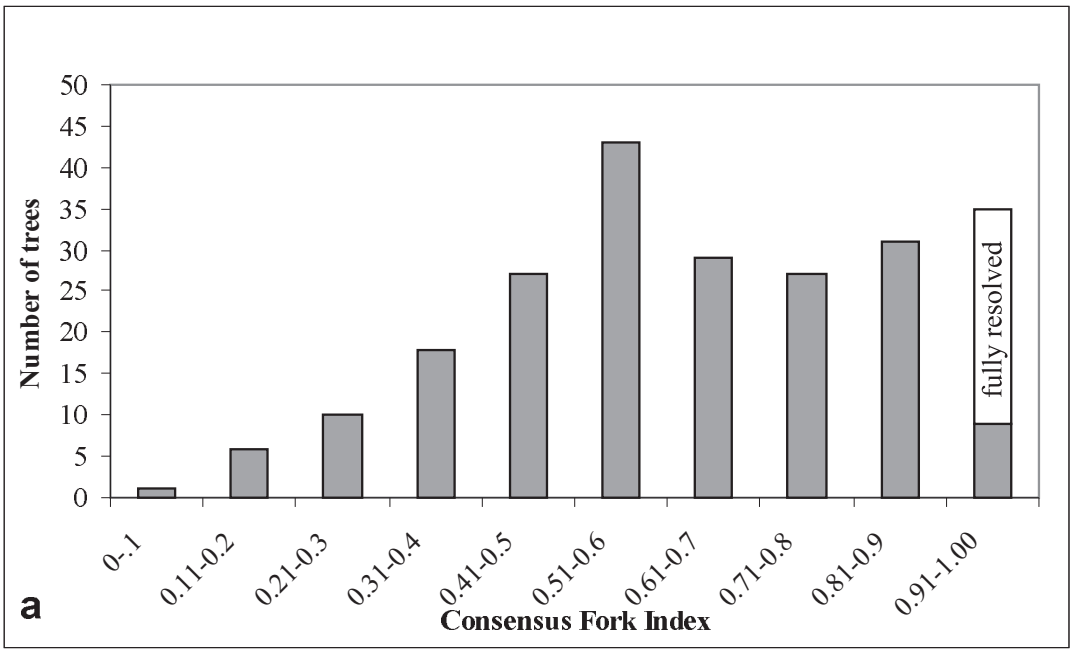


Fig. 2.1 Species trees versus gene trees. The species tree is show by the solid tree outline, and the gene lineages are shown within the species tree vessel. Situations in which gene trees can provide a misleading estimate of species trees. a. Incomplete lineage sorting. Alleles are differentially retained within the taxa sampled such that the estimated species tree is not correct. b. Gene duplication and differential sampling. Duplicate genes may be retained however incomplete or differential sampling of both copies can cause the gene tree not to reflect the species tree. c. Hybridisation. Hybridisation can result in chimeric sequences which, if unrecognised, can lead to the reconstruction of gene trees not reflecting the species tree. In this example taxon C is a hybrid between the lineage leading to taxon D and that leading to A and B. Taxon C contains sequence from both parental lineages (as illustrated by shaded boxes). Modified from Martin & Burg (2002).



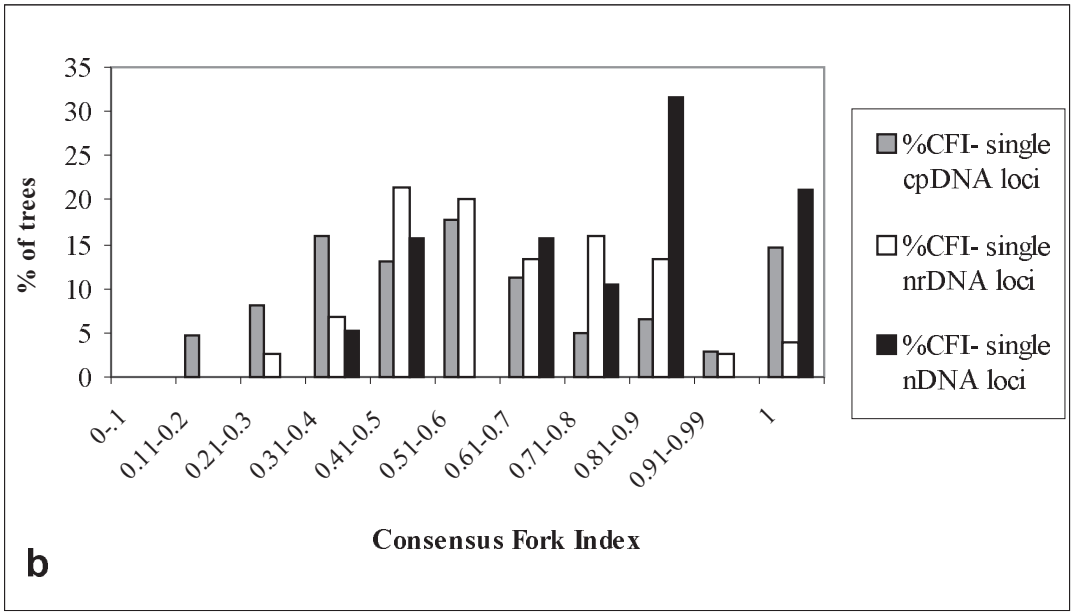


Fig. 2.2 Resolution obtained in recent species-level phylogenetic studies. Survey based on 345 matrices from 136 studies (see Appendices 2 and 4). a. Resolution obtained in recent species-level analyses across all loci: frequency distribution of consensus fork indices, including both combined and separate analyses. b. Resolution obtained using different loci: % frequency distribution of consensus fork indices for analyses of individual loci.

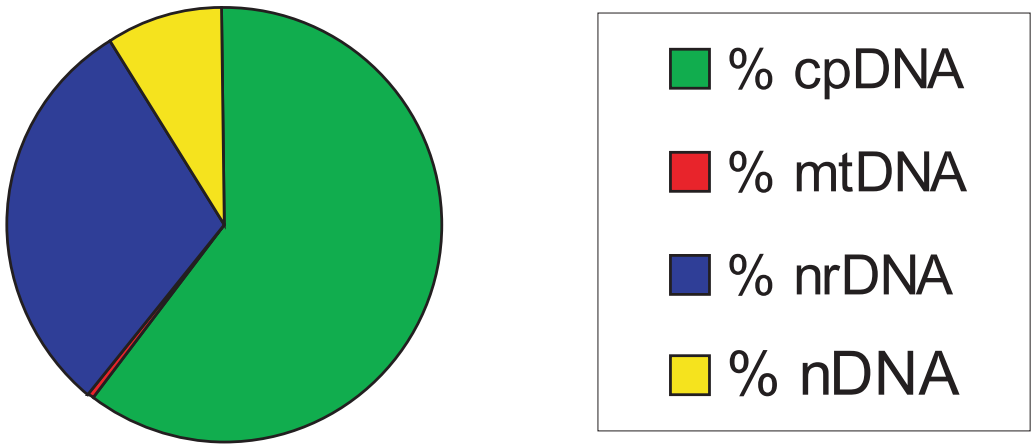


Fig. 2.3 Loci used in species-level phylogenetic studies. Survey based on 345 matrices from 136 studies (see Appendices 2 and 4). Loci currently used in species-level analyses: proportions of cpDNA, nrDNA, mtDNA and nDNA loci.

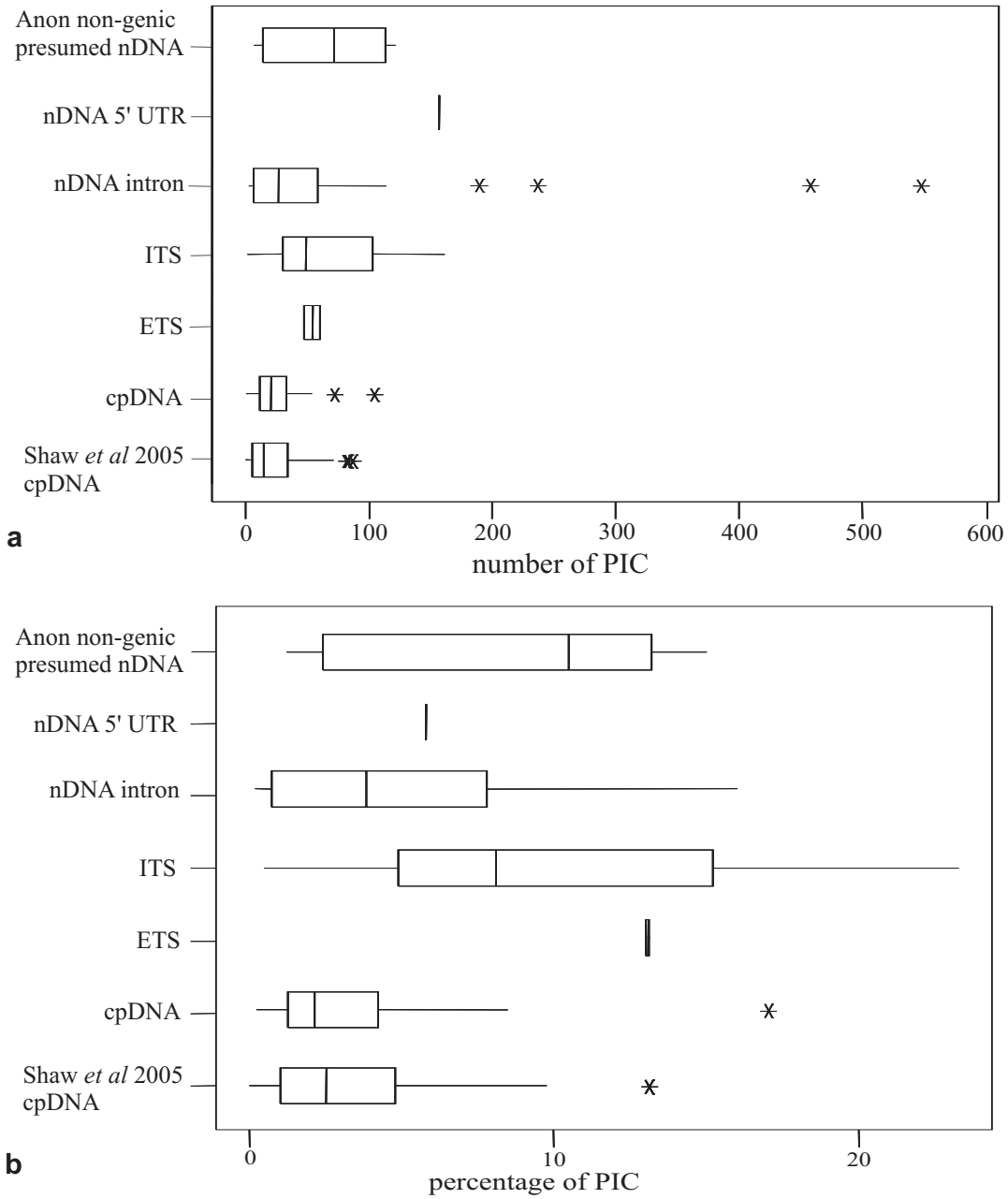


Fig. 2.4 Potential informativeness different types of DNA sequence loci in plant species-level phylogenetic analyses. a. Number of PICs per locus. b. Percentage of PICs for individual loci. Boxes encompass the 75% interquartiles, showing the median while outliers are indicated with a. Data are derived from species-level phylogenetic analyses of 226 matrices from studies of plant genera. Character statistics were calculated for congeneric terminals only, excluding outgroup taxa. Tallies of characters included both PI substitutions and PI indels coded according to the simple gap coding method of Simmons & Ochoterena (2000). Plant groups included are: *Sphagnum* L. (Shaw *et al*., 2003), *Leucaena* Benth. (Bailey *et al*., 2004 and unpublished data), *Gossypium* L. (Cronn *et al*., 2002), *Gaertnera* Lam. (Malcomber, 2002), *Sphaerocardamum* S. Schauer (Bailey & Doyle, 1999), *Passiflora* L. (Yockteng & Nadot, 2004a, b), *Glycine* Willd.-alignments assembled for nine species based on data from Doyle *et al*. (1996, 2003a, and b), and Sakai *et al*. (2003), *Paeonia* L.-alignments assembled for eight species based on data from Sang *et al*. (1995, 1997), Ferguson & Sang (2001) and Tank & Sang (2001), *Lycopersicum* Hill.-alignments assembled for 13 accessions of seven species based on data from Nesbitt & Tanksley (2002), *Viburnum L.* (Winkworth & Donoghue, 2004), *Clarkia* Pursh (Ford & Gottlieb 2003), *Hibiscus* L. (Small, 2004), *Neillia* D.Don/*Stephanandra* Siebold & Zucc. (Oh & Potter, 2003) and *Bursera* Jacq. ex L.(Weeks & Simpson, 2004). The broader survey of non-coding cpDNA loci (Shaw *et al.* 2005 b) is included as a second cpDNA bar. From Hughes *et al.* (2006), Appendix 2.

3 Phylogeny and Biogeography of *Lupinus*

3.1 Previous phylogenetic studies

A broad range of data sources (including morphology (Bentham, 1859; Gladstones, 1974, 1984; Watson, 1873), serological (Cristofolini, 1989; Nowacki & Jaworski, 1978), flavanoids (Nicholls & Bohm, 1983), isozymes (Wolko & Weeden, 1990a, b), cpDNA RFLPs (Badr, 1994) and, more recently, DNA sequences (Ainouche & Bayer, 1999; 2000; Drummond, 2004; Kass & Wink, 1996; 1997a; 1997b; Ree *et al.*, 2004)) have been utilised in attempts to understand relationships within *Lupinus* but few have explicitly utilised phylogenetic methods (Ainouche & Bayer, 1999, 2000; Badr, 1994; Bayer, 1996; Kass & Wink, 1996, 1997a; Ree *et al.*, 2004; Wink *et al.*, 1999). Prior to this study, only cpDNA rbcL and nrDNA ITS sequence data had been utilised in phylogeny reconstruction of *Lupinus* (Ainouche & Bayer, 1999, 2000; Kass & Wink, 1996; 1997a; 1997b). During the course of the project additional cpDNA loci (trnT-L spacer, trnS-G and trnL intron - Drummond, 2004) and a nDNA locus (LEGCYC - Citerne *et al.*, 2003; Ree *et al.*, 2004) were developed.

All *Lupinus* phylogeny reconstruction studies to date have sampled very few of the South American taxa (Ainouche *et al.*, 2004; Kass & Wink, 1997a; Merino *et al.*, 1999; Ree *et al.*, 2004; Wink *et al.*, 1999). For the most part this reflects sparse overall taxon sampling and taxonomic confusion over species delimitation but there has also been a bias towards economically important species, most of which are distributed in the Old World. For example, even the most densely sampled analysis by Ainouche and Bayer (1999) (Table 3.1 and Fig. 3.1) included all but one of the Old World species, yet sampled just four South American taxa from an estimated total of 85 species (Planchuelo, 1984). The Mexican species were also extremely poorly sampled, represented by just two species (Fig. 3.1). Furthermore, only two of the ca. 20 species of non-digitate leaved species have been sampled.

Previous phylogeny estimates have been limited not only by sparse and uneven taxon sampling, but also by lack of resolution and low statistical support for clades. Despite differing analysis methods and taxon sampling among analyses, the monophyly of the genus has been consistently supported, but relationships between the Old World species, between the Old and New World species and the presence of two independent South American lineages, remain tentative (Fig. 3.1) (Ainouche & Bayer, 1999). The most informative source of data used to date has been ITS. While this has provided a framework of relationships within the genus, ITS fails to resolve branching order at the base of the tree or relationships between New World taxa, most of which form a large unresolved group that includes North American, Mexican and Andean species (Fig. 3.1) (Ainouche & Bayer, 1999). Scrutiny of the ITS data matrix shows that within *Lupinus* the lack of resolution in the ITS consensus tree is a consequence of low DNA sequence variation rather than character conflict. One general observation from previous studies is that greater variation be it of isozymes or base pair changes, is observed between the Old World species than between the New World species (Ainouche & Bayer, 1996; Weeden *et al.*, 1996; Wolko & Weeden, 1990a). However, given the sparse New World taxon sampling this could be a sampling artefact.

There has been considerable speculation about the historical biogeography of *Lupinus* including a suggestion of polyphyletic origin (Cristofolini, 1989). Recent poorly-resolved phylogenies have fuelled rather than quelled this speculation. Regions as disparate as South America and east Asia have been proposed as the geographic origin for the genus (Cristofolini, 1989; Dunn, 1971; Kurlovich & Repjev, 1994; Nowacki & Jaworski, 1978; Plitmann, 1981). One example highlights the lack of consensus. In 1999, two presentations at the same conference with partially shared authorship drew quite different conclusions about the biogeographic origin of the genus (Merino *et al.*, 1999; Wink *et al.*, 1999). Merino *et al.* (1999) suggested eastern South America as the probable centre of origin of the genus, while Wink *et al.* (1999), analysing a larger dataset, concluded that lupins evolved in the Old World.

Alongside this there has been discussion of the tribal placement of the genus. Placements in the Crotalarieae, Genisteae or Thermopsideae have been proposed (Dunn, 1984; Gross, 1986; Plitmann, 1981). However, *Lupinus* is now firmly accepted as a member of the Genisteae as supported by a series of molecular studies (Ainouche *et al.*, 2003; Crisp *et al.*, 2000; Doyle *et al.*, 2000; Dawn Edwards, unpublished data; Kass & Wink, 1997b; Wink & Mohamed, 2003; Wojciechowski *et al.*, 2004). However, within the Genisteae the sister group of *Lupinus* remains unknown. *Lupinus* has been resolved as basal to the rest of the Genisteae (Kass & Wink, 1997b), sister to a clade containing *Argyrolobium* Eckl. & Zeyh. and *Retama* Raf. (Crisp *et al.*, 2000), or as sister to the remainder of the Genisteae (Fig 3.5, Chapter 3) (Wink & Mohamed, 2003). In a recent unpublished study which densely sampled the core Genistoids, *Lupinus* was nested within one clade of a polyphyletic *Argyrolobium* (Edwards, unpubl. data). Finally, it has been suggested, based on morphology, that *Lupinus* may be closely related to the South American genera *Anarthrophyllum* Benth. and the poorly known *Sellocharis* Taub. (Lewis *et al.*, 2005a) but this is not supported by the few molecular studies that include sequences of *Anarthrophyllum* (Edwards, unpubl. data; Wojciechowski *et al.*, 2004). The poorly-known monotypic endemic southern Brazilian genus *Sellocharis* has not been sampled in any molecular studies to date. This is part of a wider uncertainty surrounding relationships within the Genisteae and has been reflected in the changing generic delimitation and overall classification of the tribe (Ainouche *et al.*, 2003; Bisby, 1981; Polhill, 1976, 1994).

3.2 Age of *Lupinus*

The age of *Lupinus* has been estimated using different methods. Kass and Wink (1997a) estimated the age of the *Lupinus* stem node as 12-14 Myr based on the genetic diversity observed in ITS sequences and calibrated with a base substitution rate derived from the split between the subfamilies Caesalpinioideae and Papilionoideae at 60 Myr. A second age estimate of 16.00 ± 1.87 (range 11.96 - 21.16) Myr for the *Lupinus* stem node comes from Lavin *et al.'s* (2005) penalised likelihood evolutionary rates analysis but was not presented in

the paper. This analysis involved a large-scale matK and rbcL legume phylogeny which was calibrated with 12 well-studied and carefully selected fossils (Lavin *et al.*, 2005). A third study using non-parametric rate smoothing (NPRS) of ITS data for the core Genistoids with a single fossil calibration point estimated the *Lupinus* crown node to be 31.8 – 32.1 Myr (Edwards, unpubl. data). It is unclear why there is such a large difference (ranging from 12 – 32 Myr) between the age estimates using ITS and matK but three possible factors can be identified. First, the rate of ITS sequence evolution appears to be extremely slow in the core Genistoids as a whole which could produce older estimates (Edwards, pers. comm.). Second, the core genistoid study is much more densely sampled than either of the other analyses, a factor which would also be expected to produce older dates. Also the NPRS algorithm implemented in the core genistoid study is well known to introduce bias towards older date estimates (Linder *et al.*, 2005; Sanderson *et al.*, 2004). However, all of these age estimates are inconsistent with a vicariant explanation for the distribution of *Lupinus* since Laurasia and Gondwana split ca. 150 Ma and Gondwana broke up ca. 90 Ma (Lavin *et al.*, 2004). However, recent diversifications rather than initial diversifications are thought to account for a large majority of extant diversity particularly in Leguminosae (Calvillo-Canadell & Cevallos-Ferriz, 2005; Lavin *et al.*, 2001b) suggesting that the age of at least part of *Lupinus* could be far younger than those dates estimated for the genus.

3.3 Methods

3.3.1 Taxon sampling

Appendix 5 lists the material used along with geographic, source, voucher information where available, and Genbank number.

Sequence data from three nuclear loci are used in this study. For ITS this involved sequencing 148 additional accessions to build on the 117 sequences available from Genbank and 18 unpublished sequences for Mexican accessions (Norma Oropeza unpubl. data). The GPAT1 locus was developed specifically for *Lupinus* as part of my study (see 3.4.5) and

all sequences are newly generated. Helène Citerne developed locus-specific LEGCYC1A primers during a study of the evolution of *CYCLOIDEA*-like genes in legumes (Citerne, 2006; Citerne *et al.*, 2003). *CYCLOIDEA*-like genes are divided into three types *CYC1*, *CYC2* and *CYC3* (Howarth & Donoghue, 2006). Current research indicates that there may be five *CYCLOIDEA*-like genes in some legumes (two LEGCYC1 genes and three LEGCYC2 genes) (Howarth & Donoghue, 2006). *LEGCYC1A and LEGCYC1B* have been characterized and sequenced in the Genistoid legumes including *Lupinus* (Citerne *et al.*, 2003; Ree *et al.*, 2004). Twenty-one LEGCYC1A sequences were generated by Ree *et al.* (2004). Colin Hughes used these and additional primers to generate a further 126 accessions LEGCYC1A sequences for *Lupinus.* Sequences from the LEGCYC1A dataset are incorporated into the 3-locus analysis presented here.

New taxon sampling concentrated on southeastern USA and South America, which were previously under-sampled in previous studies, to ensure that all major geographic regions are well-represented. Existing phylogenetic trees (Ainouche & Bayer, 1999; Kass & Wink, 1997a) were used as a guide to ensure inclusion of all previously recognised lineages. To test the monophyly of the Andean species, 36 species were sampled spanning the range of life forms (Fig. 3.21a–j), geography, and altitudes. Sampling was greatest for the ITS dataset which comprises 283 sequences representing 156 species i.e. around 54% of the genus (see Table 3.1 for breakdown according to geographical origin). The GPAT1 dataset sampled 93 species due to limited availability of North American material and a conscious choice to limit Andean sampling when little sequence variation was found across a broad geographical and morphological span. Where possible more than one accession and where necessary more than one clone was sequenced for each species. The smallest dataset (combined 3-locus matrix) includes taxa from all the major lineages. Two 3-locus matrices were analysed. One with no missing data and the other with seven taxa with missing data (see Appendix 5) so as to include all Old World groups and representatives from series *Platycarpos* (See Chapter 7). This is justified as it has been shown that small amounts of missing data are not detrimental to the accuracy of phylogenetic reconstruction and whilst

this allows greater taxon sampling, it may also increase accuracy (Weins, 1998, 2006; Weins & Reeder, 1995).

Table 3.1 Taxon sampling. Numbers of sequences and species (in parenthesis) from each geographic region in each data matrix.

	ITS (Ainouche & Bayer, 1999)	**ITS**	**GPAT1**	**GPAT (both copies aligned)**	**LEGCYC1A**	**3-loci (no missing data)**	**3-loci (missing data)**
Total no.	49 (46)	283 (156)	177 (98)	212 (99)	158 (107)	108 (83)	115 (87)
Out Groups	5 (5)	22 (22)	7 (5)	7 (5)	8 (8)	3 (3)	3 (3)
Old World	12 (10)	32 (13)	6 (5)	12 (6)	11 (8)	3 (3)	7 (7)
North America	28 (27)	125 (73)	55 (32)	68 (32)	52 (31)	29 (22)	31 (24)
South America	4 (4)	104 (59)	109 (57)	125 (57)	88 (60)	73 (55)	74 (55)
Total no. *Lupinus*	44 (41)	261 (134)	170 (93)	205 (94)	150 (99)	105 (80)	112 (84)

3.3.2 Outgroup selection

Outgroup comparison are the most commonly utilised method to test monophyly, identify sister taxa and root phylogenetic trees (Nixon & Carpenter, 1993). Outgroups were chosen on the basis of previous studies (Kass & Wink, 1997a; Pardo *et al.*, 2004). All outgroups in the LEGCYC1A and GPAT1 datasets were from the Genisteae. The ITS dataset additionally contained outgroups from Crotalarieae and Thermopsidae. Ideally gene trees for LEGCYC1A and GPAT1 would have been rooted with a sequence from a taxon in which the genes have not duplicated. However, since both duplications are likely to be prior to the divergence of the Genisteae alignment of such a sequence would be difficult. Technical difficulties prevented uniform outgroup sampling in LEGCYC1A and GPAT1 such that only three outgroups were common to all three loci.

3.3.3 DNA Extraction

DNA extractions were carried out following a modified CTAB method (Doyle & Doyle, 1987) or the Quiagen DNeasy Plant Mini Kit (see protocols 1 - 8, Appendix 6).

3.3.4 Amplification of ITS

The entire ITS region was amplified using the primer pairs ITS4 & ITS5 or ITS4 & 17SE

(see protocol 9, Appendix 6) (Sun *et al.*, 1994; White *et al.*, 1990). The latter combination was easier to amplify and sequence in some of the lowland South American taxa. Products were cleaned using a range of techniques (see protocols 14 – 16, Appendix 6).

3.3.5 Development and amplification of GPAT1

The enzyme glycerol-3-phosphate acyltransferase (GPAT) catalyses the first step in phospholipid and triglyceride synthesis (Ishizaki *et al.*, 1988; Nishida *et al.*, 1993). GPAT is thought to play a role in chilling resistance in plants by incorporation of selective fatty acids into chloroplast lipids and subsequently chloroplast membranes (Bhella & Mackenzie, 1994; Ishizaki *et al.*, 1988). Different isozymes have been isolated from mitochondria, chloroplast and cytoplasm (Murata & Tasaka, 1997). Most research has focused on the chloroplast form which is thought to be encoded in the nucleus. This form has been found to be single copy in *Arabidopsis thaliana* L. (Nishida *et al.*, 1993), *Pisum sativum* L. (Weber *et al.*, 1991), *Carthamus tinctorius* L. (Bhella & Mackenzie, 1994), *Spinacia oleracea* L. (Wolter, unpublished) and *Cucurbita moschata* Duchesne (Ishizaki *et al.*, 1988).

In *Arabidopsis* nDNA GPAT has 13 exons (Fig. 3.2) (Nishida *et al.*, 1993; Tank & Sang, 2001). I became interested in GPAT when Tank and Sang (2001) successfully utilised sequence data from a 2.3-2.6 kb fragment, including a ca. 2 kb intron, of GPAT to build a highly resolved and robustly supported species-level tree for *Paeonia*. I started to work with this gene because the degenerate primers designed by Tank and Sang (2001), which cover the large intron in *Paeonia*, amplified DNA in some *Lupinus* accessions. Furthermore, this locus has been found to be single-copy in a range of angiosperm families including the Leguminosae.

The locus referred to as GPAT1 in this study is a ca. 500 bp fragment (covering the 5' end of intron 5) of one GPAT copy type (Fig. 3.2). The procedures used to develop this region are summarised below.

Degenerate general primers (GAF1 and GAR1) (Tank & Sang, 2001) were initially used to amplify and partially sequence a 2.5 kb region from exons four to nine (Fig. 3.2) from *Lupinus mutabilis* (CEH2326 & CEH1999) and *Lupinus sp*. (CEH2004). Intron and exon boundaries were identified by comparison with *Pisum, Phaseolus* and *Vicia* mRNA GPAT sequences (GenBank numbers X59041, X79722 and AF090734 respectively). Exons five, six and seven were found to be easily alignable across the Papilionoideae allowing six lupin specific primers, GPAT_5F, GPAT_5R, GPAT_6F and GPAT_6R, GPAT_G7F, GPAT_G7R, to be designed in these exons (Fig. 3.2). Appendices 4-5 list details of the protocols and primers used. Copy specific amplifications (see below) were carried out using Amersham Biosciences pure Taq Ready-To-Go PCR Beads (Number 27-9558-01) to obtain sufficient PCR product.

In order to assess amplification efficiency, intron length and sequence variability, different primer combinations and PCR conditions were tested. A ca. one kb region including a large 820 bp intron (intron 5) between exons five and six was chosen for further development as it showed high variability and could be easily amplified using the primers GPAT_5F and GPAT_6R. Dirty sequence traces with double peaks indicative of multiple products (Scherson *et al.*, 2005) were observed in some sequence traces. These products were cloned and a tree-based approach (Bailey *et al.*, 2003) to identify paralogous copy types revealed two distinct sequence types which segregated into two well-supported clades regardless of which outgroups were used to root the tree. The size of the intron is conserved between the two copies and both copies are equally variable. The more frequently obtained copy type was called GPAT1 and the rarer type GPAT2. Copy-1 specific primers, GPAT1_I5R4 and GPAT1_I5R6 were designed for use with the GPAT_5F primer to amplify the 5' end of intron five (Fig. 3.2). As few products produced clean sequences despite the use of copy-specific primers most sequences were obtained by cloning.

Outgroups were amplified using the GAF1 and GPAT_6R primers. Products were 3-4 kb in size, larger than the corresponding region in *Lupinus*. In some taxa (*Cytisus* CEH 2185,

Spartium junceum CEH 2190) double bands were clearly visible (Fig. 3.3). Partial sequences were obtained through direct sequencing for some taxa allowing outgroup-specific primers to be designed (GPATOG_I5F1 and GPATOG_I5R2, Fig. 3.2). Although some product was obtained with these primers, poor amplification meant that sequence was not obtained for all taxa. Multiple attempts to isolate and clone two copies of GPAT from outgroups were unsuccessful probably at least in part due to the large size of the PCR product.

3.3.6 Development of SCAR-based DNA sequence regions

The method outlined by Bailey *et al.* (2004) (Fig. 3.4) was followed but modified to utilise inter simple sequence repeat (ISSR) primers instead of RAPD primers. The aim was to investigate the potential to develop *Lupinus*-specific primers for a novel SCAR-based hyper-variable sequence locus and consider screening strategies to deal with the large number of potential loci which this approach can produce. Although molecular data including genome sequencing projects and EST databases are available for several model/crop legume species (*Glycine max, Lotus japonicus* (Regel) K. Larsen*, Medicago truncatula* Gaertn.*, Phaseolus vulgaris* L.) these taxa are sufficiently distant from *Lupinus* (Fig. 1.1) (Doyle & Luckow, 2003; VandenBosch & Stacey, 2003; Young *et al.*, 2003) to complicate direct application of these data in locating molecular loci for species-level phylogeny reconstruction.

ISSR primers were used instead of RAPD primers as they are reported to be more repeatable (Culley & Wolfe, 2001) and attempts to amplify RAPDs from *Lupinus* DNA were unsuccessful. Twenty-four primers were utilised, including both new and modified existing ISSR primers, covering a range of lengths, repeat numbers, repeat sizes, different repeat sequences, different length anchors and both five and three prime anchors (see Appendix 10). Amplifications were unsuccessful with single primers but products were obtained using primer pairs (see protocol 13, Appendix 6 and Appendix 11).

Nine species representing the phylogenetic diversity of *Lupinus* (distantly and putatively closely related species) and the morphological range of Andean lupins were included in the preliminary sampling. These were *L. albus, L. microphyllus, L. mutabilis, L. piurensis* C.P. Smith*, L. ramossisimus, L. sierrae-blancae* Wooron & Standl.*, L. sp*7 CEH2107*, L. texensis* Hook. and *L. weberbaueri* Ulbr.. Taking into consideration the studies published at the time (Bailey *et al.*, 2004; McLenachan *et al.*, 2000) and my specific objectives the following criteria for screening potential loci were implemented:

- Sequence length 500 – 1000 bp,
- equal size PCR products across species,
- single bands and clean sequence with specific primers,
- repeatable amplifications,
- sequences which are easy to amplify and sequence i.e. not containing microsatellites,
- levels of sequence variation (numbers of parsimony informative characters) greater than GPAT1 and LEGCYC1A.

Bands between 500-1000 bp amplifying in more than one species were gel cleaned and cloned (see protocols 16 and 19-20, Appendix 6). Sequence variation was compared with the ITS, GPAT1 and LEGCYC1A loci. Fragment-specific primers were designed from the aligned cloned sequence data.

3.3.7 DNA sequencing

Sequencing reactions were carried out following protocol 17 (Appendix 6). Sequencing products were cleaned and dried using ethanol precipitation (protocol 18, Appendix 6). Sequencing was carried out by the sequencing facility of the Department of Biochemistry, University of Oxford (http://polaris.bioch.ox.ac.uk/dnaseq/index.cfm?page=home.index). Products were sequenced using amplification primers and both forward and reverse reactions were carried out for all products.

3.3.8 Sequence verification and alignment

Sequence traces were viewed, edited and contigs assembled using Sequencher (Gene Codes Corporation, 1991-1998) or BioEdit (Hall, 1999). Alignments were initially carried out using ClustalX (Thompson *et al.*, 1997) and manually adjusted in BioEdit (Hall, 1999) or WinClada (Nixon, 2002). Alignment was straightforward for both ITS, GPAT1 and GPAT1/ GPAT2 together. Gaps were coded following the simple gap-coding method of Simmons and Ochoterna (2000) either manually or using SeqState (Müller, 2005) and included in the data matrices for parsimony analysis. Sequences generated for this study have been deposited in GenBank under the accession numbers given in Appendix 5. Alignments are presented in the electronic appendices on CD-ROM accompanying this thesis.

3.3.9 Data set compatibility and combination

Datasets were analysed individually and all three were concatenated for a simultaneous analysis of the combined data (Bull *et al.*, 1993; Huelsenbeck *et al.*, 1996; Nixon & Carpenter, 1996; Siddall, 1997). Congruence between loci was visually assessed from consensus tree topologies. The Incongruence Length Difference (ILD) test (Farris *et al.*, 1995; Mickevich & Farris, 1981) was implemented in WinClada (Nixon, 2002). Sequences from the same accession were combined, where possible, in the 3-locus data matrix (Nixon & Carpenter, 1996). Where multiple sequences for the same species were available in one dataset, through cloning or multiple accessions, the most complete sequences were chosen in preference. The sequence with the most autapomorphies was chosen for combination when sequences were equally complete. This strategy is different from normal concatenation approaches which fuse individual accessions into a single polymorphic terminal representing each species (Hughes *et al.*, 2002; Nixon & Carpenter, 1996) and was chosen because the monophyly of species has not been rigorously tested. Although this disregards information, polymorphic characters have been found to often contain less phylogenetic signal than fixed characters (Weins, 1995).

3.3.10 Parsimony analyses and support values

Parsimony analyses were carried out using Nona (Goloboff, 1999) spawned from WinClada (Nixon, 2002) using a heuristic multiple tree bisection-reconnection (TBR) and TBR search strategy. Uninformative characters were excluded prior to analysis. Preliminary analyses used 100 random addition sequences, tree bisection and reconnection (TBR), holding 1 tree per replication until 1000 trees were found. Longer analyses to provide more exhaustive tree searches using 1000 replicates and held 10 trees per replication and attempting to swap to completion were also carried out. Analyses were stopped before completion since preliminary data suggested that swapping all equally most parsimonious trees to completion would not be possible. A random survey of the equally parsimonious trees showed that rearrangements of taxa in a large polytomy in the consensus tree accounted for the majority of the different topologies. It is likely that longer analyses would continue to recover different topologies within this clade without finding a shorter tree. To test this, the parsimony rachet (Nixon, 1999), which searches a wider area of tree space, was implemented in WinClada to search for shorter trees. Two hundred iterations per ratchet were performed perturbing ca. 20% of the informative characters, constraining 10% of the nodes and holding one tree per iteration (standard guidelines as given in Nixon, 1999). Five repeat ratchets were run for each matrix. Strict consensus bootstrap analysis was also carried out in Nona using 1000 replications each with 100 random additional sequences and 100 starting trees held per replication. Trees were rooted on *Aspalathus cephalotes* Benth. (ITS), *Retama monosperma* Boiss. (LEGCYC1A), *Genista hirsuta* Vahl. (GPAT) and *Genista anglica* L. (GPAT1 and 3-locus) (Nixon & Carpenter, 1994). Alternative rootings did not affect the topology of the ingroup (Milinkovitch & Farris, 1998; Sanderson & Shaffer, 2002; Tarrio *et al.*, 2000). Strict consensus trees (referred to as Nelsen trees in WinClada) were constructed in WinClada.

3.3.11 Bayesian analyses

Mr Model Test (Nylander, 2004) was implemented to ascertain the most suitable model of

sequence evolution for each data partition. Using the Akaike Information Criterion (Posada & Buckley, 2004) GTR+G was deemed the most appropriate model for ITS and GPAT whilst GTR+G+I was optimal for LEGCYC1A. Bayesian analyses were carried out on individual ITS and GPAT1 (Colin Hughes carried out the individual LEGCYC1A analysis), the 2-locus ITS and LEGCYC1A and the combined 3-locus datasets (see Appendix 8 for Bayes Blocks) in Mr Bayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). Taxon sampling was reduced compared to the parsimony analyses to reduce computational time required (see Table 3.2). The combined ITS/LEGCYC1A dataset was reduced to one accession per species to ensure that evolutionary rates were being compared between the same taxonomic units (Sanderson, 2002). Four chains were used and heated to 0.01 for ITS, ITS/LEGCYC1A, 0.001 for GPAT1 and 0.001 for the combined 3-locus dataset in order to sample tree space more widely (Archibald *et al.*, 2003; Sanderson, 2002). Trees were sampled every 100 generations. Tracer (Rambaut & Drummond, 2003-2005) was used to graphically assess stationarity for each dataset. Three individual runs were carried out for each dataset for three million generations. Burn-in trees were discarded before the 50% majority rule consensus tree was created in Mr Bayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). Posterior probabilities are calculated automatically during Bayesian analysis (Ronquist & Huelsenbeck, 2003).

Table 3.2 Breakdown of terminal taxa (and species) sampling for each dataset utilised in the Bayesian analysis.

	ITS	GPAT1	1-locus LEGCYC1A	2-locus ITS & LEGCYC1A	3-locus Missing data
Out Groups	14 (14)	7 (5)	8 (8)	1 (1)	3 (3)
Old World	11 (11)	7 (6)	10 (7)	7 (7)	7 (7)
North America	53 (49)	53 (29)	52 (31)	30 (30)	31 (24)
South America	77 (47)	107 (55)	88 (58)	51 (50*)	74 (55)
Total no. terminal taxa (Total no. species)	155 (121)	174 (95)	158 (104)	89 (88)	115 (87)

* Two accessions of one species were accidentally retained.

3.3.12 Dating

The combined ITS and LEGCYC1A trees were used for evolutionary rates analysis. This two-locus combination provided relatively high levels of resolution while retaining wider geographic and taxon sampling than the 3-locus dataset. The log likelihood ratio test (Shimodaira & Hasegawa, 1999) implemented in Mr Bayes rejected a molecular clock for the two-locus consensus Bayesian tree (LR = 3313.72, df = 142, p=0.00). The program r8s (Sanderson, 2002) was selected a from range of alternatives (Rutschmann, 2006) to assess variation in substitution rates and incorporate this into lineage age estimation. Branching order and branch lengths from 100 Bayesian trees sampled every 10,000 generations after stationarity were analysed. Cross-validation (see r8s block in Appendix 8) and penalised likelihood analysis (see r8s block in Appendix 8) yielded an average smoothing parameter of 13 which was used in final dating analyses. Relative ages estimated with r8s were converted to absolute ages by fixing the age of the *Lupinus* stem node using mean (16.01 Myr) and maximum (21.16 Myr) dates for the most recent common ancestor of *Spartium* and *Lupinus* from the legume-wide analysis of *matK* sequences (Lavin *et al.*, 2005). Mean and standard deviation for each clade were obtained, where possible, for each tree (Age estimates for major clades recorded in Appendix 9).

3.3.13 Species diversification rates

Net species diversification rates, assuming an equal rate of random speciation logarithmic tree model, were calculated as SR = $(\ln n_1 - \ln n_0)/t$, where n_1 = number of extant species, n_0 = initial species diversity, here taken as 1, and t = time in Myr (McCune, 1997). Upper and lower standard deviations of age estimates were used in calculations of speciation rate.

3.3.14 Biogeography

The geographic distribution of *Lupinus* was partitioned into five areas: Old World, North America, Mexico, Andes and lowland South America. These regions were chosen to test

their monophyly, assess geographical patterns and implement a dispersal/vicariance analysis. North America and Mexico were considered separately to test the polarity of North-South American movements. The Andes and lowland South America were delimited as separate areas to allow investigation of the origin of Andean lupins. Two distinct South American groups have also been recognised previously (Planchuelo, 1996a; Planchuelo & Dunn, 1984). These areas were utilised in a dispersal-vicariance analysis using DIVA (Ronquist, 1996). DIVA represents an event-based approach to biogeographic inference and thus does not require an area-cladogram (Ronquist, 1996). The 3-locus (with missing data) consensus tree was pruned to make it fully resolved as required for DIVA analysis (Ronquist, 1996). Additionally the same areas were optimised onto the 3-locus phylogeny (with missing data) using WinClada (Nixon, 2002). For this the strict consensus tree was used instead of one of the equally parsimonious trees because it represents elements of uncertainty surrounding the phylogeny. Nodes at which geographic transitions take place are present in all trees and thus adequately represented by the consensus tree.

3.4 Results and discussion

3.4.1 Phylogenetic utility of ITS for *Lupinus* species-level phylogenetic reconstruction

The ITS matrix is the most densely sampled of the three matrices, comprising of 283 terminals representing 159 species. This is the most densely sampled data matrix for *Lupinus* to date with nearly two and a half times greater taxon sampling than any previous analysis. The aligned length of the ingroup was 641 bp and 133 (20%) characters were parsimony informative (see Table 3.3). ITS1 ranged in length from 233 to 238 within the ingroup (232 to 239 including outgroups) whilst ITS2 varied from 207 to 220 within the ingroup (207 to 222 including outgroups). All sequences generated for this study included the 5.8S subunit. 60 sequences from Genbank consisted of ITS1 and ITS2 only. The 5.8S region was 163bp in length and contained only nine parsimony informative characters. Indels were small reaching a maximum length of four bp and were confined to the spacers.

Overall the strict consensus ITS gene tree is poorly resolved (Fig. 3.5). *Lupinus* alone has a consensus fork index of just 0.37 (Table 3.3). Congruent with previous studies (Ainouche & Bayer, 1999; Ainouche *et al.*, 2004; Kass & Wink, 1997a) *Lupinus* is resolved with 99% bootstrap support as a monophyletic group. The base of the strict consensus ITS tree is a trichotomy between a small clade (A Fig 3.5) and two larger clades (F and H, Fig 3.5). These major clades contrast in their structure. Clade F is almost fully resolved with generally moderate to high bootstrap support (most unresolved nodes are within species) whereas clade H, containing over three times as many terminals, represents a near polytomy. The ITS tree suggests some unique relationships, the most notable are clades A, D, F and I (Fig. 3.5) but all four of these clades, although present in the strict consensus tree, receive no significant bootstrap support. The most significant difference from the 3-locus tree is the presence of clade F. This is important because clade G is not sister to clade H as see in subsequent trees. However, clade F is unsupported (16% bootstrap support).

Although showing a higher than average intrageneric ITS percentage of PICs (Table 3.3 and Fig. 2.4, Chapter 2), ITS provides insufficient PICs to produce a well-resolved and robustly supported gene tree for *Lupinus.* Although some of the major clades are resolved with high support, the lack of resolution across the gene tree more generally reduces the utility of ITS to infer species-level relationships within *Lupinus*. This is not unusual as numerous other studies (Bayer, 1996; Emshwiller & Doyle, 1998; Harris *et al.*, 2000; Hodges & Arnold, 1994; Miller *et al.*, 1999; Richardson *et al.*, 2001a; Samuel *et al.*, 1998; Schilling *et al.*, 1998; Sonnante *et al.*, 2003; Vargas *et al.*, 1998; Wink *et al.*, 1999) demonstrate low levels of ITS sequence evolution and low resolution in consensus trees. This phenomenon has often been attributed to rapid species diversification (Harris *et al.*, 2000; Klak *et al.*, 2004; Richardson *et al.*, 2001a; Richardson *et al.*, 2001b; Wojciechowski *et al.*, 1999).

Table 3.3 Data matrix and consensus tree statistics for parsimony analyses. Ingroup only values are reported first followed in parenthesis by values including outgroups. In an attempt to take multiple species accessions into account calculations only include species with infraspecific nodes excluded. Missing data refer only to the ingroup. *Retama monosperma* CEH2182 and *Spartium juceum* CEH2190 sequences of LEGCYC1A are partial, comprising only the first 611 bps. CFI = number of resolved nodes in the strict consensus tree/n-2 where n = number of terminal taxa. The Andean clade is defined as clade K in Fig. 3.6 in the 3-locus phylogenetic tree. No. nodes resolved </> 50% BS is a direct count of BS values. Direct comparison of data matrix statistics are only strictly valid if taxon sampling is identical as increased taxon sampling could potentially increase the number of PIC or alter the CFI.

	ITS	GPAT1	GPAT (both copies aligned)	LEGCYC1A	3-locus (no missing data)	3-locus (missing data)
No. terminals	261 (283)	170 (177)	212 (205)	150 (158)	105 (108)	112 (115)
No. accessions	261 (283)	126 (131)	131 (136)	150 (158)	105 (108)	110 (113)
No. species	134 (156)	93 (98)	94 (99)	99 (107)	80 (83)	84 (87)
Sequence length (bp)	613-622 (613-622)	299-512 (299-512)	I=299-512 II=127-507	954-1217 (954-1217)	n/a	n/a
Aligned Length (bp)	641 (670)	549 (558)	676 (676)	1279 (1300)	2528 (2528)	2528 (2528)
No. PIC	133 (201)	141 (171)	222 (293)	305 (332)	404 (484)	449 (537)
% PIC	21% (30%)	27% (31%)	33% (43%)	24% (36%)	16% (19%)	18% (21%)
No. informative indels	14 (25)	21 (24)	43 (44)	12 (12)	21 (27)	24 (30)
No. excluded sites	0 (0)	0 (0)	0 0	26 (26)	26 (26)	26 (26)
CFI	0.37 (0.39)	0.33 (0.34)	n/a	0.38 (0.39)	0.59 (0.57)	0.57 (0.58)
CFI excluding Andean clade taxa	0.51 (0.50)	0.51 (0.52)	n/a	0.59 (0.58)	0.93 (0.91)	0.98 (0.86)
CFI of Andean taxa	n/a	n/a	n/a	0.03	0.21	0.21
No. Nodes resolved >50%BS	36 (44)	20 (23)	n/a	25 (28)	37 39	37 39
% Nodes resolved >50%BS	0.73 (0.73)	0.66 (0.70)	n/a	0.68 (0.68)	0.80 (0.81)	0.79 (0.81)

3.4.2 Phylogenetic utility of GPAT for species-level phylogenetic reconstruction

GPAT was reported, on the basis of single-species investigations, to be single copy in five angiosperm families (Asteraceae, Brassicaceae, Chenopodiaceae, Cucurbitaceae, and Leguminosae) (Bhella & Mackenzie, 1994; Ishizaki *et al.*, 1988; Kunst *et al.*, 1988; Weber *et al.*, 1991). Tank and Sang (2001) found some evidence in the form of a divergent pseudogene copy in one species to suggest possible ancient duplication and subsequent

silencing of one GPAT locus in some species of *Paeonia*. Although it was suggested that GPAT may be single copy in legumes (Weber *et al.*, 1991) this study clearly and convincingly shows that, at least for a subset of the core Genistoids, at least two copies are present.

Two copies of GPAT have been recovered from 19 lupin accessions representing all major phylogenetic lineages indicating that all lupins have duplicate copies. Despite several attempts to clone PCR products, sequences were not recovered from the two products amplified in outgroups (Sections 3.3.5, Fig. 3.3). The strict consensus tree (Fig. 3.6) indicates that copy1 was only retrieved from *Genista anglica CEH2339* since this sequence is sister to *Lupinus* copy1 sequences whilst the remaining outgroup sequences are sister to a clade containing copy one and two sequences. Since some outgroups also show multiple copies (Fig. 3.3) the duplication event must be deeper than *Lupinus. Lupinus* and possibly some of the other Genistoid genera are ancient polyploids (Section 6.5.4) providing a possible explanation for the presence of duplicate genes. It is not unusual for multiple copies of a gene to coexist and it is widely acknowledged that gene duplication and loss is a widespread phenomenon (Lynch & Conery, 2000). As homologous sequences are required for phylogenetic reconstruction care must be taken to detect and distinguish multiple copies. Once it became clear that GPAT is multicopy in *Lupinus,* subsequent sampling was concentrated on copy1.

Nearly twice as many indels were invoked in the alignment of all the GPAT sequences compared to GPAT1 alone. The strict consensus parsimony tree places sequences into two strongly-supported clades (Fig. 3.6) even though the separation of the two sequences types is not immediately obvious from the alignment. The topologies of the GPAT1 and GPAT2 gene trees do not conflict, but this is not comprehensively tested given the sparse taxon sampling in the GPAT2 gene tree. Although based on a limited sampling, these data indicate that GPAT2 is as variable as GPAT1. Given that both copies were providing comparable numbers of PICs, GPAT1 was chosen for this study as more GPAT1 sequences were obtained in the preliminary study and it proved to be easier to amplify and clone than GPAT2.

The aligned length of the ingroup for GPAT1 sequences was 549 bp and provided 141 (27%) PICs. Several large indels are implied by the alignment but they were not problematic to align manually. In terms of percentage PICs this region is notably high compared to the majority of other nuclear introns utilised in species-level phylogeny studies (Fig. 2.4, Chapter 2) (Appendix 2, Hughes *et al.*, 2006) but as it is relatively short, it provides relatively few PIC in total. The ingroup topology remained constant whatever combination of outgroups or copy two sequences were used to root the tree. The topology of the GPAT1 strict consensus tree (Fig. 3.7) is broadly congruent with the ITS tree but is more resolved and has generally higher bootstrap support (Table 3.3 and Fig. 3.5). It is also broadly congruent with the ITS and LEGCYC1A gene trees. The consensus fork index is 0.33 overall but rises to 0.51 when the Andean clade is excluded (Table 3.3).

Resolution is seen at the base of the GPAT1 consensus tree, Clade C is resolved as sister to the rest of the genus (Fig 3.7). The next node subtends a trichotomy of two small clades and one large clade. The large clade (L Fig 3.7) is strongly supported (91% bootstrap support) and is also resolved in the LEGCYC1A and 3-locus consensus trees. Clade L is divided into two subclades. One clade is weakly supported and shows despite a basal polytomy, high levels of support and resolution. The other contrasts markedly as it is strongly supported but virtually unresolved and what resolution there is, is poorly supported.

The GPAT1 consensus tree (Fig. 3.7) conflicts with the LEGCYC1A tree (Fig. 3.8) because it suggests that clade G is deeply nested within clade L. However, this incongruence is unsupported in the GPAT1 tree and all three nodes supporting this relationship are each supported by only a single homoplasious character. Thus, this incongruence is not of great concern because if these weak nodes are ignored clade L would form a polytomy which is not in conflict with the other datasets.

There is however one notable case of strongly-supported incongruence between the GPAT1 and both the individual ITS and LEGCYC1A consensus trees. This is the position of *L.*

parvifolius. In the GPAT1 tree (Fig. 3.7) this taxon is placed in an unresolved clade with *L. linearis* whilst ITS and LEGCYC1A (Figs. 3.5 and 3.8) suggest a sister relationship to *L. guaraniticus* (Hassl. C.P.Sm.) (this is also seen in the 3-locus consensus trees Figs. 3.9 and 3.10). Unifoliolate leaf morphology also supports a close relationship between *L. parvifolius, L. guaraniticus* and other South American unifoliolate species (See Chapters 5 and 6). *Lupinus linearis* has digitate leaves. The sequence is not highly divergent thus is not likely to represent a pseudogene, and hybridisation among these Brazilian species has not been recorded. Thus, a more likely explanation for this incongruence might be human error in the laboratory but so far this has not been checked.

The GPAT1 locus has clear potential as a nuclear locus for species-level phylogeny reconstruction provided the dynamics of gene duplication are taken into account. GPAT1 performed better in terms of higher percentage of PICs, greater resolution and support than ITS and similarly to LEGCYC1A but despite this, GPAT1 was insufficient alone to resolve relationships to an informative level within *Lupinus*. It has also been used to reconstruct a robust phylogeny of *Paeonia* (Tank & Sang, 2001). The multi-copy nature of GPAT needs to be borne in mind when this locus is being considered for other plant groups. It should also be noted that only a very small region (3/4 of intron 5) of the total GPAT1 locus was used in this analysis and developing primers to access a wider part of the gene would be feasible and could be productive.

3.4.3 Phylogenetic utility of LEGCYC1A for species-level phylogenetic reconstruction

LEGCYC1A was the longest of the individual loci with an aligned length of 1279 bp, providing 305 PICs within the ingroup (24%). Consistent with Ree *et al.*'s (2004) smaller dataset, LEGCYC1A is more informative than ITS. The strict consensus LEGCYC1A tree (Fig. 3.8) is predominantly congruent with the individual ITS, GPAT1 and combined 3-locus trees.

The basal node subtends two clades, clade L (68% bootstrap support) seen in the 3-locus trees and a clade containing everything else (29% bootstrap support). Congruent with the 3-locus consensus trees, clades G and H are resolved as sister clades but unlike the ITS dataset do not differ so dramatically in levels of resolution. Particularly poor levels of resolution are restricted to a nested clade K (Fig. 3.8, also seen in the 3-locus trees). The consensus fork index for the ingroup is 0.38. A surprising and possibly anomalous placement of the North American species *L. nanus* Dougl Ex. Benth. with the Andean taxa is seen. The source of this accession ('Pixie Delight' from Thompson & Morgan) was a commercial seed company presenting the possibility that this may be a hybrid or incorrectly identified and despite efforts to verify this, it has not been possible to check this finding using other accession of *L. nanus*. Due to low sequence variability the ITS dataset is not informative about this discrepancy.

LEGCYC1A is among the fastest nuclear loci developed for species-level phylogeny (see Chapter 2). It provides a relatively well-resolved and robust phylogeny for *Lupinus* as a whole but lacks resolution not only in the large clade (K, Fig. 3.8) but also at three other nodes which form polytomies of seven or more branches. Similar to GPAT it required considerable development to identify the individual LEGCYC loci (Citerne *et al.*, 2003) and design copy-specific primers (Citerne, 2006).

3.4.4 Phylogenetic utility of multiple loci and combined analysis for species-level phylogenetic reconstruction

This study highlights the fundamental importance of using multiple loci for phylogeny reconstruction. Each of the three loci differs in percentage of parsimony informative characters, levels of resolution, support and in minor ways topology. Acceptance of any one of the gene trees alone to represent the phylogeny of *Lupinus* would convey less informative and slightly different hypotheses. Despite the fact that the two low-copy nuclear loci performed better than ITS on all measures, they were still insufficient alone to produce

anywhere near 100% resolution. A combined analysis was carried out to investigate whether a stronger or different phylogenetic signal could be obtained.

Visual inspection showed individual trees from the three loci to be broadly congruent. Most of the limited incongruence that was apparent was restricted to the basal branches and relationships of the Old World taxa and all incongruent nodes were unsupported except the placement of *L. parvifolius*. The ILD test found no significant incongruence between any combinations of datasets (P = 0.00 for all tests). The combined matrix has an aligned length of 2528 bp and 449 (18%) PICs in the ingroup (110 terminal taxa representing 84 species) alignment. The combined 3-locus dataset contains nearly four times as many PICs as the ITS matrix alone.

The small amount of missing data in one of the 3-locus datasets made minimal difference to the topology of the consensus tree (compare Figs. 3.9 - 3.10). The consensus tree built from the dataset with no missing data resolves eight additional nodes, but these are all poorly supported, and do not conflict with the consensus built from the dataset with missing data and loses one node. Overall the topology is retained. The consensus built from the matrix with limited missing data (Fig. 3.10) represents the best estimate of *Lupinus* phylogeny as it represents all previously recognised lineages, samples lupins from across their range and additionally represents the diversity in morphology observed within the genus.

Although the parsimony rachet did not find shorter trees it appears to be searching a different part of tree space since *L. angustifolius* L. was resolved as the basal branch in all rachet trees (not shown). Otherwise the topology has congruent with the initial tree search. The incongruence may stem from the reduced character sampling implemented in the rachet and lack of non-homoplasious characters supporting the placement of *L. angustifolius* as sister to *L. hispanicus* Boiss. & Reut. and *L. luteus* in all other analyses (Figs. 3.5-3.10 and 3.12-3.16).

The combined 3-locus (ITS, GPAT1 and LEGCYC1A) analysis provides significantly greater resolution and support compared to the individual gene trees (Table 3.3 and Fig 3.11). Percentage resolution (CFI – 0.57) in the 3-locus (with missing data) consensus tree is much higher than any of the loci alone (ITS=0.37, GPAT1=0.33 LEGCYC1A=0.38). This is even more apparent when the large unresolved Andean clade is excluded, in which case over 80% of the possible number of nodes in the 3-locus tree (with missing data) are resolved (Fig 3.11b). This phylogeny reconstruction tree represents a considerable advance over previously published estimates (Fig. 3.1). This demonstrates that the combined 3-locus dataset is performing well across the genus as a whole except within the Andean clade suggesting that the evolution of this clade may be very different from the rest of genus. Species-level phylogenies with a consensus fork index of 90-100% are rare, represented by only ca. 20% of the total number of published trees of plant species (Appendix 2, Hughes *et al.*, 2006).

Bootstrap support values are also higher in the 3-locus (with missing data) consensus tree relative to the individual gene trees (Fig 3.11). The ingroup contains six (seven and a half if the Andean clade is excluded) times as many nodes supported by bootstrap values above 90% than ITS, the dataset with the lowest support. Table 3.4 shows that increased support values can also be seen at specific nodes.

Table 3.4

Node subtending	**ITS**	**GPAT1**	**LEGCYC1A**	**3-locus(with missing)**
Clade G	84%	87%	92%	100%
Clade H	51%	-	63%	97%
(*L. crotalarioides*, *L. subsessilis*, *L. velutinus*)	63%	62%	-	96%

The relationships of clade G differ between the three individual analyses. This appears to be indicative of a wider problem of resolving resolution in the basal area of the tree and with clade H. The strong support for clades L and H obtained in the 3-locus consensus trees shows that the apparent conflict between the individual datasets is due to lack of strong

phylogenetic signal in this area of the tree. Even with more PICs in the combined analysis, several basal branches remain unsupported. This short-branch length problem is likely to demand significantly greater sequence data to be resolved with high support (Rokas *et al.*, 2003; Wortley *et al.*, 2005).

3.4.5 Utility of Bayesian analyses

All analyses were judged to have reached stationarity by one million generations thus the first 10,000 trees were discarded. Trees resulting from the Bayesian analyses (Figs. 3.12-3.16) are highly congruent with the parsimony trees (Figs. 3.5-3.10) but show increased resolution within the Andean clade (Fig. 3.18). Overall resolution was not significantly increased (compare Figs. 3.11a and 3.17). Posterior probability values were greater than 0.50 for all resolved nodes. However, it is widely acknowledged that posterior probabilities are not directly comparable to bootstrap values and are inflated in comparison (Alfaro & Zoller, 2003; Douady *et al.*, 2003; Simmons *et al.*, 2004). Furthermore, posterior probabilities are more sensitive to short branch lengths but at the same time are more prone to type I errors when branches are short (Alfaro & Zoller, 2003). The percentage of nodes supported by posterior probabilities greater than 0.90 is compared across datasets (Table 3.5). The CFI of the Andean clade increases from 0.14 in the ITS gene tree to 0.46 in the 3-locus tree. This is greater than the highest CFI of 0.21 in the 3-locus tree using parsimony analyses (Tables 3.3 and 3.5).

This raises questions as to why there is additional resolution provided by Bayesian over parsimony analysis, especially within the Andean clade, how much confidence one can have in it and what inferences can be made based on the Bayesian resolution? The resolution could arise either from the model utilised or it could be a function of the tree search strategy. It has been suggested that Bayesian analysis resolves hard or near-hard polytomies in a random manner (Lewis *et al.*, 2005b). Alternatively, it is possible that the Bayesian analysis has searched less tree space than the parsimony. Heating the MCMC chains and swapping

between them should avoid becoming stuck in one area of tree space but does not preclude it. Software such as the Tree Set Visualisation module Mesquite (http://comet.lehman.cuny.edu/treeviz/) would be one way to investigate this (Hillis *et al.*, 2005). It is also possible that the burn in period is longer than revealed with Tracer (personal communication Julie Hawkins) which could mean that suboptimal trees were included in the construction of the Bayesian consensus trees. These possibilities merit further investigation but were beyond the scope of the current analyses presented here.

The two sister Andean clades resolved in the 3-locus analysis are not obviously congruent with any other lines of evidence i.e. geography, altitude and life history traits. Furthermore, in some cases multiple accessions of species are not resolved as most closely related. This suggests that this additional resolution must be interpreted with caution. However, the topologies do present some interesting and novel hypotheses of relationships of *L. mutabilis* and its allies (discussed in Chapter 4).

Table 3.5 Data matrix and tree statistics for Bayesian analyses. Ingroup results are given first followed by combined ingroup and outgroup results in parenthesis. No. nodes resolved </> 0.5 and 0.9 PP is a direct count of PP values. In an attempt to take multiple species accessions into account, calculations only include species; infraspecific nodes were excluded. While these statistics provide broad useful comparisons, they are not directly comparable as the different analyses include different numbers of terminals.

	ITS	GPAT1	LEGCYC1A	2-locus ITS & LEGCYC1A	3-loci (missing data)
No. terminals	141 (155)	167 (174)	150 (158)	87 (88)	112 (115)
No. accessions	141 (155)	131 (133)	150 (158)	87 (88)	112 (115)
No. species	14 (107)	90 (95)	96 (104)	88 (89)	84 (87)
Aligned length	641 (670)	549 (558)	1279 (1300)	1934 (1934)	2528 (2528)
No. excluded	0 (0)	0 (0)	0 (0)	26 (26)	26 (26)
CFI	0.36 (0.40)	0.44 (0.44)	0.54 (0.54)	0.63 (0.64)	0.59 (0.61)
CFI excluding Andean clade taxa	0.49 (0.53)	0.68 (0.65)	0.77 (0.74)	0.88 (0.89)	0.96 (0.94)
CFI of Andean taxa	0.14	0.22	0.21	0.29	0.46
No. Nodes resolved >0.5 PP	41 (51)	38 (40)	55 (59)	52 (53)	50 (52)
No. Nodes resolved >0.9 PP	30 (37)	18 (18)	38 (40)	35 (36)	43 (44)

% Nodes resolved >0.9 PP	73 (73)	47 (45)	69 (68)	67 (68)	0.86 (0.85)

The tree topologies are such that only very tentative conclusions can be drawn about the basal Andean taxon and the morphology of the putative first Andean coloniser. Perennial woody blue-flowered shrubs are found to be basal to both clades within the 3-locus tree (Fig. 3.16). A widespread blue-flowered woody alliance of species is distributed across the length and breadth of the Andes and also occurs throughout Mexico and Central America. Thus it is plausible that the lupin coloniser of the Andes was a woody blue-flowered shrub.

3.4.6 Phylogeny and biogeography of *Lupinus*

The *Lupinus* phylogeny reveals a high level of geographical phylogenetic structure (Figs. 3.19-3.20). For example, South American species occur in two clades, one distributed in mainly lowland eastern South America, the other confined to the Andes (Fig. 3.21) Optimisation of geography onto the 3-locus consensus tree is unambiguous (Fig. 3.19). Despite some uncertainty surrounding the order of the basal branches of *Lupinus,* inferences about the biogeographic origin of the genus can be made. The Old World is reconstructed as the ancestral area in both DIVA and optimisation analyses (Figs. 3.19 – 3.20). This is consistent with the distribution of the majority of the Genisteae in the Mediterranean region and supports the hypothesis of Wolko and Weeden (1990a; 1990b). However, this depends on distribution of the sister group of *Lupinus*, and this remains uncertain. The recent suggestion that *Sellocharis* and *Anarthrophyllum,* from southern Brazil and Argentina/Chile, respectively, may be sister group to *Lupinus* (Lewis et al., 2005a) could clearly change the ancestral area of *Lupinus*. If the sistergroups are coded as lowland South America the ancestral area becomes ambiguous under unambiguous coding but remains as the Old World under both ACCTRAN and DELTRAN optimisation. DIVA invokes seven dispersals to explain the current distribution (Fig. 3.20). This can be taken as strong evidence of dispersal given that DIVA is biased towards reconstructing vicariant events (Cook & Crisp, 2005). Dispersal rather than vicariance is also suggested by the age of *Lupinus* (11.96 - 21.16 Mya) which is much younger than the estimated dates for the split

of Gondwana from Laurasia ca. 150 Mya and the break-up of Gondwana ca. 90 Mya. Furthermore, placement of the main Andean clade deeply within a large western North American paraphyletic grade also implies dispersal following the progression rule (Morrone & Crisci, 1995). Despite short branch lengths and lack of support for the basal branching order within *Lupinus,* the phylogeny implies two independent dispersals (clades C and L, Figs. 3.19 – 3.20) of *Lupinus* to the New World from the Old World, with starkly different outcomes. One New World lineage (clade C), comprises ca. four species restricted to south-eastern U.S.A. while the other (clade L) includes ca. 275 species, i.e. the vast majority of species diversity in two major New World lineages (clades G and H, Fig. 3.19), distributed from Alaska to Argentina. The most phylogenetically diverse region for *Lupinus* is south-eastern North America where species from all three New World lineages are found.

The two robustly supported (100% and 98% bootstrap support) major New World clades G and H (Fig. 3.19 and 3.21) occupy predominately allopatric distributions, one in the eastern New World and the other in the western New World and both include North and South American components. These groups are also congruent with morphology and cytology (Chapter 6). The characteristics of these two clades are very different. Clade G represents ca. 32 species the majority of which are distributed in lowland South America. Relationships between the clade G species are generally well-resolved and robustly supported indicative of strong genetic divergence. Since the North American clade is sister to a South American clade directionality of transition between the continents cannot be ascertained although the ancestral area of G and H together is unambiguously North American (Figs. 3.19-3.20). In contrast, clade H represents ca. 250 species including those from both the major New World centres of species diversity in the Rocky Mountains and the Andes. Although relationships are well resolved in part of this clade a large unresolved polytomy with short branch lengths indicative of a rapid diversification is also seen. All Mexican taxa sampled are included in clade J. ITS (albeit still with limited taxon sampling) suggests that the Mexican lupins form a monophyletic group (clade I, Fig. 3.5) although without bootstrap support. Whilst this is not seen in the combined tree they do form a grade

between the North American and Andean species suggesting polarity of movement in a southerly direction. Additional sampling from Central America and the northern Andes would be desirable to test the monophyly of the Andean clade K.

The Andean species do not form a monophyletic group but instead are placed in both of the major New World clades. Three independent movements of lupins into the Andes are implied (clades M, N and K, Fig. 3.19). Four species, represented here by *L. bandelierae* C.P.Sm. (M, Fig. 3.19), are nested in clade G. These species are distributed in the south-central Andes in northern Argentina and southern Bolivia (Fig. 7.3, Chapter 7). One species, *L. microcarpus* Sims. (N, Fig. 3.19), which is not strictly Andean in distribution as is grows in the coastal Mediterranean climatic region of Chile as well as in the Argentinean Andean foothills has a disjunct distribution between the Mediterranean regions of California and Chile. As with many plants which have amphi-tropical distributions, this could be explained by bird migration along the coast of the Americas (Raven, 1963) although recent human introduction and naturalisation can not be excluded. This region is well-known for it's distinct Mediterranean flora and may merit recognition as a separate geographic region. The majority (ca. 81) of the Andean species are represented by a large well-supported monophyletic clade (85 % bootstrap value, 1.0 PP) that is largely unresolved in all except the Bayesian analyses. Several nested pairs or triplets of species or accessions are seen mostly representing multiple accessions of the same species. The attributes of this large polytomy are discussed in Section 3.5.7, but the lack of sequence divergence among these species is striking given the wide geographic distribution from Venezuela to Argentina and Chile and extensive morphological diversity in this clade (Fig. 3.21). Considering the levels of resolution obtained in the rest of the genus (up to 88% CFI) this large polytomy immediately suggests the evolution of this clade was more rapid than the rest of *Lupinus.*

Another notable result from the *Lupinus* phylogeny relates to the disjunct distribution of non-digitate-leaved species. These species have long been thought to form a natural group (Dunn, 1971; Monteiro & Gibbs, 1986a; Planchuelo & Dunn, 1984) and have

been classified as such in previous infrageneric classifications (Agardh, 1835; Taubert, 1894; Watson, 1882). Non-digitate-leaved species are found in both North and South America. Long distance bird dispersal was proposed (Dunn, 1971) to explain this disjunction. However, the monophyly of this group is refuted by the phylogeny presented here suggesting that non-digitate leaves evolved twice independently within the genus (Fig. 6.1, Chapter 6). It is notable also that the North American non-digitate-leaved species, with a distribution centred in Florida, form an independent lineage from the rest of the New World taxa and have closer relationships with the Old World species (a relationship in line with chromosome numbers (Section 7.2, Chapter 7).

The phylogeny of the Old World species remains uncertain. Although two major clades are strongly supported by both morphological and molecular data (clades E and B, Fig 3.9) the relationships between these and *L. albus* and *L. micranthus* differ amongst the individual analyses and the 3-locus tree and even in the combined analysis remain poorly supported. This pattern may be representative of large scale extinction and/or isolation perhaps in refugia as seen in other Mediterranean taxa (Besnard *et al.*, 2002; Zhang *et al.*, 2001). These data suggest that phylogenetic signal may have been lost with extinct taxa or through sequence saturation and clarifying these relationships will be challenging.

3.4.7 Radiation of Andean *Lupinus*

Multiple lines of evidence suggest that recent and exceedingly rapid radiation of Andean lupins provide the best explanation for the large polytomy observed in the individual and combined gene trees. These are:

- High number of species – ca. 81
- Monophyly – suggesting a single colonisation from North America
- High morphological diversity – diverse life forms from prostrate herbs, acaulescents, shrubs to small trees (Fig. 3.21a-j)
- Lack of sequence divergence – low genetic differentiation

- Short branch lengths – leading to recent age estimate (Fig. 3.15)
- Pattern of molecular evolution common to several independent loci

Many other high Andean upland genera e.g. *Draba* L., *Espeletia* Mutis ex Humb. & Bonpl., *Huperzia* Bernh., *Hypericum* L. and *Lysipomia* H.B. & K. are also large and diverse. It is suggested that *Lupinus* may be an exemplar genus illustrating the origin of much of the high Andean plant diversity (Appendix 3, Hughes & Eastwood, 2006).

The age of the Andean crown node is estimated at 1.47 ± 0.29 Myr (1.93 ± 0.35 Myr using the older calibration). This is consistent with the origin of the páramo habitat 2-4 Mya (Burnham & Graham, 1999; Gregory-Wodzicki, 2000; Simpson, 1975; Van der Hammen & Cleef, 1986) (see Appendix 9 for date estimates). This age estimate is robust to even older calibrations. Calibration of the *Lupinus* crown node at 37.87 Myr (the oldest date estimated from Edwards unpubl. ITS analysis of Genistoids – see Section 3.2) only increased the Andean age to 3.19 Myr.

Rates of Andean *Lupinus* species diversification are exceptional and higher than other plant radiations reported to date (Table 3.6). They also approach the rates reported for cichlid fish which are an order faster than other documented plant or animal radiations. These high rates may be attributed to the recency of radiation given that there is evidence for speciation to be concentrated in the early phases of radiations (Gavrilets & Vose, 2005; McCune, 1997). Comparative data (Table 3.6) are available for two Andean genera *Valeriana* L. (Bell & Donoghue, 2005) and *Gentianella* Moench (Von Hagen & Kadereit, 2001). Although the monophyly of these groups has not been confirmed due to sparse sampling they also suggest high rates of diversification supporting the idea that rapid species diversification is common to many high elevation plant Andean genera.

Table 3.6 Diversification rates for a selection of rapid plant and cichlid fish radiations

Taxa	Number of species	Age (myr)	Net speciation rate (sp myr^{-1})	Reference
Andean Lupinus	81	1.47 average 1.76 max 1.18 min	2.99 2.50 3.72	This study.
Andean Gentianella	170	3.0 max 1.6 min	1.71 3.2	(Von Hagen & Kadereit, 2001)
Páramo Valeriana	55	5.0 max 3.5 min	0.8 1.14	(Bell & Donoghue, 2005)
Ruschioideae	1563	3.8 min	1.94	(Klak et al., 2004)
Aquilegia	70	5.2	0.82	(Hodges & Arnold, 1994)
Inga	300	10	0.57	(Richardson et al., 2001a)
Hawaiian Silverswords	28	6	0.56	(Baldwin & Sanderson, 1998)
Lake Tanganyika Cichlids	85	2.4-12.0	0.37-1.85	(McCune, 1997)
Lake Malawi Cichlids	400	0.7-2.0	3-9	(McCune, 1997)
Lake Victoria	300	0.12-0.2	28-475	(McCune, 1997)

It is notable that although there is considerable morphological diversity within the Andean *Lupinus* clade there are no obvious common morphological or physiological traits that might have precipitated rapid diversification. Accelerated morphological divergence of this sort is associated with many island or island-like radiations (Seehausen, 2006) and has been attributed to opportunities provided by depauperate environments where competitors are fewer (Schluter, 2000). Indeed the great diversity of life forms displayed by Andean lupins (Fig. 3.21a-j) mirrors that seen following many plant radiations on islands (Baldwin & Sanderson, 1998; Böhle *et al.*, 1996). High rates of plant species diversification in the Andes can be attributed to four key factors. The continental **scale** of the Andean *Lupinus* radiation, spanning more than 4,000km from Venezuela to Argentina (Fig. 3.21k), far exceeds typical island radiations which occupy extremely confined geographical areas, providing unusually extensive opportunities for geographical isolation. Higher species diversification rates on larger islands have been attributed to increasing opportunity for geographic isolation with island area (Losos *et al.*, 1998). Secondly, repeated **fragmentation** of high altitude habitats due to altitudinal shifts in Andean vegetation zones prompted by Quaternary climate fluctuations are likely to have contributed to geographical isolation and

diversification in the Andes (Simpson, 1975; Van der Hammen & Cleef, 1986; Vuilleumier & Monasterio, 1986) (Hooghiemstra & Van der Hammen, 2004). Thirdly, the extremely dissected **topography** and steep tropical montane environmental gradients associated with altitude (Janzen, 1967) further enhance opportunities for geographical isolation that is evident from the present-day distributions of the majority of Andean *Lupinus* species. Finally, **resource heterogeneity** in terms of habitats and environmental conditions is very high in the mid to high elevation Andes.

Discussions surrounding whether plant radiations are faster on islands or continents have failed to reach a definitive conclusion since examples of rapid diversification from both have been documented (Baldwin & Sanderson, 1998; Böhle *et al.*, 1996; Kay *et al.*, 2005; Klak *et al.*, 2004; Richardson *et al.*, 2001a; 2001b). Also the discovery that so-called "non-adaptive" radiations can have diversifications equal to the lower boundaries of proposed adaptive radiations has added to the discussion (Kozak *et al.*, 2006). The high Andes can be viewed as an island archipelago albeit on a continental scale (Appendix 3, Hughes & Eastwood, 2006). The factors influencing lupin diversification suggest that it was driven by ecological opportunities. These are similar to those afforded by the availability of new habitats and absence of competition on newly-formed islands and island-like formations (Baldwin & Sanderson, 1998; Böhle *et al.*, 1996; Gavrilets & Vose, 2005; Klak *et al.*, 2004; McCune, 1997; Schluter, 2001; Verheyen *et al.*, 2003) and not linked to key morphological or physiological innovations which are more commonly associated with continental radiations (Kay *et al.*, 2005; Klak *et al.*, 2004). The exceptional rates of diversification found for high elevation Andean plants suggest that island-like diversification driven by ecological opportunity may indeed outpace diversification on continents (Appendix 9, Hughes & Eastwood, 2006).

It is commonly acknowledged that different loci evolve at different rates making them applicable to different levels of phylogenetic study (Pennington, 1996). However, many species-level studies involving putative rapid radiations, have found that even the most

rapidly evolving loci are insufficiently variable to produce resolution (Fishbein *et al.*, 2001; Fishbein & Soltis, 2004; Hodges & Arnold, 1994; Knox & Palmer, 1995; Malcomber, 2002; Richardson *et al.*, 2001a; Wojciechowski, 2005; Wortley *et al.*, 2005). This problem is also seen in animal phylogenetics (Albertson *et al.*, 1999; DeFilippis & Moore, 2000; Lara *et al.*, 1996; Meyer, 1993; Rokas *et al.*, 2005). Simulation studies suggest that with substantially greater lengths of DNA sequence data it will possible to resolve some of these short branch length difficulties (Wortley *et al.*, 2005) whereas other studies have turned to molecular markers (Després *et al.*, 2003; Pelser *et al.*, 2003).

3.4.8. Utility of the SCAR approach and candidate loci

The work presented here to develop novel hypervariable SCAR-based loci is very much a preliminary study to investigate the potential of this method within *Lupinus* and to define effective screening criteria for the identification of hyper-variable nuclear loci in general. In terms of these aims the study was successful. Sequences were obtained from multiple species for eight loci. However, rigorous assessment of the loci was not straightforward because of differential taxon sampling, which is an inevitable by-product of the random primer method. The eight loci all showed some variation but only one performed potentially better than ITS, GPAT1, or LEGCYC1A. The challenges associated with locating hyper-variable loci encountered here are mirrored by other studies with the same aims that utilised different methods (Scherson *et al.*, 2005; Whittall *et al.*, 2006).

Considerable technical problems were encountered with the ISSR primers. Although results tended to be repeatable, amplifications were persistently poor, producing little product despite optimisation. A further problem encountered was a predominance of PCR products smaller than 500 bp. There was no observable pattern between those primers which performed well and those which did not (Appendix 11). Bands were frequently amplified only in a few taxa. Additionally it was found through cloning that bands which had been cut out and cleaned often contained more than one product.

Sixty-seven primer combinations were attempted (Appendix 11). Seventeen of these produced at least one band between 500-1000 bp shared by more than one species. Sufficient clean product was obtained from seven of these primer combinations allowing gel purification. Two bands were retrieved from one primer combination. Cloning obtained sequence from 2-5 species for each locus. The eight loci had aligned lengths between 677 bp and 1002 bp.

Comparison of sequences was made with GenBank sequences, ITS, GPAT1 and LEGCYC1A datasets and across species amplified for each locus. All sequences obtained through cloning were BLAST searched in Genbank (Altschul *et al.*, 1997). Matches were common to small sections of genomic and mitochondrial model legume DNA of unknown function. Three sequences blasted to known loci: *Pisum* mtDNA-nad2 (12:17), *Pisum* mtDNA tRNA-ser gene and *Lotus* L. genomic DNA (BC) and *Arabidopsis* and *Solanum demissum* Lindl. putative protein sequences (R1). The numbers of variable and PICs were compared with existing datasets (Appendix 12). Although all loci had fewer variable characters than ITS, GPAT1 and LEGCYC1A, two loci (BC and E1) had a greater number of PICs and a higher percentage of PICs. Discovery of putative multiple copies was frequent. Copy number is uncertain in three of the eight loci (Appendix 12) due to the presence of one alignable but highly divergent sequence.

Design of successful lupin specific primers was straightforward due to the highly conserved nature of all loci despite being based on as few as two sequences. Specific primers (see Appendix 10) were designed for two loci (E1 and 7:22-2), chosen on the basis of high pairwise diversity. Amplifications with some primer combinations were consistent and bright across a range of lupins. Direct sequences were clean indicating good primer specificity and single PCR products.

This method can be effective in locating novel genomic loci which can easily be turned into SCARs without prior knowledge of primers. This is important given that lineage-specific

loci are often necessary and that the majority of known nDNA introns are less variable than ITS (Appendix 3, Hughes *et al.*, 2006). Primers designed from sequences of the group of interest are generally far more successful in amplifying single copies than those designed by comparison of multiple sequences across a diverse range of angiosperms or even closely related taxa, a commonly used method to obtained sequence from nDNA introns. One disadvantage of this method is the low success rate in locating suitable regions. From a total of 67 primer combinations sequence data was only obtained from eight bands and only one of these is potentially more useful than existing datasets. However, with automation and the continued reduced cost of laboratory materials and sequencing, larger scale investigations would be feasible.

This and other studies (Scherson *et al.*, 2005; Whittall *et al.*, 2006) have shown that even adhering to strict criteria does not guarantee location of ideal hyper-variable single-copy loci. In this light it is even more important that these screening criteria are strictly applied to limit investment in unproductive loci. The only criterion used in this study which it may be worth relaxing is the selection of bands of exactly equal size. It is possible that this hampers the discovery of hyper-variable loci since it has been suggested that the majority of sequence differences between closely related species can be accounted for by indels (Britten *et al.*, 2003). Clearly, homologous sequences with large indels will not be discovered by this method following this criterion.

3.5 Conclusion

Lupinus provides a further example of a group for which ITS fails to provide resolution and support at the species-level. The utility of GPAT1 and LEGCYC1A sequences to reconstruct a resolved species-level phylogeny for *Lupinus* has been demonstrated. The major issues of sparse taxon and character sampling which have limited previous *Lupinus* phylogenetic studies have been addressed. While taxon sampling has doubled from previous studies it is still clear that sampling is still deficient for North and Central America species.

The combined 3-locus analysis presented here has increased resolution to 0.59 CFI with 52% of the nodes receiving 90-100% bootstrap support. Well-supported hypotheses of relationships between the Old World taxa also remain elusive. The Andean lupins are not monophyletic but the majority belong to one major clade representing the fastest plant diversification documented to date. Despite these areas of low resolution, the phylogeny is sufficiently well-supported to facilitate investigation of character evolution within the genus (Chapter 6) and form the framework for a new infrageneric classification (Chapter 7).

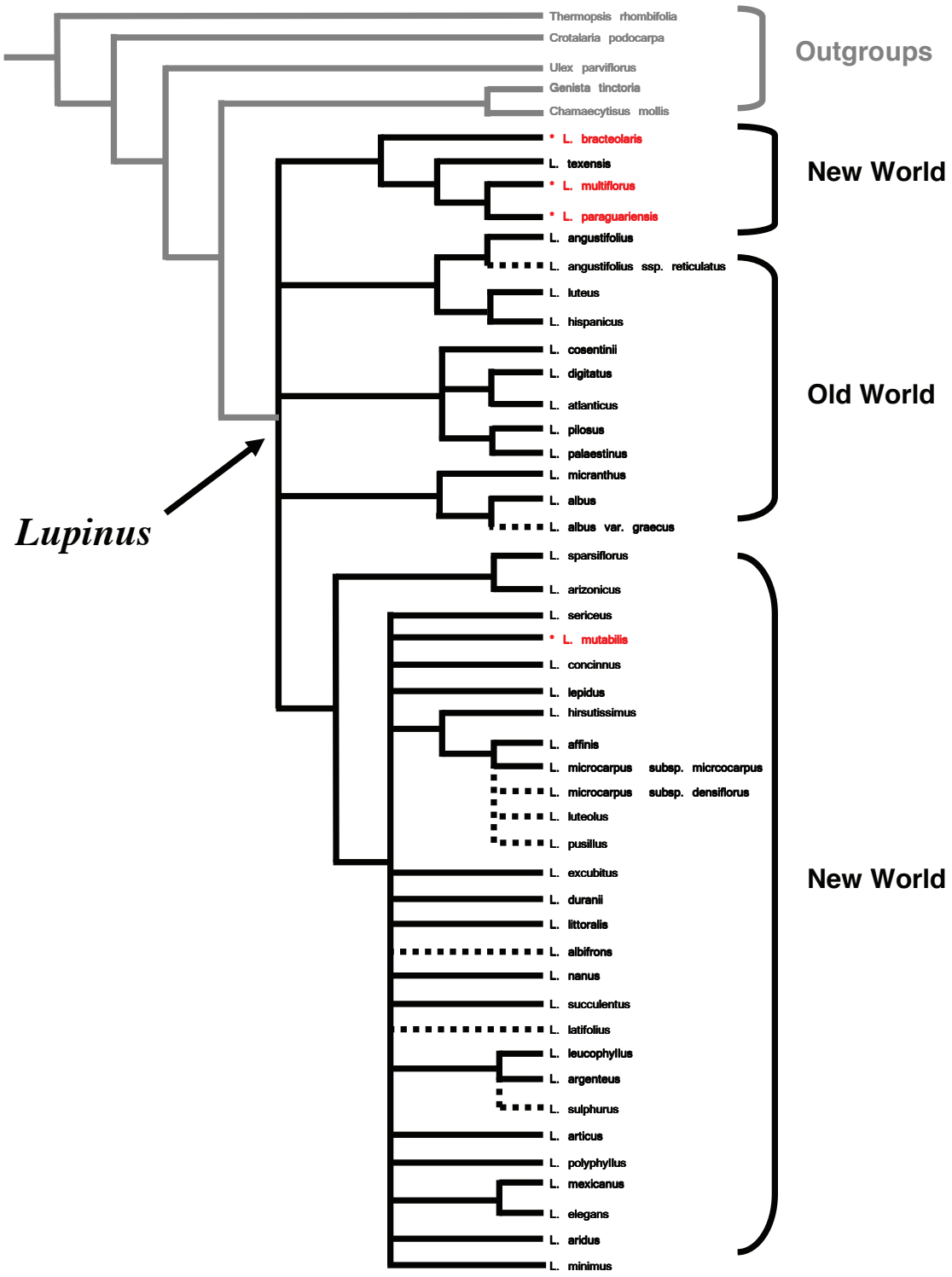


Fig. 3.1 *Lupinus* strict consensus ITS gene tree redrawn from Ainouche and Bayer (1999) N=41. The dashed branches represent taxa that were not included in the phylogenetic analyses but where later mapped on the cladogram. Only four South American accessions were included these are marked with stars and coloured red. The only Andean species sampled was the domesticate *L. mutabilis*. Support values were not presented for the consensus tree.

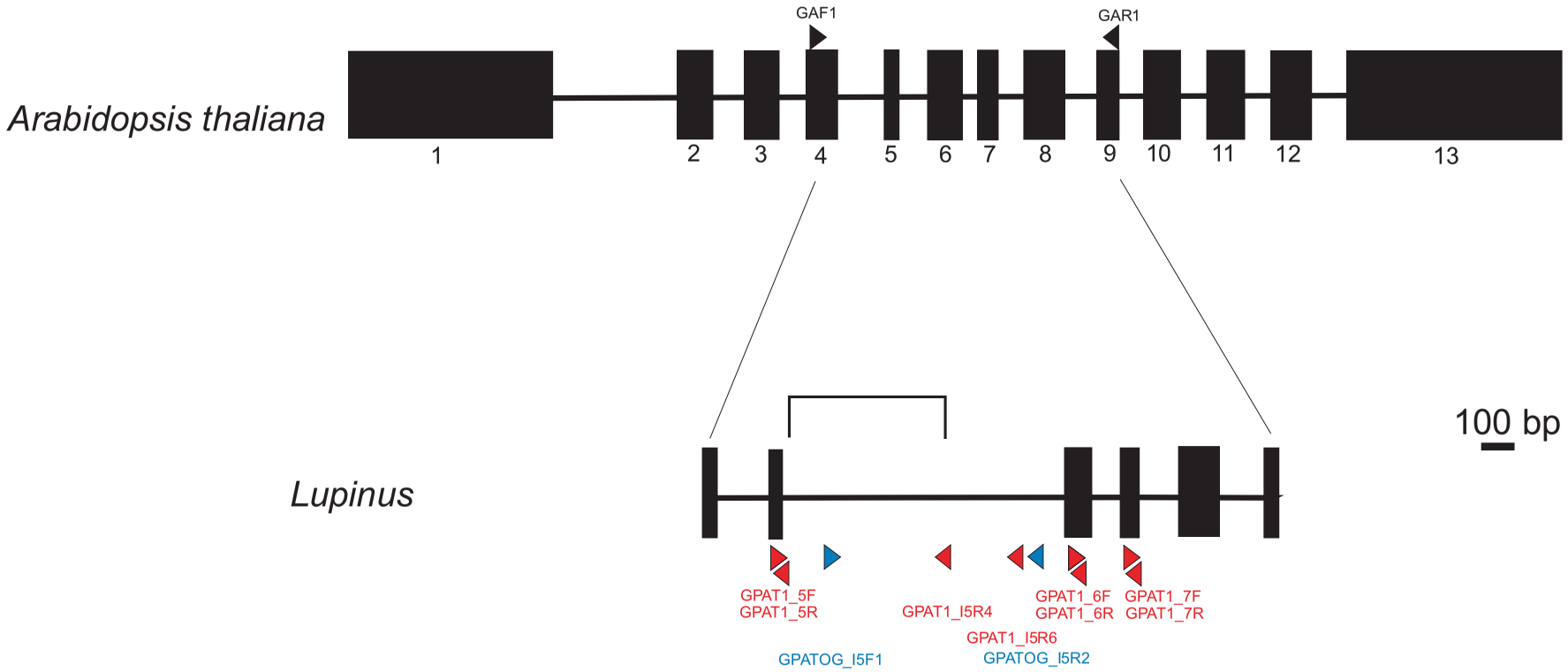


Fig. 3.2 Diagram of the full-length GPAT open reading frame in *Arabidopsis thaliana* and a fragment of the gene in *Lupinus*. Boxes, numbered from the 5' to the 3' end, represent exons, and lines between the exons represent introns. Lines connecting exons between *A. thaliana* and *Lupinus* indicate homologous exons. Arrows above or below exons indicate the location and direction of PCR primers used in this study. The bracketed region represents the region used for the phylogenetic analyses. The size of each region is indicated by the scale bar. Tank & Sang (2001) primers are in black, GPAT1 lupin-specific primers in red and outgroup primers in blue. Modified from Tank & Sang (2001).

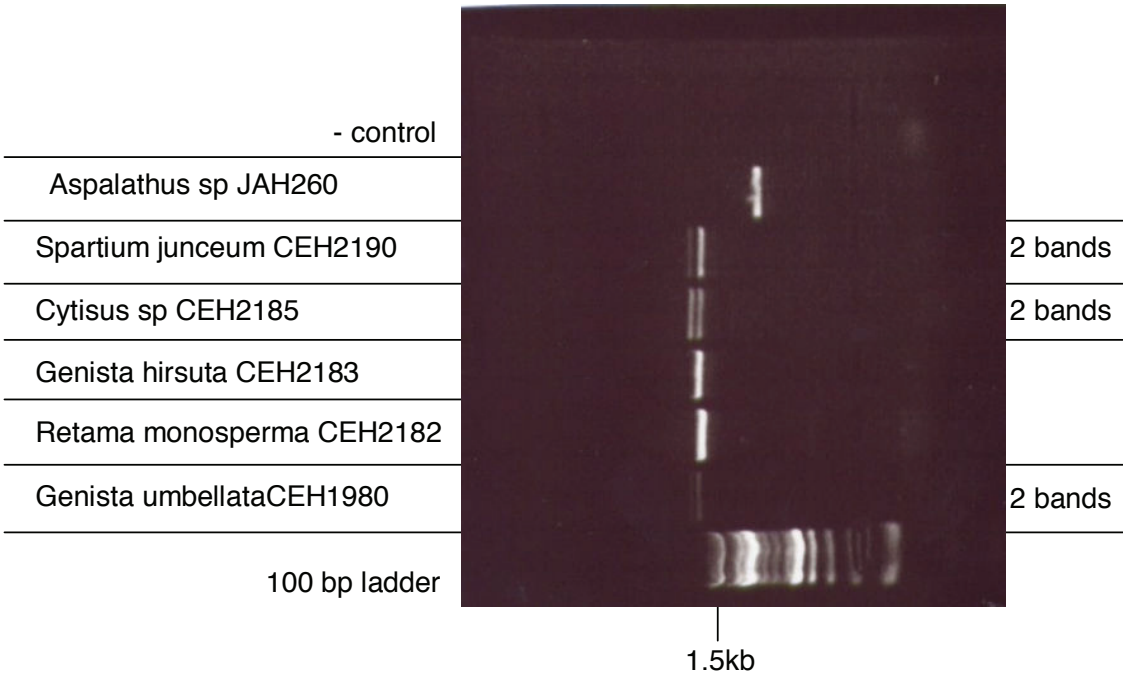


Fig. 3.3 Agar gel showing two PCR GPAT products in some Genisteae. The primer combination GPAT_F1 and GPAT6R was used in a PCR reaction utilising PCR beads. 3µl of product were run out on the gel alongside 1µl ladder.

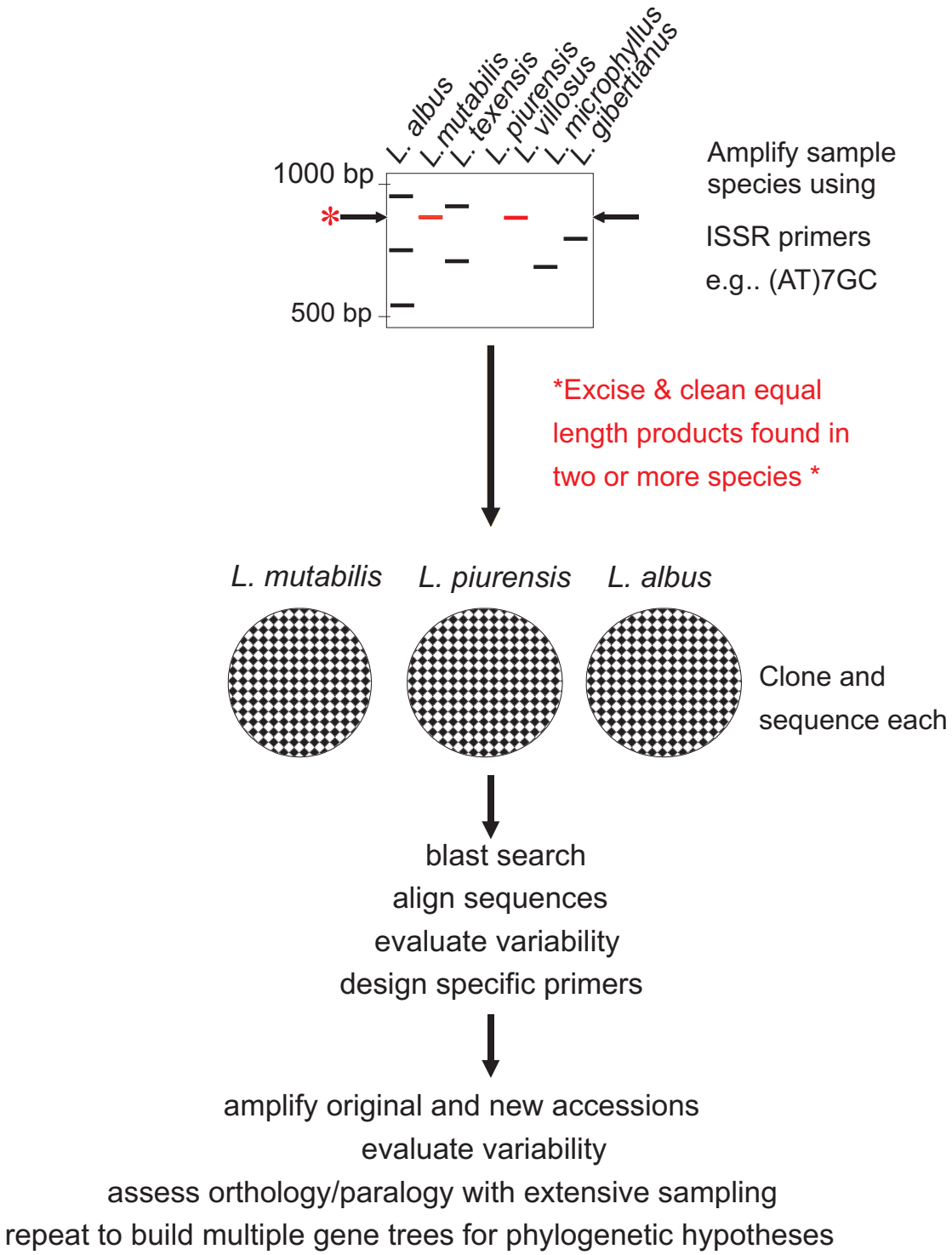


Fig. 3.4 SCAR method. This was utilised to isolate ISSR-based DNA sequence characterised amplified regions. Modified from Bailey *et al.* (2004).

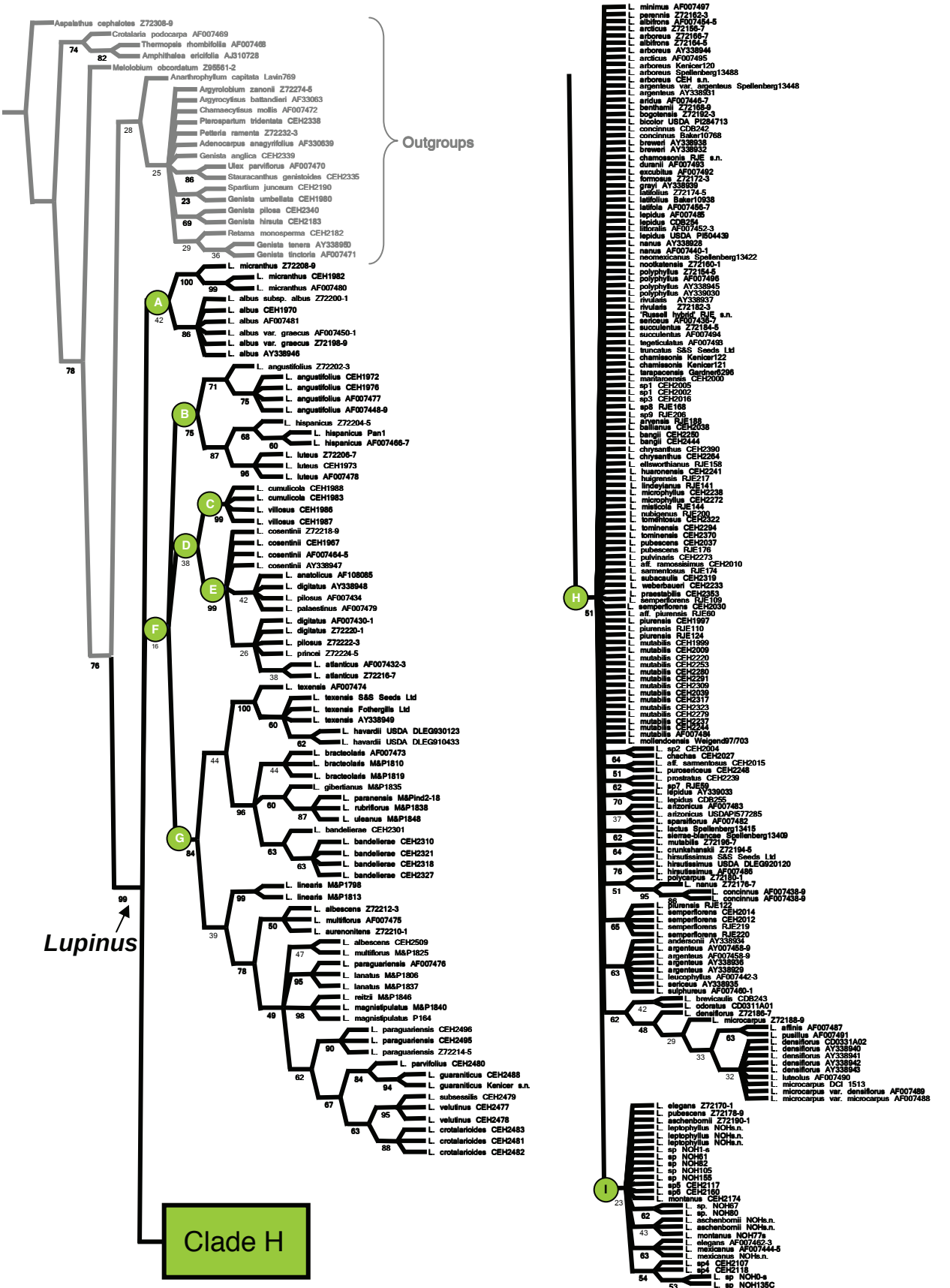


Fig. 3.5 ITS parsimony gene tree. Strict consensus of 27174 equally parsimonious trees of N=283, L=676, CI=0.45 and RI=0.85. Bootstrap values from 1000 replications are shown below branches. The parsimony rachet did not find shorter trees for this matrix. Species are named (where possible) and followed by either Genbank numbers or collector name and number. Clades referred to in the text are labelled A-I. Accession data and Genbank numbers are given in Appendix 5. The aligned matrix is available in the supplementary electronic Appendix E1.

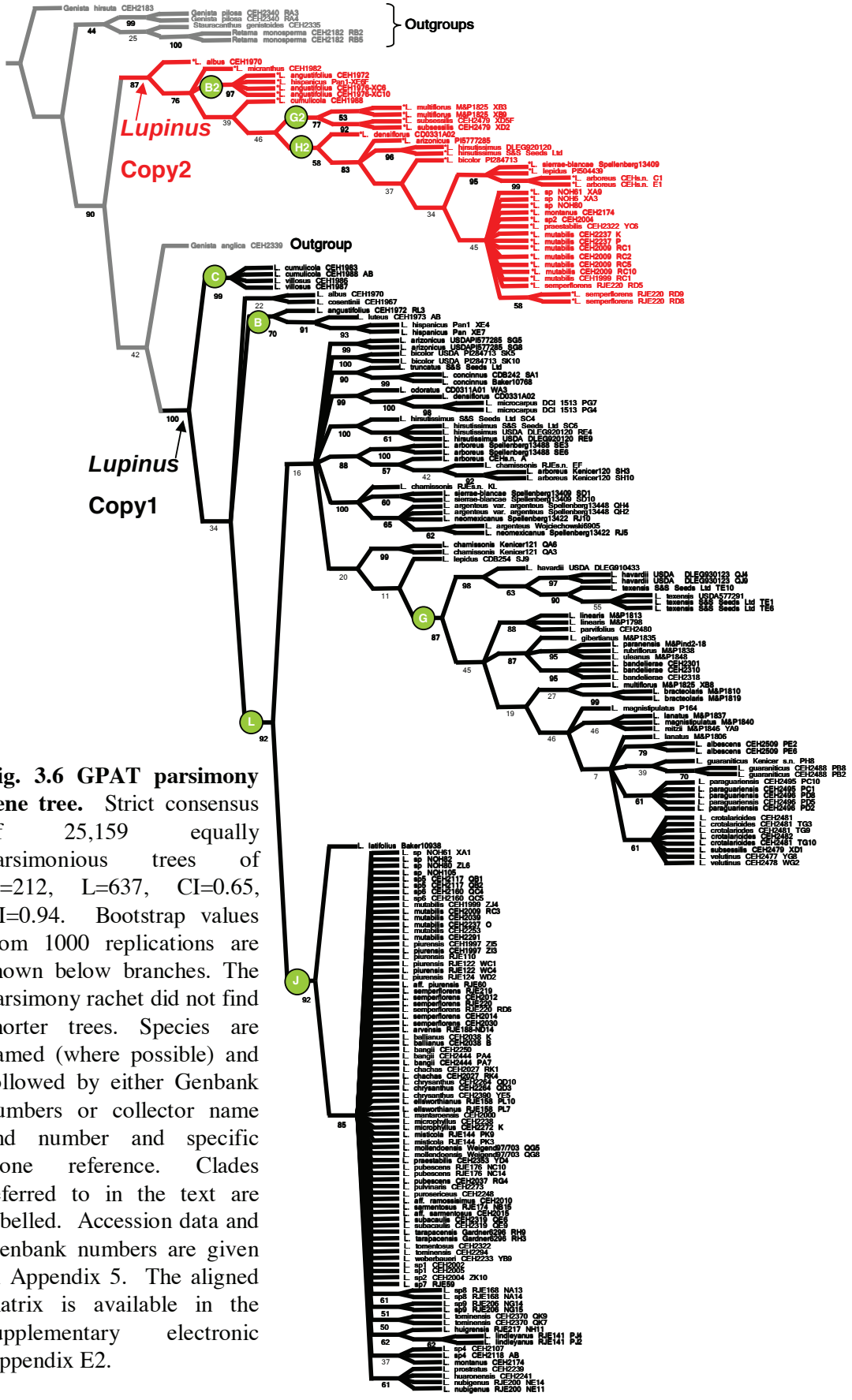


Fig. 3.6 GPAT parsimony gene tree. Strict consensus of 25,159 equally parsimonious trees of N=212, L=637, CI=0.65, RI=0.94. Bootstrap values from 1000 replications are shown below branches. The parsimony rachet did not find shorter trees. Species are named (where possible) and followed by either Genbank numbers or collector name and number and specific clone reference. Clades referred to in the text are labelled. Accession data and Genbank numbers are given in Appendix 5. The aligned matrix is available in the supplementary electronic Appendix E2.

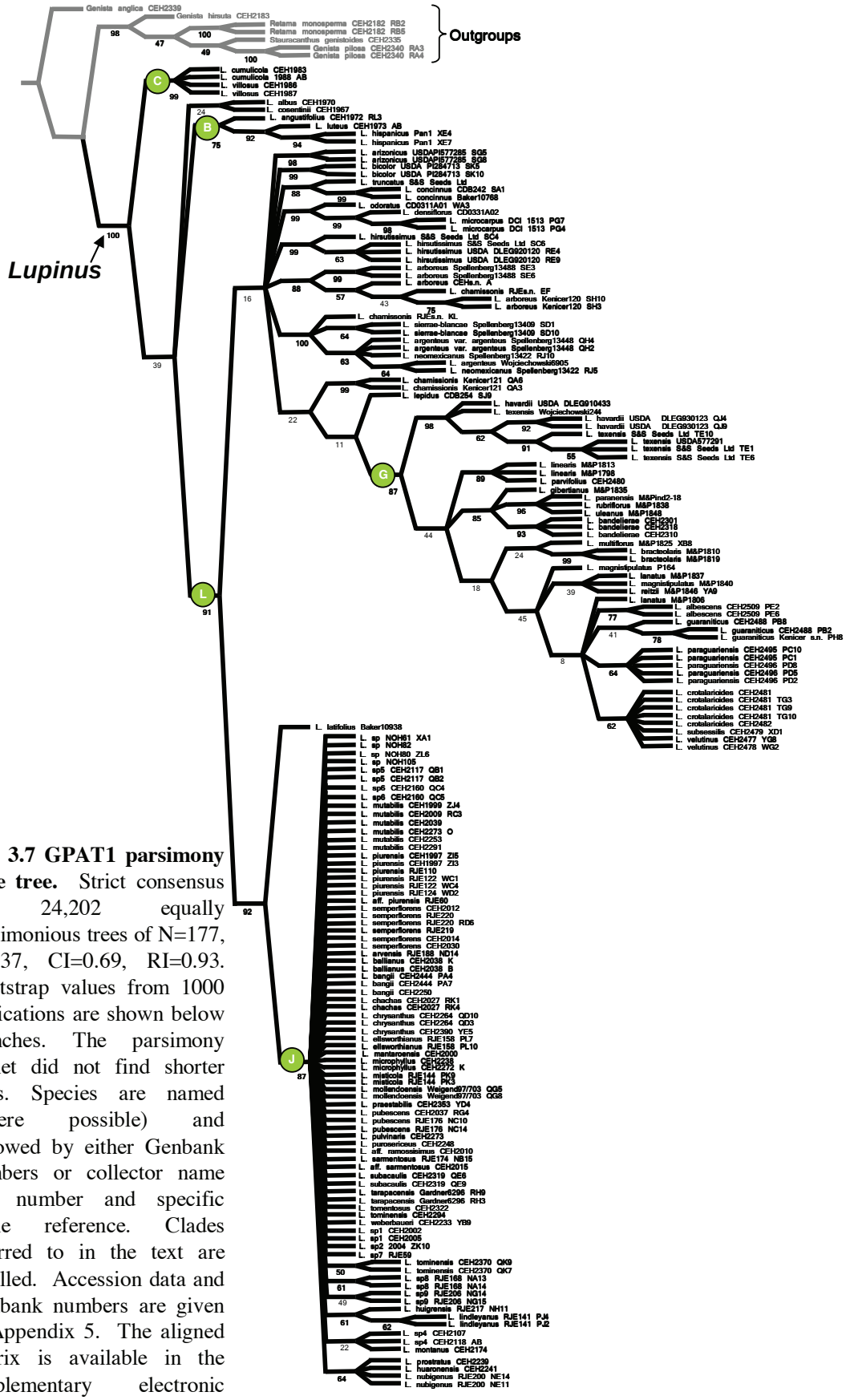


Fig. 3.7 GPAT1 parsimony gene tree. Strict consensus of 24,202 equally parsimonious trees of N=177, L=637, CI=0.69, RI=0.93. Bootstrap values from 1000 replications are shown below branches. The parsimony rachet did not find shorter trees. Species are named (where possible) and followed by either Genbank numbers or collector name and number and specific clone reference. Clades referred to in the text are labelled. Accession data and Genbank numbers are given in Appendix 5. The aligned matrix is available in the supplementary electronic Appendix E3.

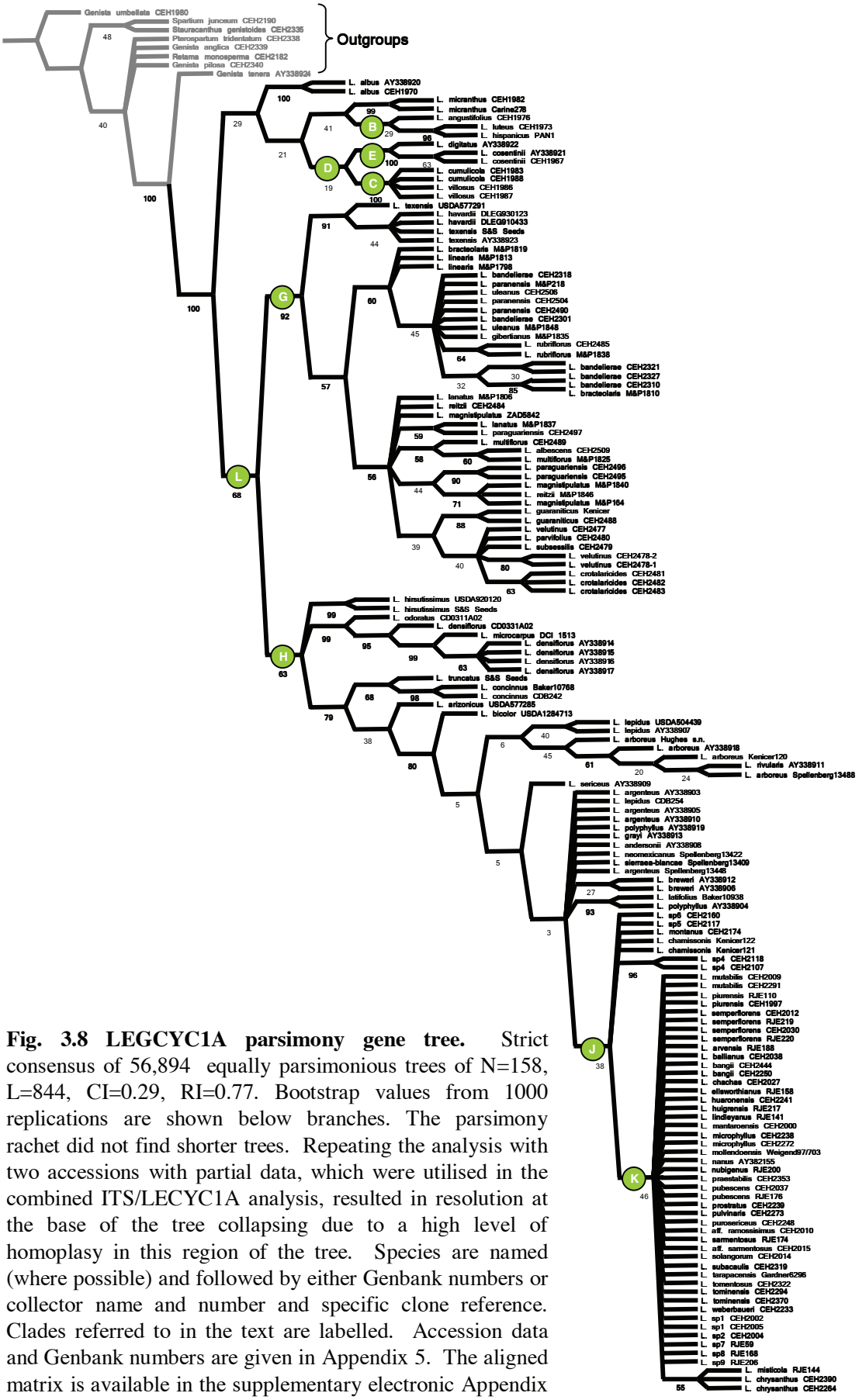


Fig. 3.8 LEGCYC1A parsimony gene tree. Strict consensus of 56,894 equally parsimonious trees of N=158, L=844, CI=0.29, RI=0.77. Bootstrap values from 1000 replications are shown below branches. The parsimony rachet did not find shorter trees. Repeating the analysis with two accessions with partial data, which were utilised in the combined ITS/LECYC1A analysis, resulted in resolution at the base of the tree collapsing due to a high level of homoplasy in this region of the tree. Species are named (where possible) and followed by either Genbank numbers or collector name and number and specific clone reference. Clades referred to in the text are labelled. Accession data and Genbank numbers are given in Appendix 5. The aligned matrix is available in the supplementary electronic Appendix E4.

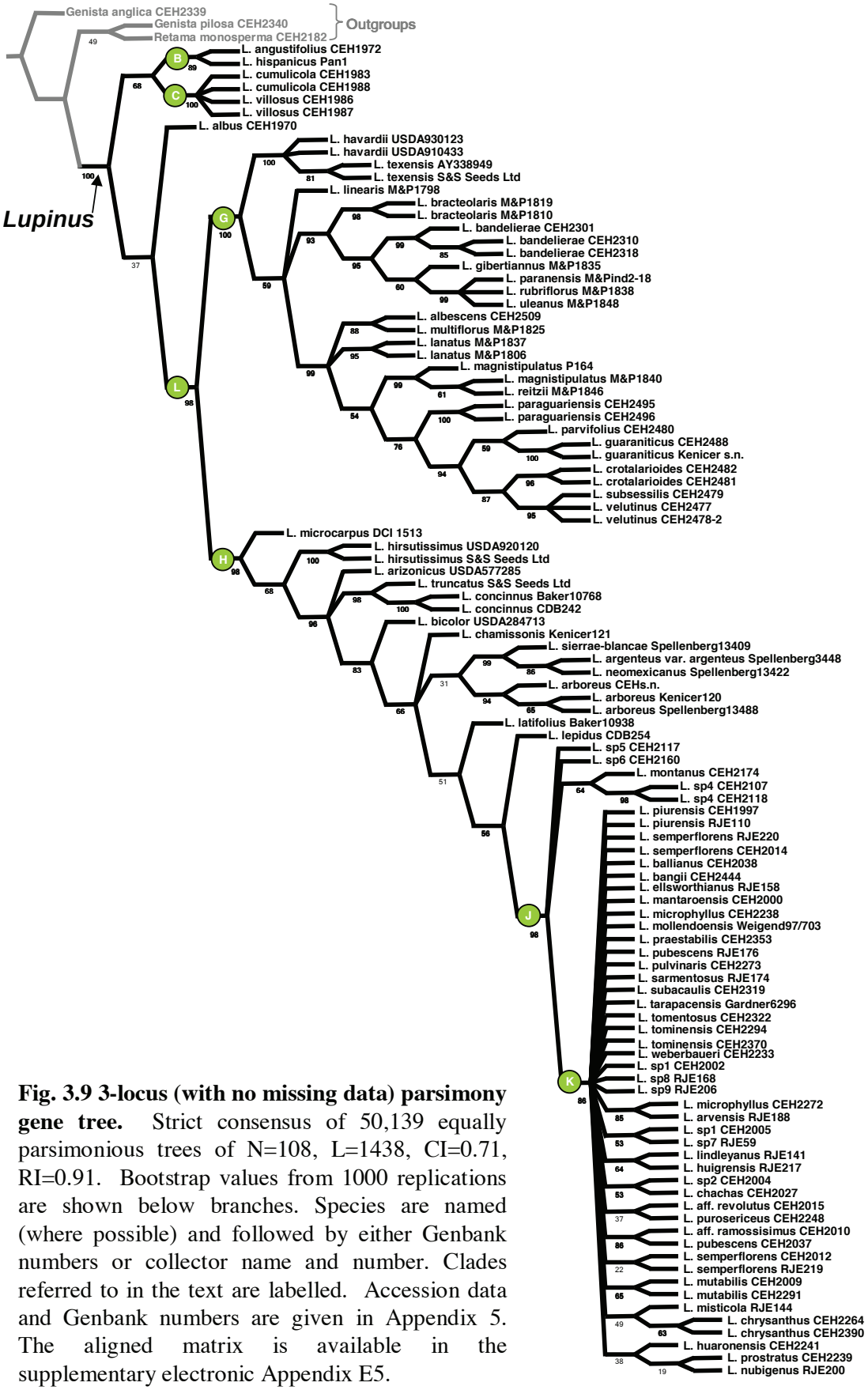


Fig. 3.9 3-locus (with no missing data) parsimony gene tree. Strict consensus of 50,139 equally parsimonious trees of N=108, L=1438, CI=0.71, RI=0.91. Bootstrap values from 1000 replications are shown below branches. Species are named (where possible) and followed by either Genbank numbers or collector name and number. Clades referred to in the text are labelled. Accession data and Genbank numbers are given in Appendix 5. The aligned matrix is available in the supplementary electronic Appendix E5.

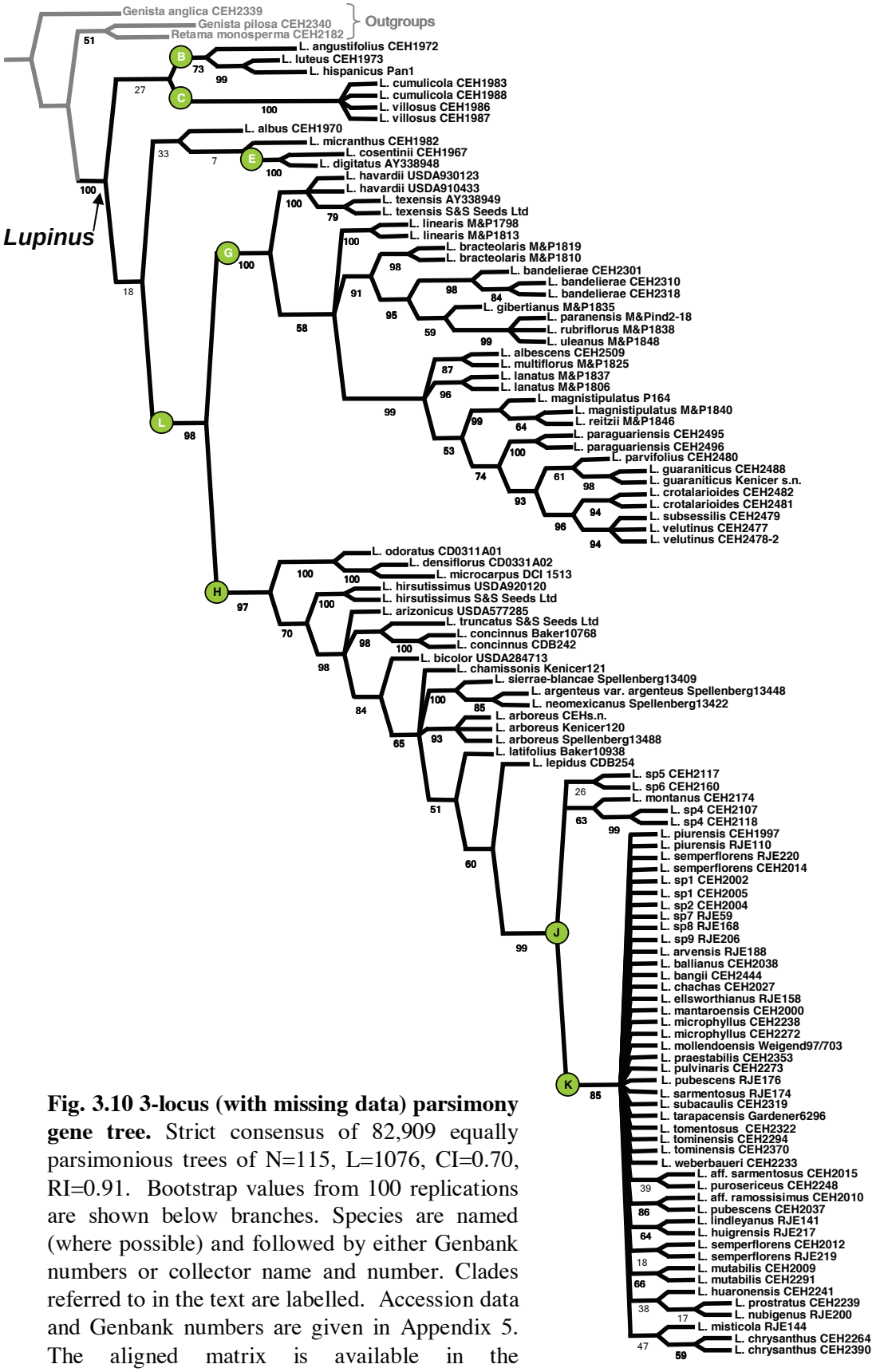


Fig. 3.10 3-locus (with missing data) parsimony gene tree. Strict consensus of 82,909 equally parsimonious trees of N=115, L=1076, CI=0.70, RI=0.91. Bootstrap values from 100 replications are shown below branches. Species are named (where possible) and followed by either Genbank numbers or collector name and number. Clades referred to in the text are labelled. Accession data and Genbank numbers are given in Appendix 5. The aligned matrix is available in the supplementary electronic Appendix E6.

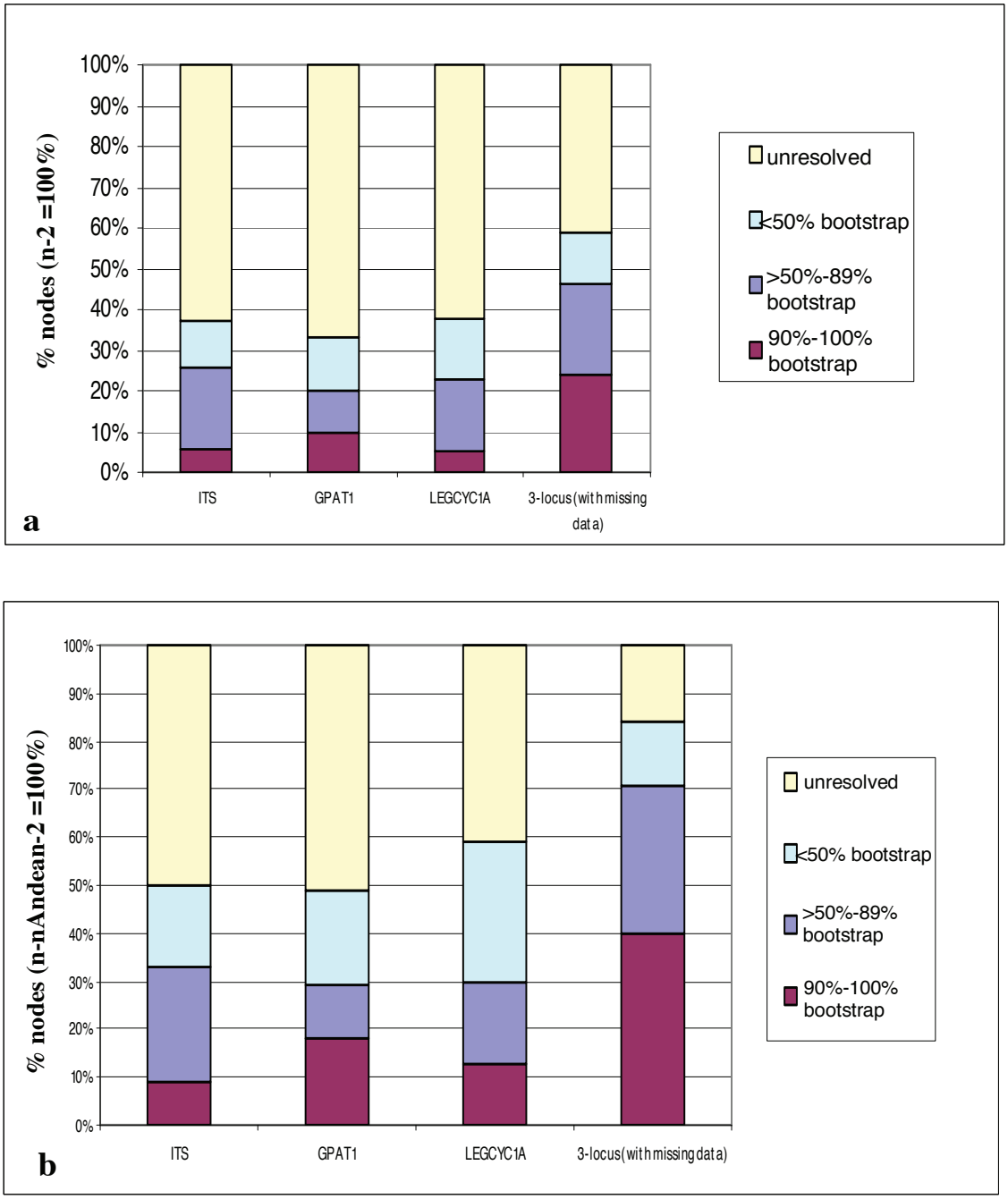


Fig. 3.11 Statistical summary of parsimony analyses. Levels of resolution and support shown by the ingroup in parsimony strict consensus gene trees from the individual ITS, GPAT1 and LEGCYC1A and the combined 3-locus (with missing data) datasets are compared. a) Ingroup dataset b) Ingroup excluding the large Andean clade (K Fig. 3.6) as defined by the 3-locus consensus tree. n=total number of terminals, nAndean=total number of terminals representing the Andean clade.

Fig. 3.12 ITS Bayesian gene tree. 50% majority rule consensus of 20,000 trees. Values below nodes are posterior probability values derived during analysis.

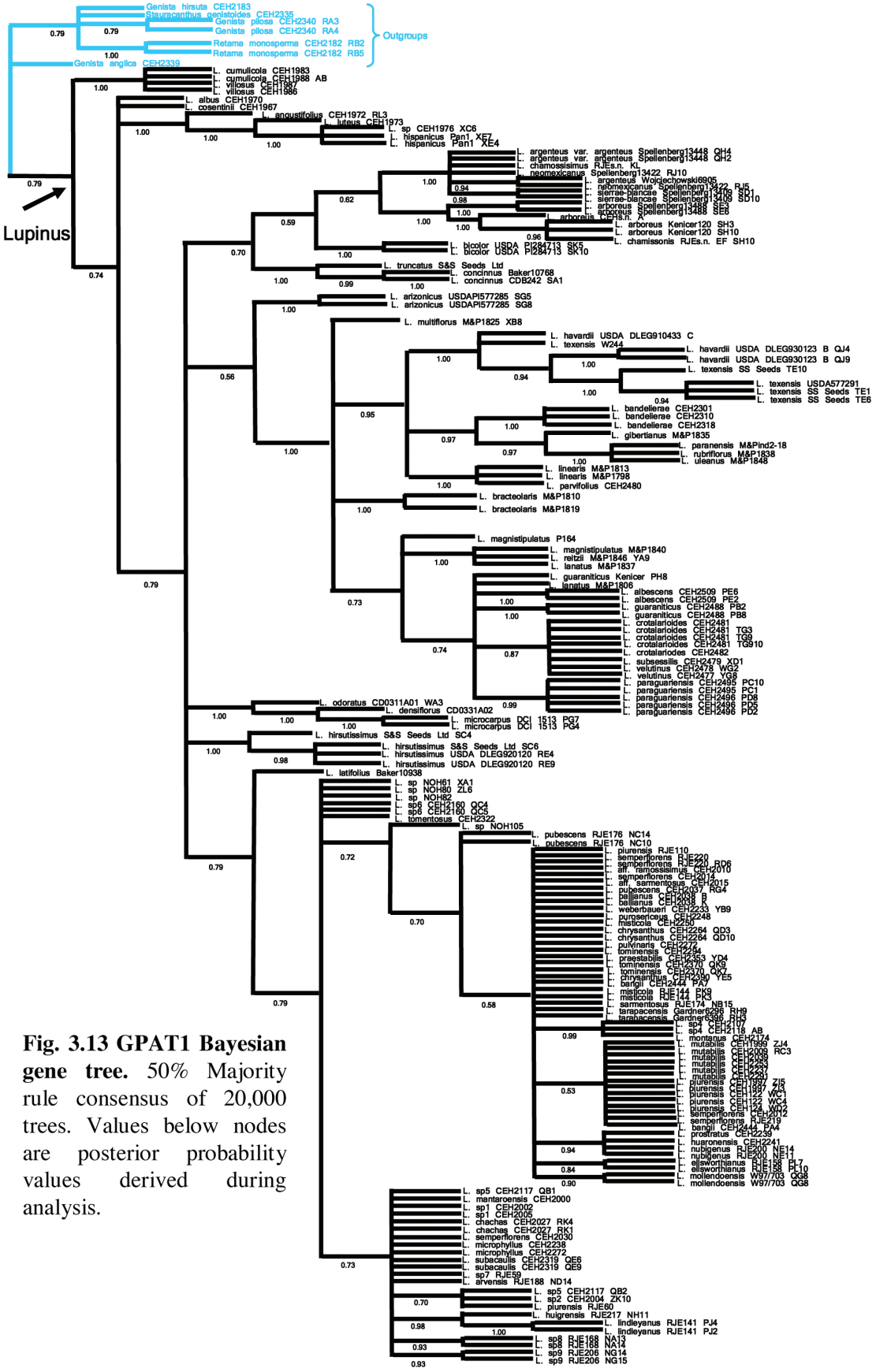


Fig. 3.13 GPAT1 Bayesian gene tree. 50% Majority rule consensus of 20,000 trees. Values below nodes are posterior probability values derived during analysis.

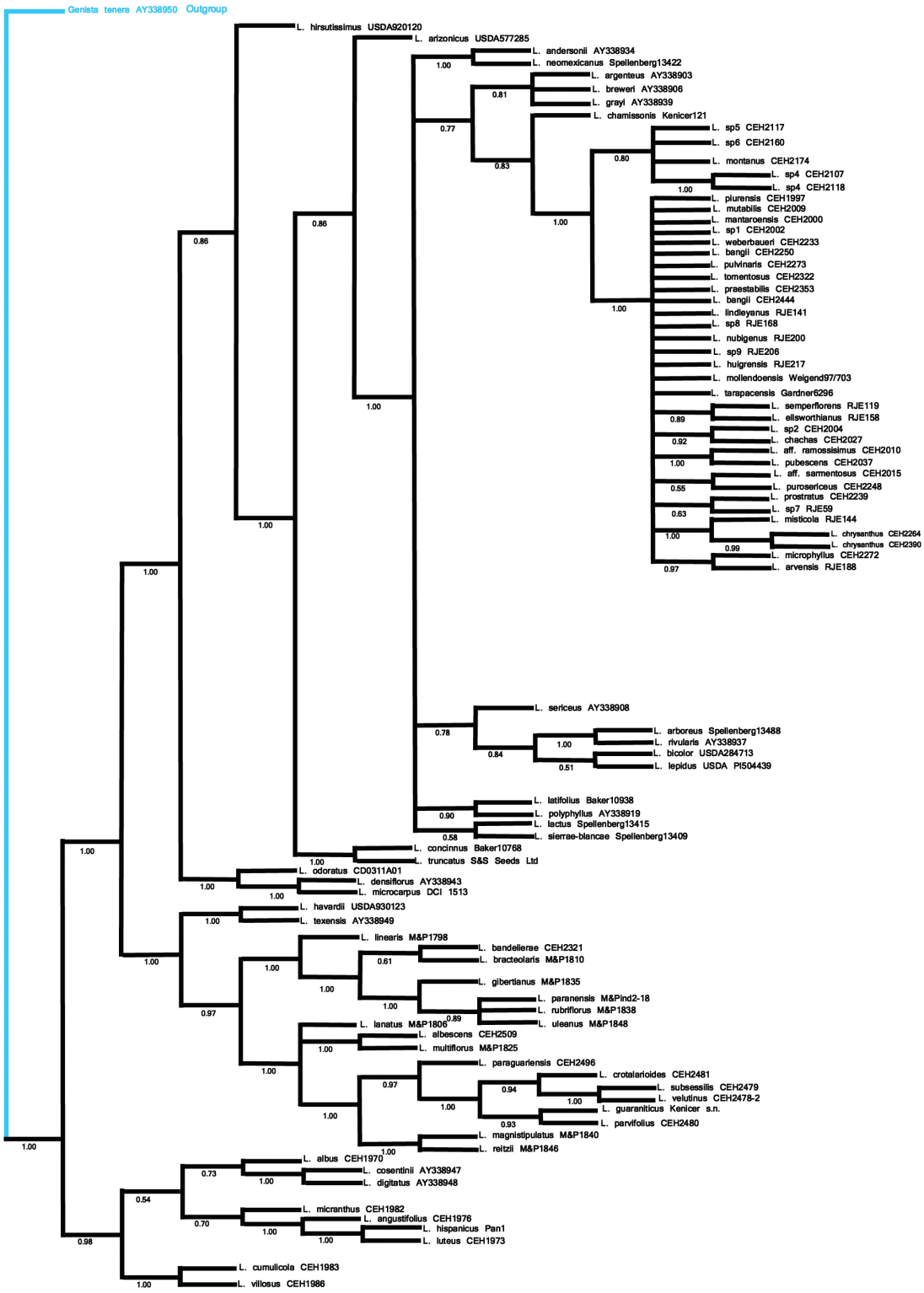


Fig. 3.14 ITS and LEGCYC1A Bayesian gene tree. 50% Majority rule consensus of 20,000 trees. Values below nodes are posterior probability values derived during analysis.

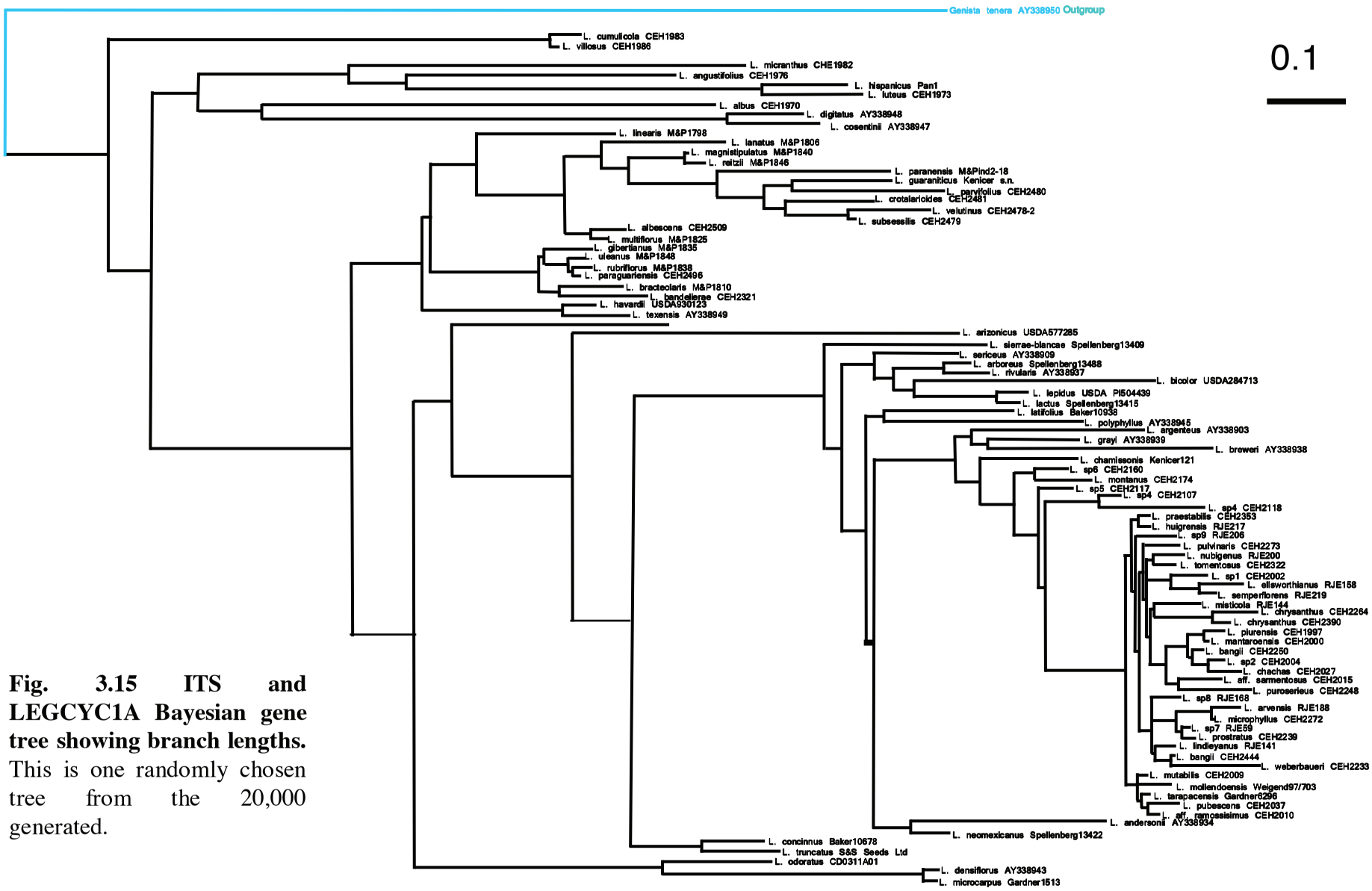


Fig. 3.15 ITS and LEGCYC1A Bayesian gene tree showing branch lengths. This is one randomly chosen tree from the 20,000 generated.

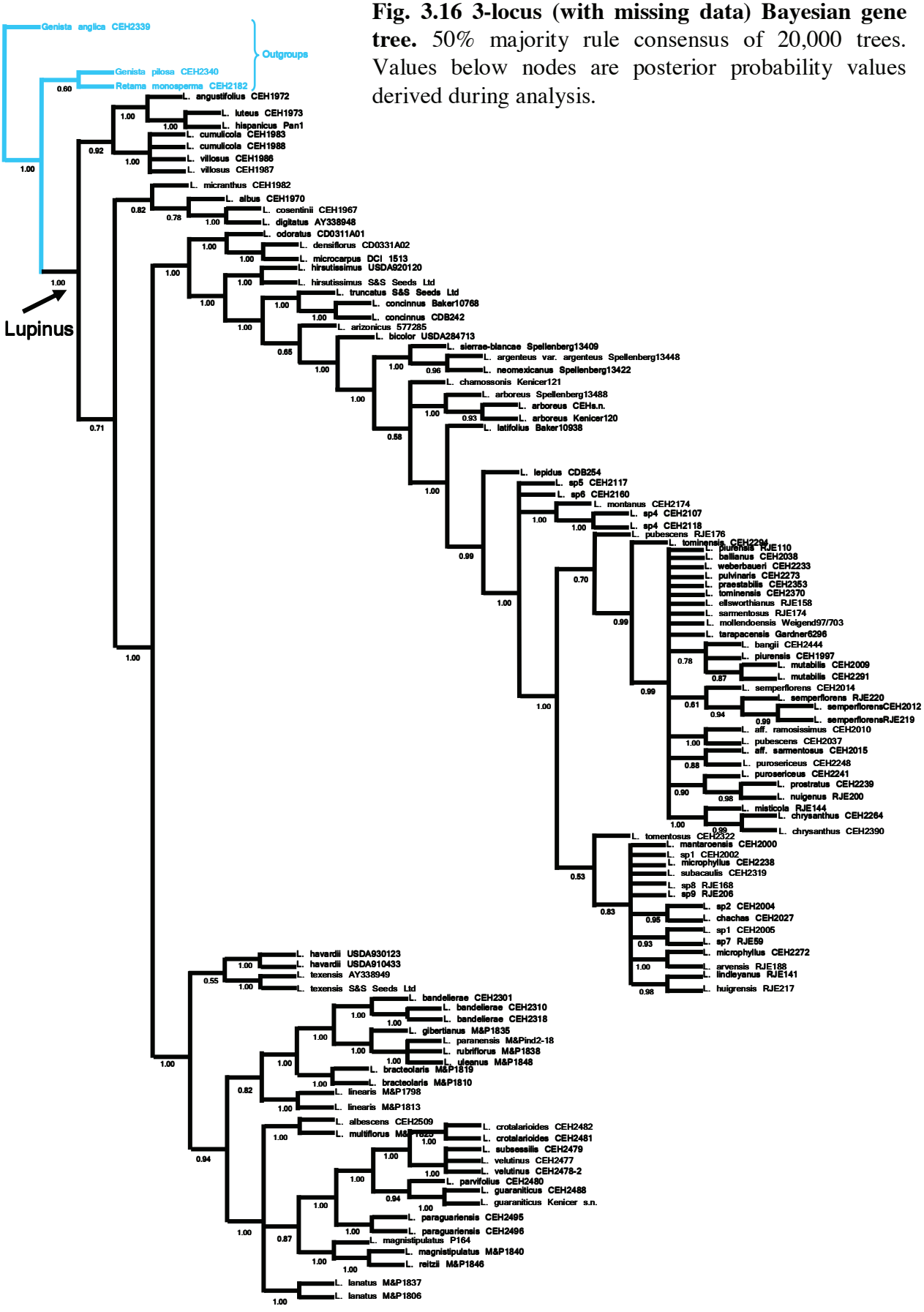


Fig. 3.16 3-locus (with missing data) Bayesian gene tree. 50% majority rule consensus of 20,000 trees. Values below nodes are posterior probability values derived during analysis.

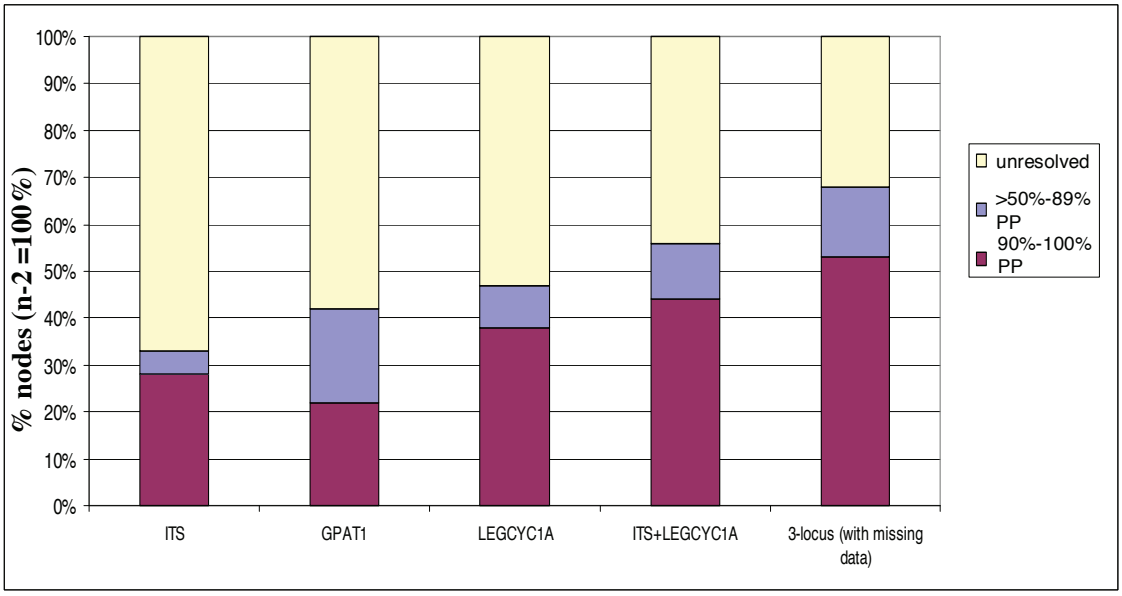


Fig. 3.17 Statistical summary of Bayesian analyses. Ingroup resolution and support is higher in the combined datasets than the individual loci. The posterior probabilities (PP) were generated from the 50% majority rule Bayesian gene trees from the individual ITS, GPAT1 and LEGCYC1A and the combined 3-locus (with missing data) datasets.

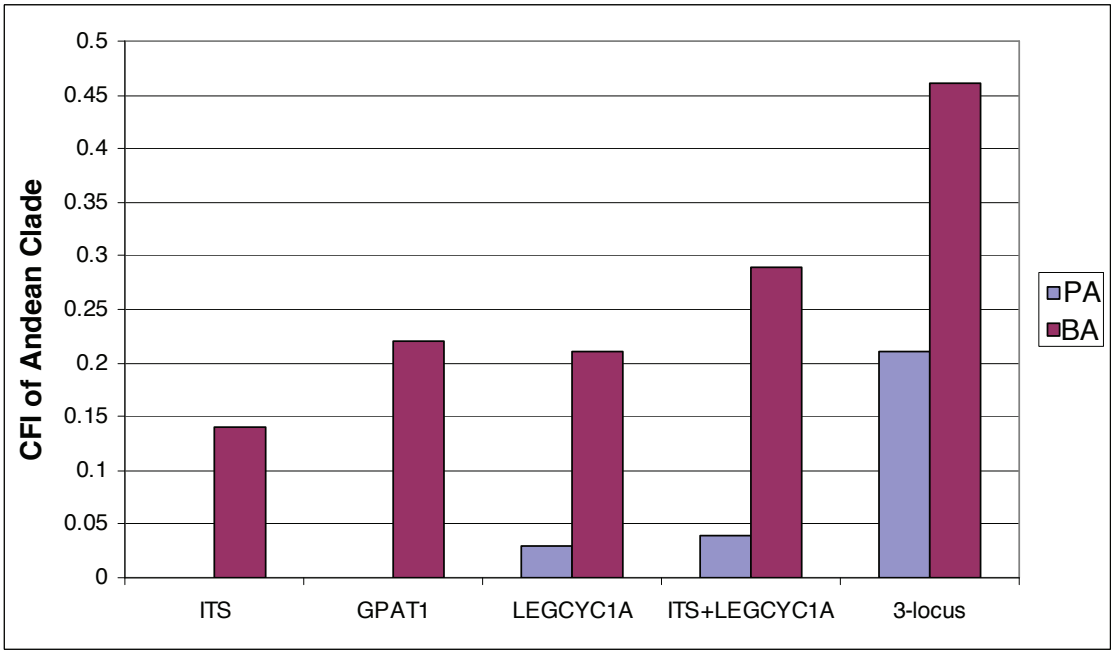


Fig. 3.18 Resolution in the Andean clade. Bayesian analyses (BA) resolved higher levels of resolution than parsimony (PA) within the Andean clade. The Andean clade was not resolved in the ITS and GPAT1 parsimony analyses.

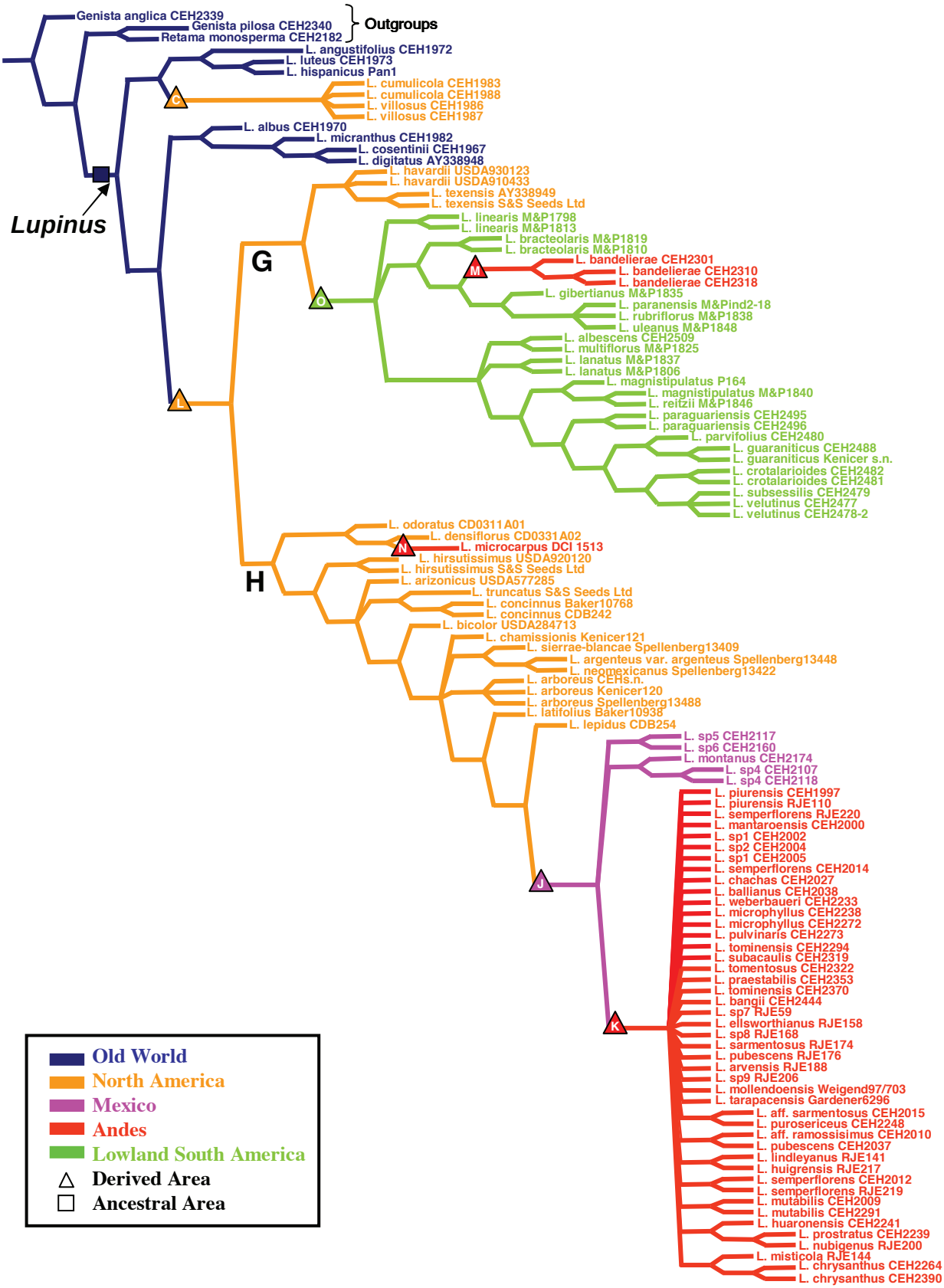


Fig. 3.19 Biogeography of *Lupinus*. Unambiguous optimisation of geography onto the *Lupinus* phylogeny. The probable ancestral area for *Lupinus* is the Old World. The two major clades G and H both have North and South American components. Clade G is largely comprised of lowland South American taxa whilst clade H encompasses the North American, Mexican and the majority of the Andean species. Andean taxa (shown in red) are found in three places on the tree (K, M, N). It remains unclear as to whether the disjunction in the *L. microcarpus* distribution between California and Chile and Argentina is natural or through human introduction.

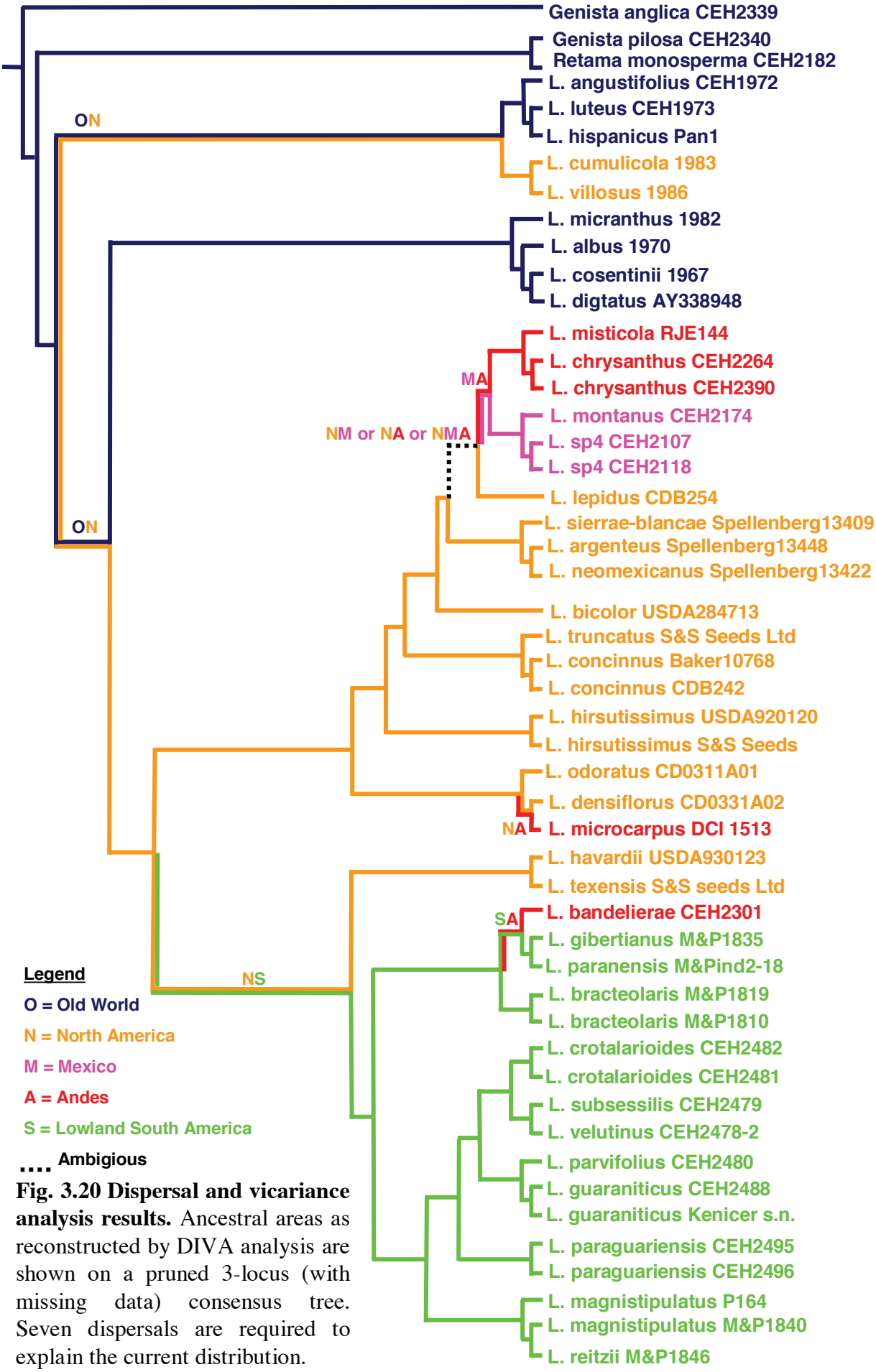


Fig. 3.20 Dispersal and vicariance analysis results. Ancestral areas as reconstructed by DIVA analysis are shown on a pruned 3-locus (with missing data) consensus tree. Seven dispersals are required to explain the current distribution.

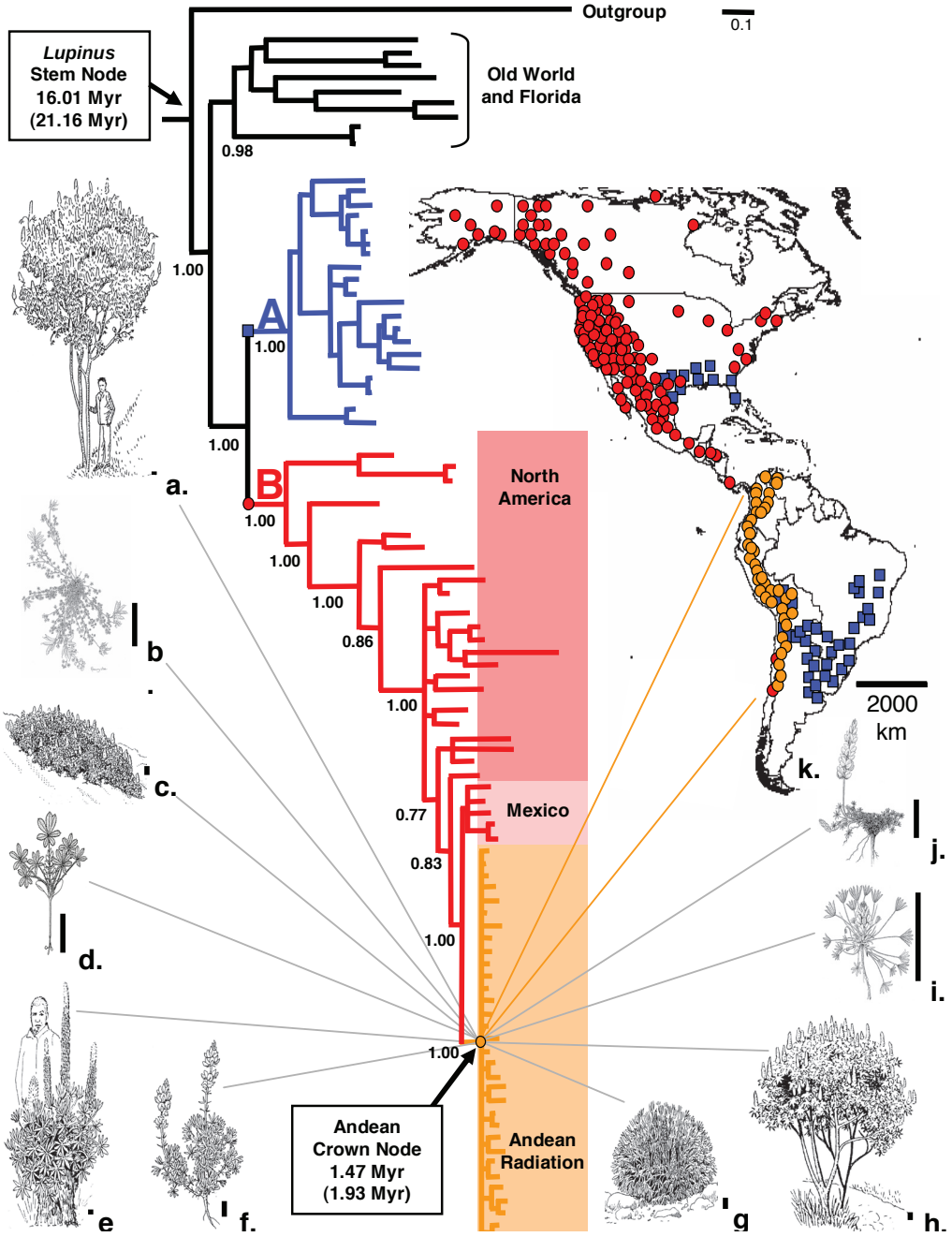


Fig. 3.21 Phylogeny, distribution and diversity of the Andean *Lupinus* radiation. 50% majority rule Bayesian tree from analysis of the combined ITS / *LEGCYCIA* data sets. Posterior probabilities of major clades are shown below nodes; all nodes had posterior probabilities > 0.5. This tree is congruent with the strict consensus derived from parsimony analysis. For clarity, taxon names have been omitted. Branch lengths are proportional to changes on the tree. A and B denote two well-supported New World clades: **A** – Eastern New World; B – Western North America, Mexico and the Andes. The map shows the distributions of these two New World clades to be largely allopatric, with limited overlap in the southern USA and the south-central Andes. The geographical extent of the main Andean radiation is shown in orange. Chilean accessions of *L. microcarpus* group with their North American counterparts at the base of Clade B, but the occurrence of *L. microcarpus* in central Chile (shown in red) is doubtfully native. Line drawings illustrate life forms encompassed by species in the Andean *Lupinus* radiation: (a) treelet, *L. semperflorens*; (b) prostrate herb, *Lupinus sp nov*; (c) perennial woody shrublet, *L. smithianus*; (d) ephemeral annual herb, *L. mollendoensis*; (e) giant stem rosette, *L. weberbaueri*; (f) woody perennial shrub, *Lupinus sp nov*; (g) acaulescent rosette, *L. nubigenus*; (h) perennial woody shrublet, *Lupinus sp nov*. (i) dwarf acaulescent rosette, *L. pulvinaris*; (j) prostrate herb, *L. prostratus*. Scale bars = 5 cm. The aligned matrix is available in the supplementary electronic appendix E7.

4 Domestication of *L. mutabilis*

4.1 Domestication

Domestication of crop plants remains poorly understood (Diamond, 2002) despite the fact that feeding the growing population is of great global concern (Food and Agriculture Organization of the United Nations 1996) and that domestication is regarded as one of the most fundamental developments in human civilisation (Cook, 1925; Hawkes, 1983). Understanding the morphological and molecular changes brought about through domestication relies on comparison of domesticates with their wild relatives and more specifically with their progenitors. Identifying progenitors can be challenging (Smartt, 1980) but a wide range of data sources and approaches are available including archaeology, molecular and morphological systematics, art, linguistics, taxonomy, evolution and developmental studies.

4.2 Approaches to understanding domestication

Investigations into domestication of plants and animals extend over a hundred and fifty years (Zeder *et al.*, 2006) but are still generating interest today, particularly as models for understanding the genetic basis of morphological change and diversity (e.g. Kislev *et al.*, 2006; Konishi *et al.*, 2006; Li *et al.*, 2006; Tanno & Willcox, 2006). Notable early research was carried out by Darwin (1868), de Candolle (1884) and Vavilov (1926; 1951; 1957). Important insights into domestication have been gained from integrating data from molecular phylogenetics, linguistics, artwork (Zeven & Brandenburg, 1986), morphology (Zohary & Hopf, 1994), biogeography and archaeology (Smith, 2001). Erickson *et al.* (2005) in their investigation into the origin and dispersal of the bottleneck gourd integrate morphological observations of plant remains from archaeological sites across America with genotype data from modern and ancient accessions epitomising the benefits to be gained from the 'consilience' of biological and archaeological approaches as advocated by Smith (2001), Zeder *et al.* (2006) and others.

Investigations into domestication have benefited from a wide range of technical advances (Brown, 1999) e.g. scanning electron microscopy (SEM) (Kislev *et al.*, 2006), more accurate techniques for dating (Kaplan & Lynch, 1999), new and more sophisticated molecular techniques and use of ancient DNA (Erickson *et al.*, 2005) to gain new insights, but are hindered by the growing globalization and intensification of agriculture, that is reducing the diversity of minor crops and landraces of major crops persisting in cultivation and threatening the habitats of wild species (Heiser, 1969). Biological (as opposed to anthropological) studies of domestication aim to identify the closest wild relatives including progenitors (undomesticated forms) of domesticates.

In the context of domestication molecular data can be used for a variety of purposes:
A To analyse sister group relationships. Through the identification of sister-taxa/groups of domesticates molecular data have reinforced previous hypotheses of relationships or proposed new hypotheses of relationships for crop species (e.g. Friesen *et al.*, 1999; Ingram & Doyle, 2003; Kochert *et al.*, 1996; Paris *et al.*, 2003). For example, sequence data from the nuclear-encoded but chloroplast-expressed glutamine synthethase gene identified two wild tuber-bearing taxa as candidate parents of the domesticated polyploid Andean tuber crop oca (*Oxalis tuberose* Molina: Oxalidaceae) (Emshwiller & Doyle, 1998; 1999; 2002). Previously, the origins of oca and the relationships of this domesticate with wild species were completely unknown. Similarly, ITS data have confirmed that the cultivated lentil (*Lens culinaris* Medik subsp. *culinaris*) is sister to its putative progenitor *L. culinaris* subsp. *orientalis* (Boiss.) Ponert (Leguminosae) (Sonnante *et al.*, 2003).

B To test whether a domesticate has a single (monophyletic) origin. Both tree and network building can be applied to molecular data to infer numbers of independent domestication events within a study group (Chacón *et al.*, 2005; Eubanks, 1997; Fofana *et al.*, 1999; e.g. Gepts, 1996; Gimenes *et al.*, 2002; Larson *et al.*, 2005; Londo *et al.*, 2006; Matsuoka *et al.*, 2002).

C To identify region(s) of origin. The topology of a phylogenetic tree can be informative about the origin of domestication. An example is shown by Olsen and Schaal (1999) where the nuclear-encoded glyceraldehyde 3-phosphate dehydrogenase gene was sequenced from 212 cultivated and wild cassava (*Manihot esculenta* Crantz) accessions. The geographic distribution of the haplotypes suggested an origin for domesticated cassava on the southern border of the Amazon basin (Olsen & Schaal, 1999).

D To investigate hybridisation and introgression and polyploidy. Molecular data can highlight hybridisation and introgression and provide information on the relationships between polyploids and diploids which is not always apparent from morphology (Wendel, 2000). Many important domesticated crops, e.g. wheat, are polyploids (Petersen *et al.*, 2006; Simmonds, 1986). For example the origin of the allotetraploid African cereal crop tef (*Eragrostis tef* Trotter) was investigated using sequence data from the nuclear gene *waxy* and the chloroplast region *rps16* and a likely source for one of the parental diploid genomes was confirmed (Ingram & Doyle, 2003).

E To estimate levels of genetic diversity. Domesticates characteristically exhibit low levels of genetic variation compared to their progenitors or other wild relatives (Weeden *et al.*, 1996). This is often explained with reference to domestication bottlenecks and the effects of selection. However, some domesticates, such as maize (*Zea mays*) and cucumber (*Cucumis sativus* L.) show surprisingly high levels of base pair polymorphism. Patterns of genetic variation can sometimes be partially explained by reproductive characteristics but can also be informative about domestication processes. For example, the high level of genetic variation found in *Phaseolus vulgaris* is often attributed to two separate domestications maintaining diversity (Chacón *et al.*, 2005). Weeden *et al.* (1996) compared the contrasting levels of genetic diversity in three legume crops, pea (*Pisum sativum*) cowpea (*Vigna unguiculata* (L.) Walp.) and lupins (*Lupinus spp.*). Cowpea followed the common pattern of very low genetic diversity in cultivated accessions compared to wild accessions. The level of genetic diversity within domesticated peas was not significantly lower than wild pea accessions

suggesting that domestication has been a relatively gradual process in which little genetic polymorphism has been lost and that some of the seemingly wild populations are naturalised (Weeden *et al.*, 1996). Isozyme data exhibited high variability in some wild *Lupinus* and low variability in others.

4.3 Morphological change associated with pulse domestication

Pulses were amongst the first domesticated crops alongside cereals, and were domesticated in both the New and Old Worlds e.g. lentils (*Lens culinaris*) in the Middle East, luna beans (*Phaseolus lunatus*) in central and South America, and soya beans (*Glycine max*) in Asia. A diverse range of genera (e.g. *Cicer, Lens, Lupinus, Phaseolus, Pisum, Vicia*) are utilised as pulse crops and provide a rich source of protein for human or animal consumption (Isely, 1982). A common pattern of morphological changes associated with the domestication of pulse crops has emerged. These are thought to be due to both unconscious and conscious selection (Hancock, 2004b). This pulse domestication syndrome involves several or all the following morphological and physiological characters:

- Increase in growth rate,
- Increase in pod and seed size,
- Development of indehiscent (non-shattering) pods,
- Increased testa permeability,
- Reduction in seed pigmentation,
- Reduction in seed dormancy,
- Loss of defence mechanisms,
- Reduction in seed/fruit alkaloid content,
- Increased self pollination.

Additional characters seen in some legumes are:

- smooth seed coats,
- change from perennial to an annual life cycle,
- day-neutral flowering,
- reduced morphological and physiological variability.

Many of these character changes can also be generalised to non-legume domesticates (Hancock, 2004c; Hawkes, 1983; Konishi *et al.*, 2006; Li *et al.*, 2006). Zohary (1989) has suggested that this is because the key traits of domestication are the breakdown of the wild mode of seed dispersal and the loss of germination regulation which could be unconsciously selected for during early transitions from foraging to cultivated agriculture regardless of the species involved.

The increase in the seed size seen in domesticated pulse legumes can be striking (Ladizinsky, 1979; Zohary & Hopf, 1973). However, evidence of intermediates is often absent from the archaeobotanical record. For example, neither beans recovered from the Guitarrero cave nor those recovered from coastal sites in Peru from 11,000 BP are significantly different in size from those found today (Kaplan, 1980; Kaplan & Lynch, 1999). For many species seed size is often the only reliable character which can be used as an indicator of pulse domestication in archaeological material although seed coat traits are also variable in some cases.

Within *Lupinus* the anatomical basis of increased testa permeability in the Old World domesticate *L. angustifolius* has been investigated by comparing soft (permeable) and hard (impermeable) seed variants. Only two differences were found despite extensive microscopic observations of the testa and hilar region (Valenti *et al.*, 1989). The apical and basal walls of the palisade cells in permeable seeds had transverse grooves whereas the impermeable variants had no grooves (Valenti *et al.*, 1989). An optical phenomenon known as the light line was observed in the impermeable seeds which has also been reported from impermeable seeds of *Glycine max* (Harris, 1987; Valenti *et al.*, 1989). The light line is thought to be caused by refraction of light from a denser region in the palisade layer (Dübbern de Souza & Marcos-Filho, 2001).

Relative percentages of seed coat and pod wall biomass have also been compared between wild and domesticated lupin species. However, wild sampling only included Old World

species. It was found that the domesticates *L. mutabilis* and *L. albus* had lower percentages than most of the wild species. However, this trait has not convincingly been demonstrated to have altered through domestication because a comparison of wild and domesticated *L. angustifolius* showed no differences (Clements *et al.*, 2005a) and the likely progenitors of *L. mutabilis* were not sampled. Verification of changes in seed coat and pod wall thickness due to domestication will require direct progenitor-domesticate comparisons.

The genetic basis of morphological diversity in general is poorly understood (Doebley, 2006; Ehrenreich & Purugganan, 2006) despite much recent research in this area (e.g. Bennett *et al.*, 2006; Bharathan & Sinha, 2001; Dengler, 1984; Gleissberg & Kadereit, 1999; Grigg *et al.*, 2005; Harrison *et al.*, 2005). Domesticated species provide some of the best opportunities for such studies because, like mutants, morphological changes can be directly compared in domesticates and their progenitors (when these are known) (Doebley, 2006). Numerous investigations to locate loci influencing domestication traits and their genetic dominance are in progress (Hancock, 2004a). For example, many pea mutants have been characterised including those showing domestication-like traits such as indehiscent pods (Weeden *et al.*, 2002). Quantitative trait locus (QTL) analysis, candidate gene identification and transposon tagging are utilised to investigate the genetic basis of quantitative characters (Emshwiller, 2006). Through QTL analyses it has been revealed that in many traits only a few loci are responsible for the majority of morphological changes (Clark *et al.*, 2006; Doganlar *et al.*, 2000; Konishi *et al.*, 2006; Li *et al.*, 2006; Nesbitt & Tanksley, 2002). The teosinte branched 1 gene in maize which converts the architecture of the wild type plant with long axillary branches, to the domesticated form with short axillary branches and one dominant shoot was cloned using transposon tagging (Doebley *et al.*, 1997; Wang *et al.*, 1999). Recently substantial breakthroughs in understanding the domestication of rice has been made with the identification of two major QTLs for shattering (grain falling from the plant when ripe), *sh4* and *qSH1* (Konishi *et al.*, 2006; Li *et al.*, 2006). Comparison of domesticates and progenitors provide an ideal system for elucidating the genetic basis of morphological evolution because, generally, relatively few yet dramatic morphological

changes can be observed. Such work has brought domestication centre-stage in recent developmental biology research (Doebley, 2006) spurring renewed interest in crop origins more generally (Zeder *et al.*, 2006).

Domestication continues today utilising traditional crop breeding methods as well as new technology. The characteristics sought by Australian lupin breeders in their efforts to domesticate rough-seeded lupins are the same as those seen in existing domesticates (Buirchell & Cowling, 1992). These are: soft and sweet-seeds, indehiscent pods and early flowering, key components of the pulse domestication syndrome.

4.4 South American prehistory

There has been much interest in the prehistory of the South American Andes and a large number of archaeological sites have been excavated (Fig. 4.1) (Keatinge, 1988; Pearsall, 1992). Despite this, the overall archaeobotanical record remains sparse and conclusions about the timing, progression and location of agricultural development in the Andes remain tentative. The distribution of these is biased towards the coast and uplands where conditions have been suitable for preservation of material (Hastorf & Popper, 1988). Botanical material from a small handful of sites has been extensively excavated and researched. Clearly, the range of material recovered is in almost all cases is biased towards material which preserves well. Dating of the sites can be controversial predominately because some of the items used for dating, such as animal bone, may not necessarily be directly related to human presence (Rick, 1988). Evidence of human occupation of some early sites is circumstantial leading to much uncertainty. However, even with all these caveats a large amount of information has been gathered about the prehistory of the Andes.

The earliest, albeit controversial, estimates of the time of arrival of humans in South America date to 20,000 BP (Burger, 1989; Rick, 1988). Evidence of human presence in the Andes ca. 13,000 BP, at the end of the Pleistocene, is present in at least three archaeological

sites in Peru, Pikimachay in Ayacucho, Pachamachay in Junín, and Guitarrero Cave in the Callejón de Huaylas (Fig. 4.1). A further as yet undated site, Uchcumachay also in Junín, is believed to be of similar age (Rick, 1988). Evidence consists of tools, stone-tool waste, animal bones and seeds (Rick, 1988). From 11,000 BP onwards there is strong evidence for widespread human occupation throughout the Andes. At this time it is thought that most societies were following a hunter-gatherer lifestyle (Rick, 1988). The transition to a sedentary lifestyle is thought to have been a prerequisite to the development of farming and subsequent domestication of plants and animals, or to have occurred simultaneously (Rick, 1988). There is much discussion in the literature surrounding the likelihood of a transitional phase of seasonal movement of people between contrasting localities e.g. the coast and sierra prior to complete settlement (Rick, 1980, 1988). There is evidence to suggest that in Peru settlement occurred first in the costal regions ca. 10,000 BP and later in the upland regions possibly as a consequence of the rich marine resources (Moseley, 1975).

It is clear that agricultural and wider societal development was not uniform in time or type across either the highland or coastal regions (Rick, 1988) or the different parts of the Andes. For example, agriculturally based societies apparently emerged earlier in coastal Ecuador than in the Ecuadorian highlands or Peru, but appearance and development of agriculture was as early in the Peruvian highlands as on the coast (Pearsall, 2003). This trend is also reflected in food plant assemblages which differed between early sites (Smith Jr., 1980) and crops appear to have been taken up earlier or later in different locations (Hastorf, 1999). Maize was introduced to Ecuador earlier than Peru (Pearsall, 1988) but after the formation of an agricultural society (Pearsall, 2003). The topography and potentially isolated conformation of the valleys may have influenced this by limiting early exchange of cultural knowledge (Rick, 1988).

4.5 Domestication in the Andes

The Andes are one of the eight/nine geographic regions where domestication of crops and livestock occurred independently 13,000 – 8,000 years ago (Diamond, 1997, 2002; Piperno & Pearsall, 1998; Smith, 1998, 2001). In fact the Andean region was a hotspot for domestication of both plants and animals (Cook, 1925; Wing, 1986). Some of the Andean crops e.g. potato, tomato and beans are well-known as they have become globally important. However, a whole host of other crops sometimes dubbed the "*Lost Crops of the Incas*" (National Research Council, 1989a) remain largely confined to the Andes and are little known outside that region (Cook, 1925). These plants however continue to form the staple diet of many Andean communities, now as in the past (Smith Jr., 1980), and are extensively cultivated, alongside introduced crops such as broad beans and maize, throughout the Andes. However, in practise the description as Lost Crops of the Incas is misplaced as it is thought that the majority of these crops were domesticated well before the Incas (see discussion below) even though Incas were renowned for their agricultural organisation and sophistication e.g. in storage of crops (D'Altroy & Hastorf, 1984). Andean plant domesticates come from a wide range of plant families and utilise a variety of plant organs e.g. root, tuber, fruit and seed (National Research Council, 1989b; Pearsall, 1992). As well as being adapted to the harsh Andean conditions the crops and cultivars are adapted to specific microenvironments which are well-known by Andean communities (Brush, 1976; Hancock, 2004a).

What, where and when the first crops that were domesticated in the Andes remains uncertain. The first group of food plants, although not necessarily all confirmed as domesticates, to appear in prehistoric sites includes: *Capsicum sp., Cucurbita sp., Lagenaria* Ser., *Oxalis tuberosa, Phaseolus vulgaris, Phaseolus lunatus* and *Ullucus tuberosus* Caldas. Also appearing at this stage are many fibrous plants thought to be utilized in cord manufacture (Smith Jr., 1980). The earliest date given for a domesticated plant, albeit one primarily not utilized as a food source but as a container, *Lagenaria* in the Ayacucho Cave complex, is 15,000 – 13,000 BP but both the date and identification are controversial (Erickson *et al.*, 2005;

Smith Jr., 1980). The recovery of *Phaseolus lunatus, P. vulgaris* and *Capsicum chinensis* Jacp. remains from the Guitarrero Cave ca. 11,000 BP are more widely accepted as the earliest cultivated plant remains in South America (Lynch, 1980). However, more recent direct radiocarbon dating of the *Phaseolus* remains estimated the Guitarrero Cave material at 4335+/- 55 yr BP (*P. vulgaris*) and 3495 +/- 50 yr BP (*P. lunatus*) (Kaplan & Lynch, 1999). The same study dated *P. lunatus* beans from the coastal Chilca site to be older at 5616 +/- 57 yr BP (Kaplan & Lynch, 1999).

Domestication in the Andes is thought to have blossomed from 5500 BP – 3500 BP concurrent with major development and increase in societal complexity in both Ecuador and Peru (Pearsall, 2003). During this period there is evidence for the use of a much wider variety of crops including: *Amaranthus* sp., *Capsicum chinensis, Chenopodium quinoa, Curcurbita* sp., *Gossypium* sp., *Lagenaria siceraria, Lucuma bifera, Oxalis tuberosa, Phaseolus lunatus, P. vulgaris, Solanum* sp., *Ullucus tuberosus* and *Zea mays* with no single dominating crop (Pearsall, 2003). Later, there is evidence of interaction between coastal and upland communities with remains of plants specific to each turning up in locations not suitable for their cultivation (Perry *et al.*, 2006). In total, 160 species of crop plants have been listed as domesticated for food, medicine, fabric/fibre, tools and housing in pre-ceramic Andean times (Tapia & De la Torre, 1998).

With a few exceptions, e.g. beans (Gepts, 1996), the origins of most of the Andean crops remain highly speculative. For example it has bee proposed that achira, *Canna edulis* originated in Colombia based simply on South America as a centre of diversity for this genus and an early root-crop tradition in this region (Gade, 1966). Recent work on the Andean tuber domesticate oca, *Oxalis tuberosa* also illustrates the challenges associated with identification of the progenitor due to incomplete taxonomy and low levels of molecular variation (Emshwiller & Doyle, 1998). Nevertheless, two putative tuber-bearing progenitors of oca from southern Peru and Bolivia have been tentatively identified (Emshwiller & Doyle, 2002).

Considerable research has been carried out into the origins of *Phaseolus* sp. (Chacón *et al.*, 2005; Debouck *et al.*, 1987; Delgado-Salinas *et al.*, 1999; Fofana *et al.*, 1999; Garcia *et al.*, 1997; Gepts, 1990, 1996; Kami *et al.*, 1995). These have focused on the number and location of domestication events. Multiple molecular markers suggest at least two independent origins, Mesoamerica and the Andes, for both *P. vulgaris* and *P. lunatus* (Gepts, 1996). Identification of wild *P. vulgaris* required data from morphology, distribution, hybridisation and protein type distribution (Gepts, 1990). The earliest beans found are comparable in size to modern day beans (Kaplan, 1980) thus either they were large-seeded to begin with or all those recovered were domesticated. A comparison of wild and domesticated *Phaseolus vulgaris* confirms adherence to the classical legume domestication syndrome including a decrease in morphological and physiological variation (Garcia *et al.*, 1997).

Thus, there are still many unanswered questions surrounding domestication in the Andes. No obvious temporal and geographic patterns akin to those emerging in Mesoamerica (Smith, 2001) have yet been observed. Was there one main centre for domestication, if so where? Was the assemblage of crop plants built up gradually or were there bursts of domestication? Who undertook domestication and when? Were crops domesticated multiple times? Is the particularly high number of domesticates just a result of the outstanding levels of wild plant diversity, a product of the Andean topography or attributable to other anthropological factors. For many crops the origins and changes associated with domestication have not been investigated and quantified. Until more is know about domestication events for Andean crops, answers to these questions are likely to remain elusive.

4.6 Archaeological evidence for Tarwi, *L. mutabilis* domestication

Speculation surrounding the domestication of *L. mutabilis* has been fuelled by the very limited archaeological evidence of lupin remains prior to 1,800 BP. Only eleven references to lupin seed, pollen or prehistoric illustrations have been located. These are reviewed below and illustrated in the timeline in Figure 4.2. As for many pulses, the main indicator of

domestication that can be assessed from archaeological remains of tarwi is seed size, which varies dramatically from wild to domesticate (see below). Small infrequently distributed seeds are assumed to be of wild species while those in the size range of modern *L. mutabilis* occurring more abundantly are judged to be domesticated.

The earliest report of wild lupin seed remains from the Andes is dated 9,010 – 8,125 BP. Lupin seeds occur sporadically throughout the Pachamachay (Fig. 4.1) survey horizons. Figures 9.3D and E from Pearsall (1980) confirm the identification of the seeds as lupins and they are very similar in size and shape to a range of modern collections of wild species. Given the number of wild lupin species found in this area today and lack of any firm evidence for domestication for any of those species, it seems unlikely these seeds indicate *Lupinus* in cultivation. It is not impossible to imagine the casual or passive movement of lupin seeds into the cave.

Wild lupin seeds were also discovered at the coastal site of Ancoń (Fig. 4.1) (Towle, 1961). No specific age is estimated for these but the earliest plant remains from this site date to 10,000 – 8,000 BP. The presence of lupin seeds at a coastal location is significant since all but one of the Andean lupin species are distributed at altitudes above 1000 m and the one low elevation species is distributed only in southern Peru. This suggests some movement of humans between the coast and the sierra. The discovery of shells in the sierra reinforce this movement pattern (Engel, 1970; Lynch, 1983; Ravines, 1972).

Small wild legume seeds including those of *Lupinus* species have also been found throughout the continuous occupation of the La Chimba site, Ecuador (Fig. 4.1), from 4440 - 3500 BP. At this site plant remains of domesticated maize (*Zea mays*) and potato (*Solanum tuberosum*) are found from the earliest horizons with beans (*Phaseolus* sp.), cotton (*Gossypium* sp.), oca (*Oxalis tuberosa*) and other tubers appearing from 4365 BP (Pearsall, 2003). During this period root crop and maize agriculture has been interpreted as providing the staple diet supplemented by wild gathered resources (Pearsall, 2003) suggesting lupin seeds were

initially harvested from wild species prior to domestication.

Panaulauca Cave, Junin, Peru (Fig. 4.1) has also yielded lupin seeds (Pearsall, 1988). This site is interesting as it is thought to have been occupied from 5,000 - 1,400/1,500 BP i.e. spanning the timeframe in which domestication of *Lupinus* is likely to have occurred in the Andes (Fig. 4.2). However, Lupin seeds were found only infrequent at the site and are only recorded from the cave mouth midden (Pearsall, 1988). Pearsall (1988) lists lupins as used for food at this site but the basis for this is not explicitly listed and seed size is not discussed so it is presumed that a change in seed size over time was not observed. The limited quantity of seed and its peripheral location do not suggest that lupins were cultivated or being used for food at this site.

Lupin seeds have also been reported from the Cotocollao site (Fig. 4.1) in northern Ecuador (Pearsall, 2003). The earliest date for this site is ca. 3545 BP. Other plant material found included maize, tree fragments, roots and tubers. Pearsall (2003) comments "the presence of lupine seeds adds another potential cultigen to the assemblage" but it is not clear whether these seeds are small and wild or possible domesticates. Neither is there any information on their abundance.

The abundance and size of lupin seeds in the Mantaro Valley, Peru (Fig. 4.1), provide convincing evidence to suggest that domesticated *L. mutabilis* was widely cultivated in this area by 1,800 BP alongside oca (*Oxalis tuberosa*), ullucu (*Ullucus tuberosus*), mashua (*Tropaeolum tuberosum*), potato (*Solanum tuberosum*), beans (*Phaseolus* sp.), maize (*Zea mays*) and quinoa (*Chenopodium quinoa*) (personal communication Christine Hastorf). It is estimated that yields of 119 – 514 kg/ha were produced (Hastorf, 1993) which is lower on average than production today which ranges from 300 to 2000 kg/ha (Bellido, 1984). *Lupinus* seeds have also been recovered associated with ceramic remains in Inca store houses in the Mantaro valley ca. 540 BP (D'Altroy & Hastorf, 1984), consistent with the idea that lupins were cultivated in this area as they are today.

Lupinus mutabilis seeds have been reported from Nazca tombs (ca. 1,500 BP) (Antúnez de Mayolo, 1982) but without any specific date or locality. The colour and size of these remains do not confirm these were seeds of domesticated lupin, as native species close to such a description can be found today. The morphology of the seeds is reported as 5-6 mm long, 4-5 mm wide and black. These preserved seeds have a similarly high protein content (42%) to modern seeds (40-55 %) (Kurlovich *et al.*, 2003; Torres Tello, 1976). They are large by wild lupin standards but smaller than modern *L. mutabilis* (see Fig. 4.10-12). However, the size difference could be either an artefact of extreme drying or an indication of early stage domestication.

Finally, there is reference to the cultivation of *L. mutabilis* in South America in the form of requests to tax lupin crops during the Spanish occupation (references in Antúnez de Mayolo, 1982; and Gross, 1982). It is estimated 100,000 hectares of *L. mutabilis* were grown ca. 450 BP (Gross, 1982). This is far greater than modern estimates which suggest that as little as 5,200 hectares were cultivated in the 1980s (Tapia *et al.*, 1988). This is consistent with the idea that cultivation of native crops declined with the introduction of Old World crops (e.g. broad bean, *Vicia faba*) by the Spanish.

Care should be taken with the interpretation of seed recovered from archaeological sites, even from lower levels, because it can have a number of origins both modern and historic (Minnis, 1981). Distinguishing between the two is not always easy but a number of pointers are useful. These include seed morphology, which may differ between cultivated and wild species, age, quantity, distribution and depth within site, seed composition, which may differ significantly from non-archaeological sites, and the nature of remains (Keepax, 1977). Similar problems exist for other plant remains (Wing, 1986). Charred seeds are often considered historic whereas uncharred seeds are more questionable (Minnis, 1981). Pearsall (1988) listed five possible sources of charred seed in archaeological sites which include deliberate introductions for food or fuel (within dung or earth sods) and accidental or incidental entry. Bean seeds are thought to be under-represented in the archaeological record as they rarely carbonise completely (Matthew Sayre, personal communication).

Little pollen data are available for lupins from archaeological sites or pollen cores. This is in part because *Lupinus* pollen is not readily distinguished from pollen of other Papilionoid legumes (Adams & Smith, 1977) and also because lupins are insect-pollinated. However, *L. mutabilis* pollen has been identified from coprolites of Moche skeletons at the coastal Dos Cabezas Moche site dated at 1,900-1,300 BP (Geyer *et al.*, 2003). The identification is based on strong similarities with ancient Chilean pollen records (Patrick Geyer, pers. comm.). Whether wild or domesticated lupin, pollen presence in coastal areas further suggests human interaction or movement between coastal and mountain environments but does not provide strong support for lupin cultivation. This is particularly striking since lupins are insect pollinated and little pollen is likely to be widely distributed in the environment.

Illustrations of numerous crops have been identified on prehistoric pottery remains (http://www-rohan.sdsu.edu/~bharley/images/WWWAgentsPlants.jpg) (Yacovleff & Herrera, 1934). These early depictions are essentially caricatures in terms of the emphasis placed on the utilised part of the crop. While the identification of some of these is convincing e.g. masuha (*Tropaeolum tuberosum*), oca (*Oxalis tuberosa*) and tuna (*Opuntia* sp.) (Figs. 28c, d and Fig. 39 e in Yacovleff & Herrera, 1934), others are more suspect. Two illustrations on Pacheco urns found in the coastal Nazca valleys dated at 1,250 – 1,200 BP (during the Tiahuanaco culture) have been interpreted as representing *L. mutabilis* (Figs. 4.3b and c) (Figs. 27a, b in Yacovleff & Herrera, 1934). The exact source of the illustrations is unknown and may be from the ceramics illustrated in Figures 4.3a and d. However, the illustrations (Figs. 4.3b and c) and those on the ceramics (Figs. 4.3a and d) are not identical suggesting the source material for Yacovleff and Herrera's drawings may have been other pieces of pottery as ceramics illustrating cultivated plants are relatively common.

Given the relative accuracy of other crop representations and the fact that the large seeds are the most important trait of the modern Tarwi crop, none of these illustrations provide compelling evidence for *L. mutabilis* cultivation in southern Peru. However, neither are

these a convincing match for any other Andean crop known today, although the axillary structures in Fig. 4.3c are reminiscent of *Oxalis* leaves and the long narrow structures superficially resemble chile peppers (Colin Hughes pers. comm.). It is even difficult to be sure whether the structures shown present leaves, flowers or fruit. Both pictures show two different above-ground features, one arising in the shoot axis. The narrow structures terminating the shoots have been interpreted as pods (Yacovleff & Herrera, 1934). These structures are filled with vertical lines which may represent constrictions or septae between seeds. However, then there are far too many for modern Tarwi which has 3 – 7 seeds per pod but might be consistent with the suggestion that at first the immature pods were eaten rather than the seeds (Antúnez de Mayolo, 1982). However, there is evidence of the seeds in cultivation (see above) by 1,800 BP and I would thus expect some emphasis on the actual seeds in the illustrations. Furthermore, much more convincing legume pods are depicted on other pottery items from a similar period (http://www-rohan.sdsu.edu/~bharley/images/WWWPlantConchopataA.jpg and http://www-rohan.sdsu.edu/~bharley/images/WWWPlantYandHLegumes.jpg). Further discrepancies are the angle of the pods which in both illustrations are held horizontal whereas *L. mutabilis* pods are held upright and the root morphology as lupins have long tap roots.

If the narrow terminal structures are accepted as pods, the axillary structures must be viewed as leaves or flowers. However, they not only have the wrong position but are also the wrong shape and number for either. One striking feature of lupins in comparison to other cultivated plants from the Andes is their distinctive palmately digitate leaves with 5-14 narrow oblong, obovate or lanceolate leaflets. The larger structures in Fig. 4.3b could be interpreted as digitate leaves but leaflet number is too few and the axillary structures certainly do not resemble lupin flowers or fruits. The upright habit is consistent with *L. mutabilis*.

In summary, archaeological evidence suggests that *L. mutabilis* was domesticated sometime between 3500 - 1,800 BP (Fig 4.2). The uncertainty surrounding whether the Cotocollao

seeds are wild or not, as no size or abundance data are available, accounts for this wide time span. Seeds dated earlier than 3,500 BP are too small to be domesticated but could represent wild lupins perhaps used as a forage food as early as 10,000 BP (Fig 4.2). More precise age estimates can only be achieved through new archaeological investigations which discover plant remains which can not only be identified as *Lupinus* but more specifically can be demonstrated to be domesticated. Speculatively I would suggest it is likely that domestication of *L. mutabilis* occurred towards the beginning of this period since this coincides with the appearance of a wide range of Andean crops in the archaeological record (Hastorf, 1999).

4.7 Previous hypotheses of the origin of Tarwi, *L. mutabilis*

Two hypotheses have been proposed for the origin of tarwi, *L. mutabilis* (Blanco Galdes, 1986; Kazimierski & Nowacki, 1961; Tapia & Vargas, 1982). These are described below. Neither is based on convincing evidence and neither can be accepted without further investigation.

Kazimierski and Nowacki (1961) proposed that *L. mutabilis* was a hybrid between two North American species, *L. douglasii* and *L. ornatus*. They highlighted the fact that both *L. mutabilis* and *L. albus* conform to the pulse domestication syndrome. They listed characters which hypothetical progenitors would be expected to show and used these as criteria to select wild species as potential progenitors. On this basis they identified a group of American species, but listed just two, *L. douglasii* and *L. ornatus*. They reinforced this hypothesis with similarities in alkaloid composition, chromosome number, leaflet number and shape, and number of seeds per pod. However, sampling of all these features was sparse and no other New World species were presented for comparison. They reported that seeds of *L. douglasii* were eaten by North American Indians and speculated that these could have been transported by humans to South America. Moreover, they refer to *L. douglasi* throughout the paper whereas the accepted name is *L. douglasii.*

Kazimierski and Nowacki's (1961) hybrid proposal is flawed in all respects. The characters utilised to identify the hypothetical progenitors listed are not specific i.e. slow growth, coloured seeds, possession of anthocyanins, small seeds, bifid calyx, leaf morphology, chromosome number and could apply to almost all lupins in Section Occidentalis (see Chapter 7) – ca. 250 species. The statement "Hypothetical initial forms have originated in the eastern regions of North America." is made with no firm evidence and must be taken as highly speculative. Gross (1982) referred to this hypothesis as 'daring'. I disagree with the statement that most of the cultivated species take their origin beyond the natural range of the wild form as wild species with distributions equal or greater to cultivated species are often candidate progenitors (Vavilov & Dorofeev, 1992). Observation of *L. douglasii and L. ornatus* show that on the basis of morphology they are not any closer to *L. mutablis* than most Andean species and even less so than a subset of species.

Wild tarwi plants have been reported from southern Peru, around Cusco (Blanco Galdes, 1982, 1984, 1986; Tapia & Vargas, 1982). These specimens were described as morphologically similar to *L. mutabilis* and able to successfully cross with cultivated material (Blanco Galdes, 1982). The wild plants differ from *L. mutabilis* in having smaller leaves, narrower leaflets, strong anthocyanin pigmentation, smaller (5-7 mm) black or marbled seeds (Blanco Galdes, 1982) and high outcrossing rates (Blanco Galdes, 1984). Tapia and Vargas (1982) also commented that *L. praestabilis* from the Cusco area is "very similar to tarwi" (as did Smith (1948)), but in the same paper list three species, *L. condensiflorus, L. malacotrichus* and *L. paniculatus,* which also bear a strong morphological similarity to *L. mutabilis.* It is unclear whether Tapia and Vargas (1982) and Blanco (1982) are discussing the same material. In the same year Gross (1982) commented that no wild tarwi have been identified. Tapia and Vargas (1982) speculated that the origin of tarwi is likely to have been as a spontaneous mutant or variety of one of the wild species. One further complication is the occasional use of the term tarwi to refer to wild lupin species.

Smith (1951) grouped some species into an informal group which he termed the *mutabiloids* but did not define the characters these species share. The mutabiloids are *L. eanophyllus* C.P. Sm., *L. falsomutabilis* C.P. Sm., *L. protrusus* and *L. tayacajensis* C.P.Sm. (Smith, 1953). He commented on several species, *L. falsomutabilis* (Smith, 1953) *L. haughtianus* C.P.Sm. (Smith, 1944a), and *L. tayacajensis* (Smith, 1953), which he considered to be closely related to *L. mutabilis* because they are morphologically similar in some respect. Detailed examination of these protologues and associated type material shows that they do not share informative morphological characters with *L. mutabilis* (see species account of *L. mutabilis* for greater discussion).

4.8 Methods

In order to test these hypotheses and further investigate the origin of *L. mutabilis* I have integrated molecular phylogenetic data (Chapter 3) with a morphological survey of all ca. 85 Andean species (Chapter 5). I have identified a small set of species which are morphologically similar to *L. mutabilis* in characters likely to be neutral with respect to domestication. These were vegetative characters such as leaf shape and indumentum. Seed size data (Chapter 5) were also considered on the basis that the progenitor was more likely to have been a large-seeded wild species than a small-seeded species. This idea that large seeded species could be pre-disposed to domestication derives from the observation that the progenitors of the major domesticated grasses are amongst the largest seeded grass species (Blume in Diamond, 1997). Similarly, the progenitors of Old World domesticated lupins are amongst the largest seeded species within *Lupinus.* Phylogeny data were used to test hypotheses of relationships based on morphological data but this aim was hampered by the low levels of sequence divergence found amongst the Andean species, attributable to exceptionally rapid species diversification (Chapter 3).

Measurements of weight, length, width and thickness of a sample of 25 seeds per accession were taken. Variation in seed shape in each sample was recorded along with colour and colouration pattern. Descriptive terms followed The International Board for Plant Genetic

Resources (IBPGR) "Lupin Descriptors" (1981) and the Systematics Association glossary of simple symmetrical plane shapes chart (1962). Appendix 13 lists the accessions utilised in this study along with voucher information. Additional seed weight data were retrieved from the RBG Kew seed information database (SID-http://www.rbgkew.org.uk/data/sid/) and incorporated into the dataset used in this study.

4.9 Results and discussion

4.9.1 Seed size, shape and colour

The few detailed investigations of *Lupinus* seeds in particular species or groups indicate that seed morphology can be useful for both species delimitation and an important feature in studies of character evolution (e.g. domestication) (Aniszewski *et al.*, 2001; Halpern, 2005; Heyn & Herrnstadt, 1977). Due to the agricultural importance of the Old World species, the seed morphology of this group is well-known (Gladstones, 1974). As a pulse crop the weight of seeds is important and ranges of seed weight of the Old World species and *L. mutabilis* have been recorded in detail (Cowling *et al.*, 1998). *Lupinus albus* has the heaviest seeds weighing up to 77g per 100 seeds. Differences in seed size and weight have also been reported between the wild and domesticated Old World species (Cowling *et al.*, 1998; Gladstones, 1974). A detailed study of *L. polyphyllus* seeds investigated the number, size and diversity (Aniszewski *et al.*, 2001), revealing distinct seed colour groups that could be used to revise delimitation of the *L. polyphyllus* species alliance. However, up to now data on seed morphology of the vast majority of *Lupinus* species has been lacking.

Lupinus seeds exhibit a wide diversity of sizes, shapes and textures (Figs. 4.4 - 4.13) all of which are also variable within species. These traits show continuous variation across the genus as a whole (apart from some domesticates) (Figs. 4.4 – 4.13) and are not amenable to anything but arbitrary partitioning into character states. Seed colour and pattern are not straightforward to partition into discrete states either because most seeds have multiple colours and these are variable within individuals and within species. Similarly, secondary

markings such as 'crescents', 'eyebrows' or 'moustaches' are also variable within species. Presence of an aril rim (an extension of the aril which protrudes from the hilar region) (Fig. 4.5) is notable in several lowland South American species whilst all other species observed lack an aril rim. Some seeds are occasionally attached to funicular fragments and appear superficially to have an extended aril rim. However, these fragments are not part of the seed and easily detach. More extensive sampling of the lowland South American species is needed to assess the whether the aril rim are present in all members of this clade.

Seed weight in the Andean domesticate *L. mutabilis* has increased at least six fold from the putative progenitor *L. piurensis* (see below) and is over twice that of the heaviest wild Andean seeds measured so far (Fig. 4.15). Gross (1982) stated that these large seeds are a testament to the ancient cultivation activities of the South American inhabitants. It has been shown that the wild ancestors of barley and wheat are amongst the thirteen largest seeded species within the Poaceae (Mark Blumer in Diamond, 1997) which strongly suggests that large-seeded grass species were 'pre-disposed' to domestication. Following the same principle and given that there is considerable diversity in seed size among Andean *Lupinus* species the progenitor of *L. mutabilis* more likely to have been amongst the larger seeded species. The large seed size of the Old World species may also explain domestication of several Old World lupins.

Although seed colour is variable within *L. mutabilis,* and even pure black seeds are occasionally seen (Fig. 4.9a), by far the most common colour is white/cream (Fig. 4.9o). All *L. mutabilis* seeds appear to have high alkaloid concentrations indicating that pigmentation is not linked with a reduction in alkaloids as it is in some other legume species (Dübbern de Souza & Marcos-Filho, 2001). Neither is the pigmentation change linked to increased seed coat permeability as seeds of all colours germinate equally rapidly. This would suggest that there was independent selection for pale seeds and high seed coat permeability and that the two factors are not linked in *Lupinus*. This is reinforced by the observation that the palisade cells of soft-seeded varieties of *L. angustifolius* have grooves in their tangential walls whereas

hard-seeded varieties are entire (Valenti *et al.*, 1989). Gross (1982) suggests the pale colour would have conferred a selective benefit for food use, storage and cultivation to allow quick recognition of infected or poor quality seed.

Growth of *L. mutabilis* is much faster than any of the other wild species for which data are available and the cotyledons are much larger, in line with large seed size. Since a negative correlation between seed size and seedling growth has widely been observed (Westoby & Swanborough, 1996) it appears that there has been separate selection for large seeds and rapid growth. This is reinforced by the observation that the first *L. mutabilis* eophylls possess five or more leaflets whereas many Andean species possess three indicating some level of heterchronic peramorphosis in the form of earlier development of more leaflets (Li & Johnston, 2000).

4.9.2 Domestication and origin of *L. mutabilis*

Four species, *L. ellsworthianus, L, piurensis, L. praestabilis* and *L. semperflorens* C.P. Sm., from the ca. 85 Andean species were identified as morphologically similar to *L. mutabilis* (see Appendix 1 for a detailed taxonomic account of these species). The latter three are highlighted as possible progenitors of *L. mutabilis.* They share **glaucous** stems which are most pronounced in young shoots, **thin leaflets** with a **glabrous adaxial** surface that is only occasionally sparsely hairy, **mucronate, oblanceolate** rarely narrow-linear or elliptic leaflets and **large flowers** (Fig. 4.14 and taxonomic account in Appendix 1). Interestingly, the additional resolution achieved through Bayesian analysis (Fig. 3.10) to some extent corroborates these species as putative close relative of *L. mutabilis*. In the ITS tree (Fig. 3.5, Chapter 3) a well-supported clade is found comprising *L. piurensis* and accessions of *L. semperflorens* and accessions of *L. mutabilis, L. piurensis* and *L. semperflorens* also form a clade in the GPAT1 tree (Fig. 3.7, Chapter 3). The 3-locus tree (Fig. 3.10, Chapter 3) resolves a clade containing *L. mutabilis, L. piurensis* and *L. bangii.*

Of these three species, *L. piurensis,* and specifically *L. piurensis* var. *piurensis*, is the most morphologically similar Andean species to *L. mutabilis.* Thus, based on both morphology and tentative molecular results this species is tentatively suggested to be a putative wild progenitor of *L. mutabilis.* *Lupinus piurensis* is also among the largest seeded wild Andean species (Fig. 4.15) providing further evidence that this species may be the wild progenitor of *L. mutabilis*. Parsimony analysis does not refute this (Fig. 3.10, Chapter 3) and Bayesian analysis provides some positive evidence for this close relationship. A clade containing all *L. mutabilis,* one *L. piurensis* and one *L. bangii* accession supported by a posterior probability of 0.78 was found in the 3-locus tree (Fig. 3.16, Chapter 3). Although described in 1953 (Smith), *L. piurensis* has not previously been suggested as a progenitor of *L. mutabilis*. One reason for this may be its restricted distribution in northern Peru and southern Ecuador (Fig. A2.2). Expeditions looking for progenitors of *L. mutabilis* by Blanco (1986) focused on the department of Cusco. The current distribution of *L. piurensis* and other putatively closely related species, such as *L. semperflorens* (Fig. A3.2), suggest an origin of *L. mutabilis* in northern Peru. Crops and their progenitors are often considered to be conspecific (Emshwiller, 2006) but there is little consistency in the naming of domesticates. Given that this hypothesis remains tentative, that *L. mutabilis* is clearly distinguished from *L. piurensis* and to avoid unnecessary nomenclatural changes, *L. mutabilis* and *L. piurensis* are retained as distinct species (Appendix 1).

Although the additional Bayesian resolution within the Andean clade should clearly be viewed with caution, it is interesting that *L. bangii* is also found to be closely related to *L. mutabilis*. This species is an annual from southern Peru and northern Bolivia which has large fistulose inflorescences. One of the differences between *L. piurensis* and some *L. mutabilis* accessions is a large weakly fistulose peduncle. Natural hybrids between Andean species are not unknown and hybrids are often morphologically intermediate between their parents. A hybrid origin for *L. mutabilis* can not be discounted, but these species are presently allopatric.

Based on the sequence and timing of appearance of cultivated crops and presence of wild lupin seeds amongst the early archaeological plant remains from several Andean sites (Fig. 4.2) it seems probable that that wild lupin seeds were initially gathered to supplement a diet of meat, cultivated tubers, maize and beans. Harvesting of wild legume seeds to supplement cultivation of early crop domesticates has been documented from other areas such as Mesoamerica. There is a marked and more or less simultaneous increase in the number of crops appearing more or less simultaneously in the archaeological record between 5500 – 2500 BP possibly in response to increasing population and development of more complex societies. Thus it is possible that wild plants including lupins where brought into cultivation during this period. There is firm evidence of *L. mutabilis* domestication prior to 1,800 BP by which time this crop was abundant. It is possible that the changes associated with domestication, listed below, resulted from deliberate or unconscious selection during early cultivation. This could explain why domesticated lupins still have high alkaloid levels. The high Andes are a very harsh environment where survival would have been a challenge. The primary concern with cultivated plants would have been survival and fruiting ability rather than taste. I think this is reflected across the Andean domesticated crops which are generally rather bland or slightly bitter. Furthermore, the development of "sweet" Old World lupin varieties has taken considerable breeding effort including mutagenesis (Hondelmann, 1984). Reduction in alkaloid content is thus apparently difficult to achieve and would not be expected through unconscious selection alone. Blanco (1984), however, interpreted the high levels of bitter compounds as a signature of a recent domestication of *L. mutabilis*. Indehiscent pod mutants are known in various legumes so again I suspect this mutant was seized upon when it appeared but not actively sought after.

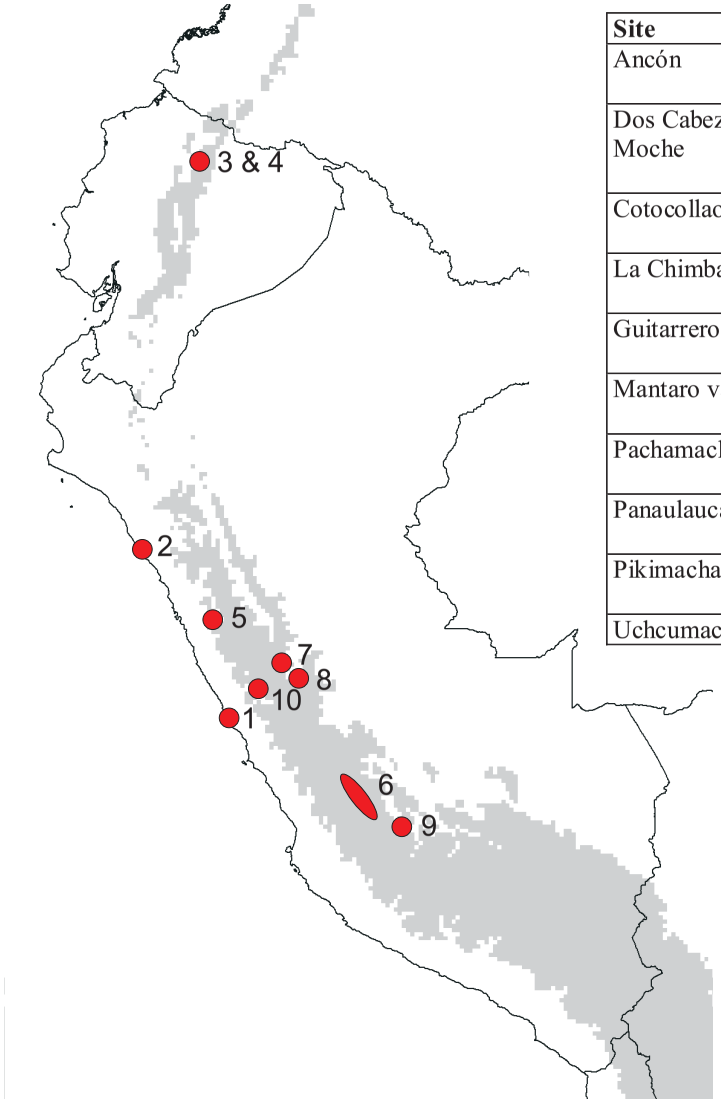
Comparison of the domesticate *L. mutabilis* with wild Andean species and specifically with the putative progenitor *L. piurensis* (Fig. 4.16) shows that morphological changes associated with the independent domestication of *Lupinus* in the Andes are similar to those observed with *Lupinus* domestication in the Old World, and that both conform to the general legume domestication syndrome described in Section 4.4. The most outstanding features

of both Old and New World domestication have been the dramatic **increase in seed size** and change from dehiscent to **indehiscent pods** (Gross, 1982). Other morphological changes associated with the domestication of *L. mutabilis* are: **reduced seed pigmentation**, **increased seed permeability** and **rapid and uniform germination and growth** (Gross, 1982).

The origin of *L. mutabilis* proposed here differs from previous hypotheses. Tapia and Vargas (1982) proposed *L. praestabilis* as the putative progenitor of *L. mutabilis*. While *L. praestabilis* may indeed be related to *L. mutabilis*, the available evidence suggests that it is not the wild progenitor, as *L. piurensis* shows greater morphological and genetic similarity. Their reasons for proposing *L. praestabilis*, include high levels of pigmentation, fertile hybrids and smaller leaves, are unconvincing. *L. praestabilis* only accrues significant levels of pigmentation when under stress, similar to other lupins. Furthermore, the extremely large stipules of this species are highly distinctive and quite unlike those of *L. mutabilis*. Given the limited genetic divergence and uniform chromosome number among the Andean species it is not surprising that *L. praestabilis* could successfully hybridize with *L. mutabilis*. The ability to produce fertile crosses has been used to infer progenitors in other species (Kochert *et al.*, 1996) but as the bulk of the Andean lupins are so closely related this attribute is unlikely to provide strong evidence. Furthermore, several successful crosses with North American species and even Old World species have been reported (Sawickia-Sienkiewicz & Brejdak, 1996). Leaves of *L. praestabilis* are larger, not smaller as described by Blanco Galdes (1986), than *L. mutabilis* and leaflets are narrower. Assuming Tapia and Vargas (1982) and Blanco Galdes (1986) are discussing the same species I disagree that *L. praestabilis* flowers and seeds are small. I agree that *L. praestabilis* seeds are between 5-7 mm but this size makes them amongst the largest wild Andean lupin seeds (Figs. 4.10-13 and 4.15). This seed range also includes two other species, *L. piurensis* and *L. semperflorens,* which show much closer morphological similarities to *L. mutabilis.* Tapia and Vargas' (1982) statement regarding other wild Andean species showing morphological similarities to *L. mutabilis* illustrates the complex morphology of this group.

4.10 Conclusion

Given that the Andean lupin radiation is the most spectacular example of explosive species diversification documented to date, it is not surprising that reconstructing a domestication event nested within it is has proved to be so challenging. Similar challenges have been faced in the investigation of another Andean domesticate oca (*Oxalis tuberosa*) (Emshwiller & Doyle, 1998, 2002) which is also a member of a large and taxonomically difficult high Andean genus. Additionally, although the Andean lupins are extremely diverse in terms of life history, most of the morphological characters used to delimit species are continuous. All characters are highly labile, subjective to code and thus unreliable to reconstruct relationships. However, morphological and genetic data have contributed to identifying allies of the domesticate *L. mutabilis* and a putative progenitor. The morphological differences observed between *Lupinus mutabilis* and its putative progenitor are: increased seed size, indehiscent pods, reduced seed pigmentation, increased seed permeability and rapid and uniform germination and growth.



Site	No.	Location	Coast/Sierra	Date	Reference
Ancón	1	Lima, Central Peru	Coast	5,300 BC -	(Towle, 1961)
Dos Cabezas Moche	2	La Libertad, northern Peru	Coast	100-700AD	(Geyer, Larson, and Stroik, 2003)
Cotocollao	3	Pichincha, northern Ecuador	Sierra	1500-500 BC	(Pearsall, 2003)
La Chimba	4	Pichincha, northern Ecuador	Sierra	2440 BC to 1500 BC	(Pearsall, 2003)
Guitarrero Cave	5	Callejón de Huaylas, Peru	Sierra	8000-7500 BC	(Lynch, 1980)
Mantaro valley	6	Mantaro valley, Peru	Sierra	1350 AD onwards	(Hastorf, 1993)
Pachamachay	7	Junín, Peru	Sierra	9850 BC - 80 AD	(Pearsall, 1980)
Panaulauca	8	Junín, Peru	Sierra	2100-1400 BC	(Pearsall, 1988)
Pikimachay	9	Ayacucho, Peru	Sierra	5800-1750 BC	(Rick, 1988)
Uchcumachay	10	Junín, Peru	Sierra	N/A	(Rick, 1988)

Fig. 4.1 Andean archaeological sites referred to in text. Shaded areas on the map represent areas above 3000m elevation.

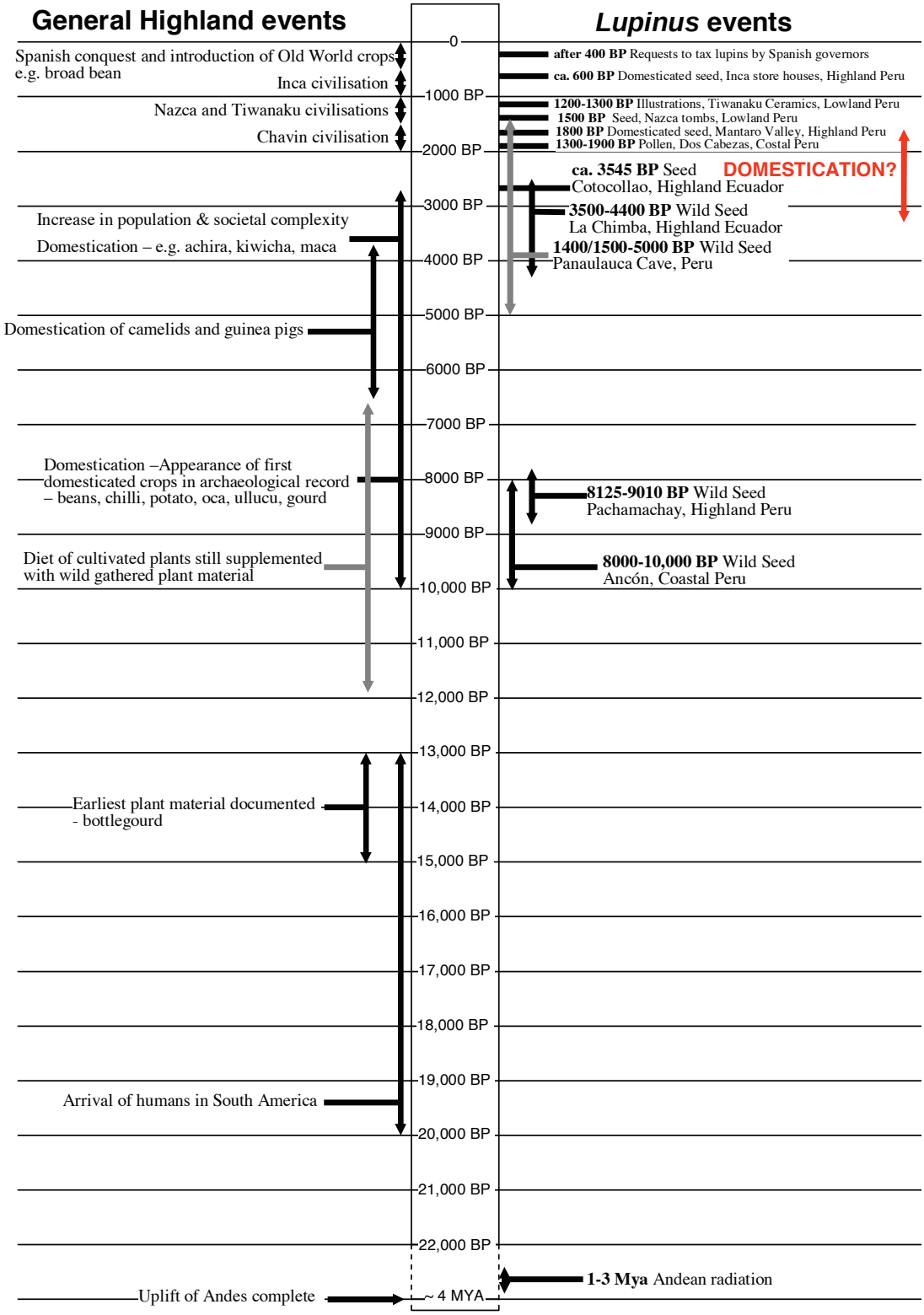


Fig. 4.2 Andean timeline showing *Lupinus* archaeological evidence and events. Approximate age estimates taken from the literature are discussed in the text. *Lupinus* domestication is estimated to have occurred between 3200-1800 years ago.

Fig. 4.3 Proposed prehistoric depictions of Tarwi, *L. mutabilis*, on ceramic artifacts. b. and c. are drawn by PJ Knobloch from Yacovleff and Herrera 1934-35:306, Fig.27a, b. The source of Yacovleff and Herrera's material is unknown. The pots shown here have similar illustrations upon them, circled in red. a. Ceramic pot with deity and edible plants, http://www-rohan.sdsu.edu/~bharley/images/WWWMINDICONDeityB.jpg, b-c Proposed Tarwi illustrations on Pacheco urns 750-800 AD, d. Ceramic bowl with feline-head handles and edible plants, Tiahuanaco-Huari 200-800 BC, National Museum of Archaeology, Anthropology and History of Peru, Lima, http://www.culturalexpeditions.com/huaripotlrg.jpg

Fig. 4.4 Seed size variation within *Lupinus.* From left to right *L. subacaulis* CEHH2319, *L. austrobicolor* CEH2001, *L. semperflorens* CEH2012 and CEH2032, *L. mutabilis* CEH2009. Scale lines at 1 mm intervals. Photograph RJE.

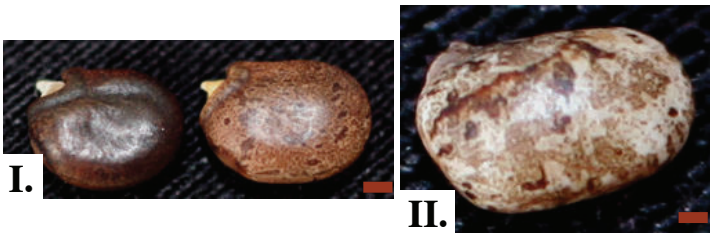


Fig. 4.5 Rim aril. I. Beaked rim aril present - *L. lanatus* s.n.30/12/02 II. Rim aril absent - *L. austrobicolor* CEH2001. Scale bars = 1 mm. Photographs RJE.

Fig. 4.6 Rough and Smooth seeds. Rough (a. and b.) and smooth (c.). a. *L. cosentinii* USDA s.n., b. *L. densiflorus* CD0331A02, c. *L. semperflorens* RJE221. Scale bars = 1 mm. Photographs RJE.

Fig. 4.7 Intraspecific seed colour and pattern variation. These three species illustrate intraspecific colour variation. A. *L.* sp CEH2099, B. *L. cumulicola* CEH1998, C. *L. truncatus* S & S Seeds. Scale lines at 1 mm intervals. Photographs RJE.

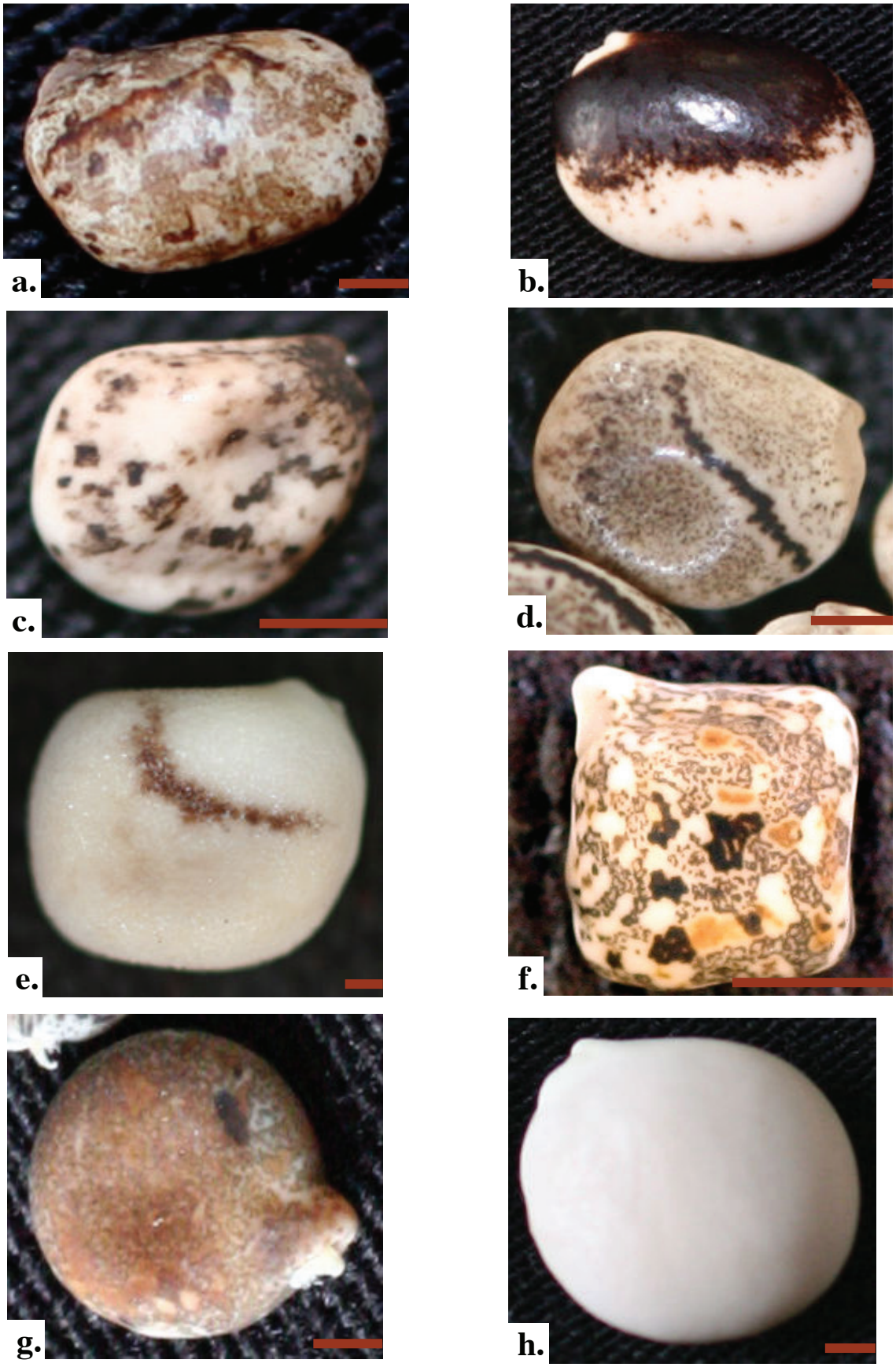


Fig. 4.8 Variation in *Lupinus* seed shape. a. Flattened oval *L. austrobicolor* CEH2001, b. Oval *L. mutabilis* CEH2326, c. Flattened oval *L. sp1* CEH2099, d. Flattened cuboid *L. huaronensis* CEH2241, e. Flattened cuboid *L. cosentinii USDA s.n.*, f. Flattened cuboid *L. havardii* USDA DLEG930123, g.

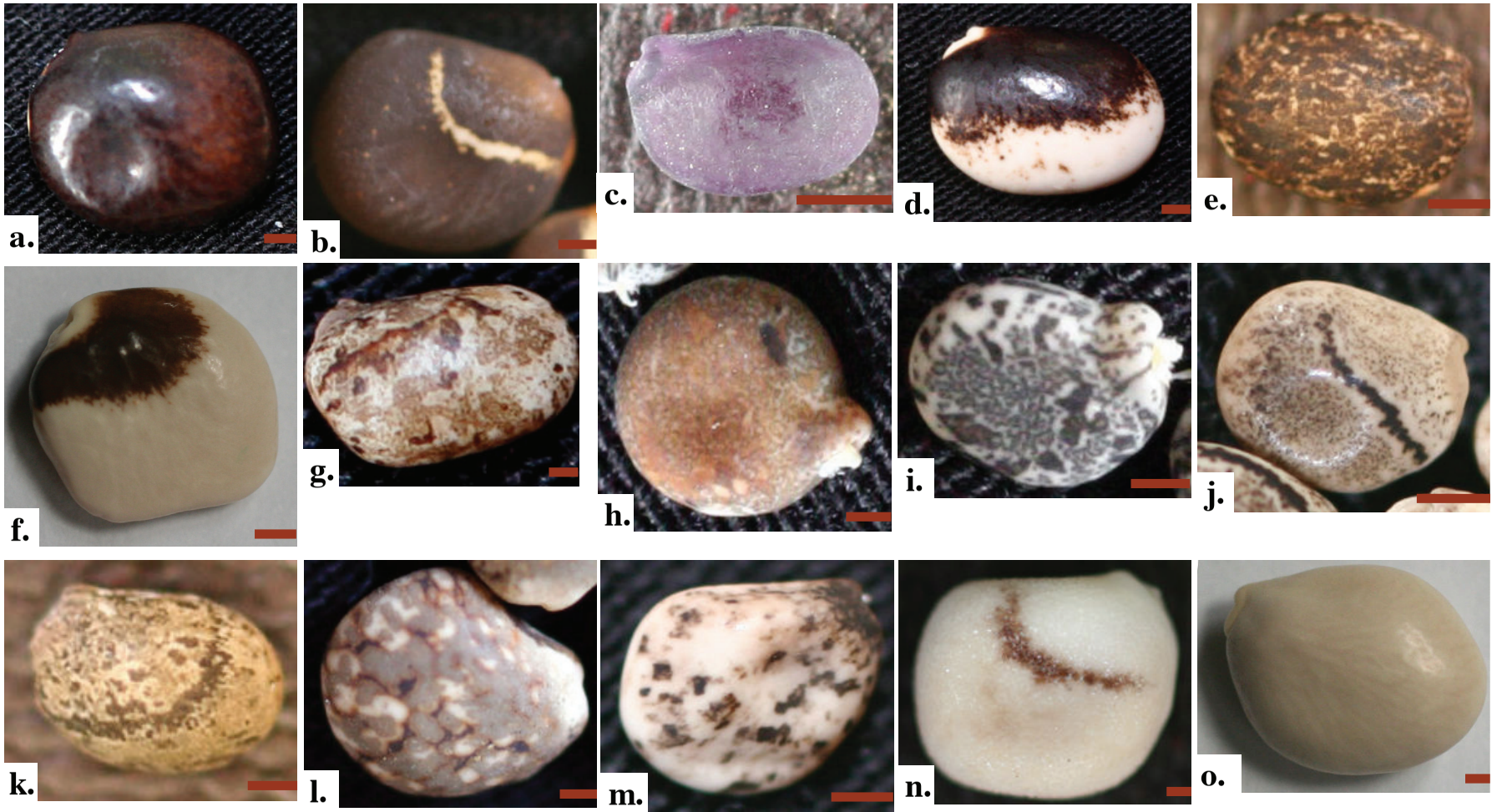


Fig. 4.9 *Lupinus* seed coat colour and pattern variation. a. Black *L. mutabilis* CEH2009, b. Dark brown with cream eyebrow *L. luteus* Lopez and Guerrero s.n.23/3/91, c. Purple *L. smithianus* RJE201, d. Black and white with white moustache *L. mutabilis* CEH2326, e. Marbled dark browns *L. semperflorens* RJE221, f. White and brown with cream moustache *L. mutabilis* P26961, g. Marbled cream and brown with faint brown eyebrow *L. austrobicolor* CEH2001, h. Marbled rusty browns *L. cumulicola* CEH1998, i. Marbled grey and cream *L. cumulicola* CEH1998, j. Marbled grey and cream with black eyebrow *L. huaronensis* CEH2241, k. Marbled light browns with mid brown eyebrow *L. semperflorens* RJE221, l. Marbled grey, brown and cream *L. angustifolius* Lopez s.n.19/3/91, m. Cream with brown mottles *L. sp* CEH2099, n. Cream with brown eyebrow - *L. cosentinii* USDA s.n. o. White *L. mutabilis* P28725. Scale bars = 1 mm. Photographs RJE.

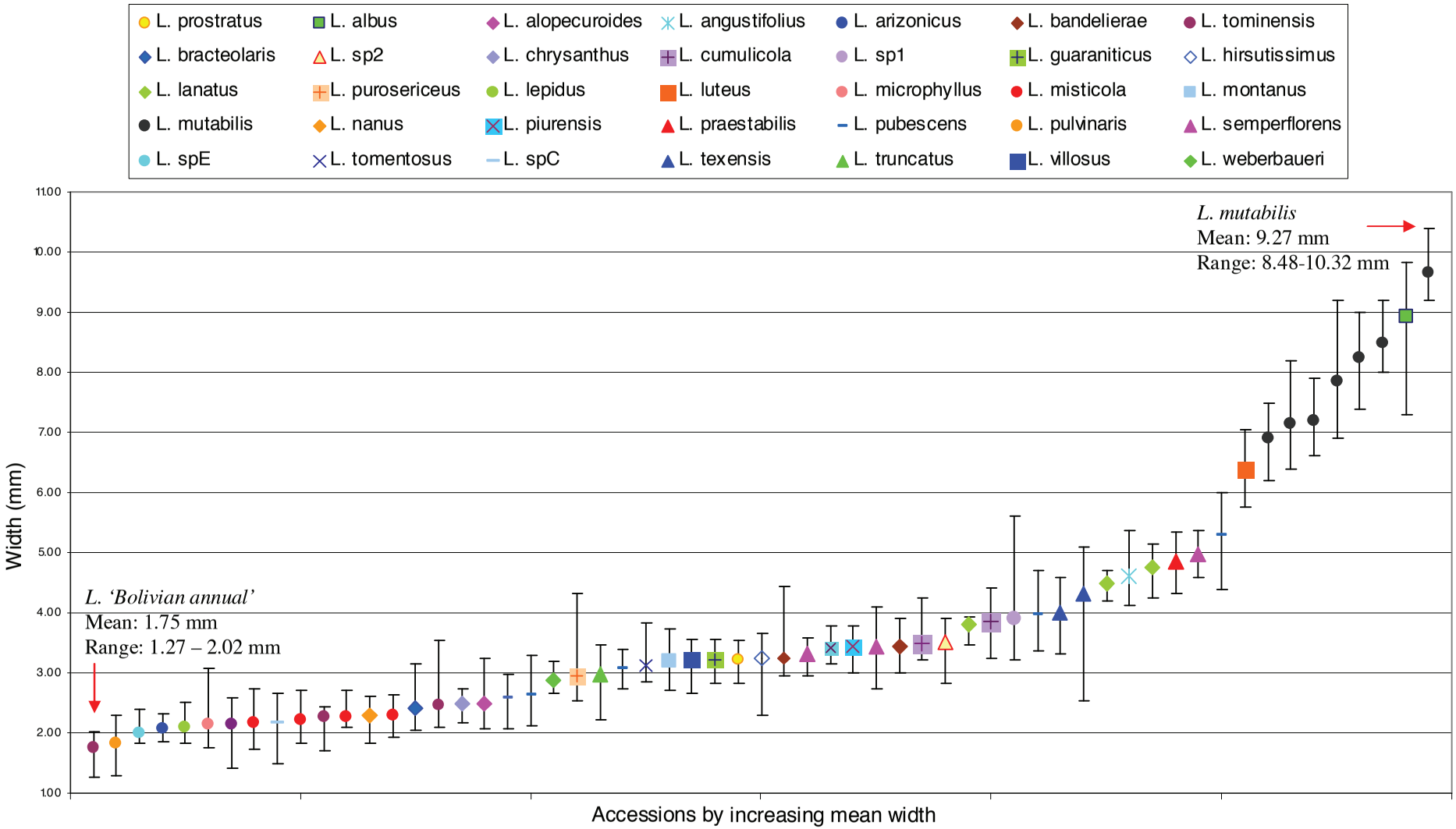


Fig. 4.10 Variation in seed width within *Lupinus*. Data points represent mean widths from a sample of 25 seeds. Range of widths is indicated by the y-bars. Where multiple accessions of the species were available these are represented by the same shape and colour data point. Data available in Appendix 13.

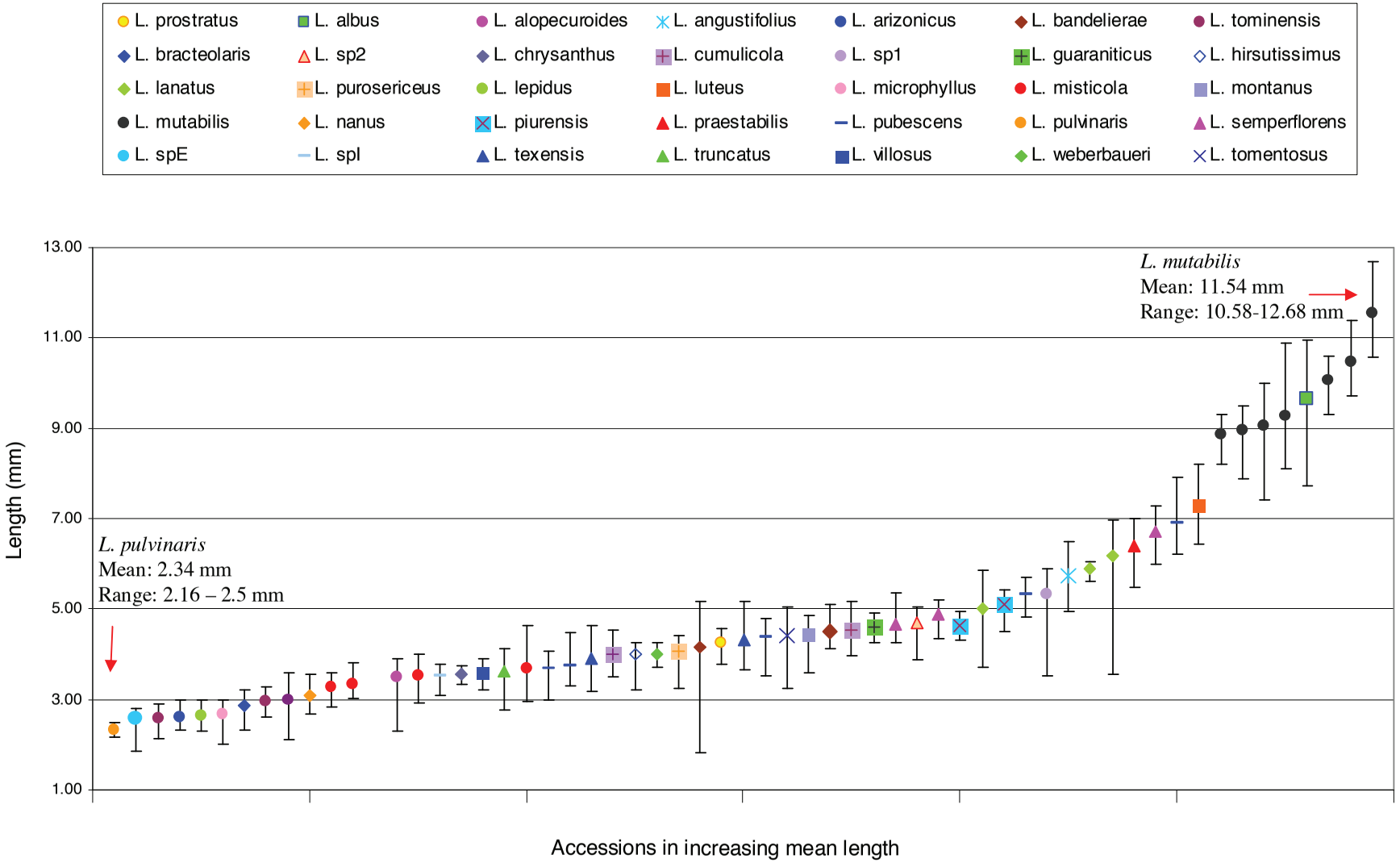


Fig. 4.11 Variation in seed length within *Lupinus*. Data points represent mean lengths from a sample of 25 seeds. Range of lengths is indicated by the y-bars. Where multiple accessions of the species were available these are represented by the same shape and colour data point. Data available in Appendix 13.

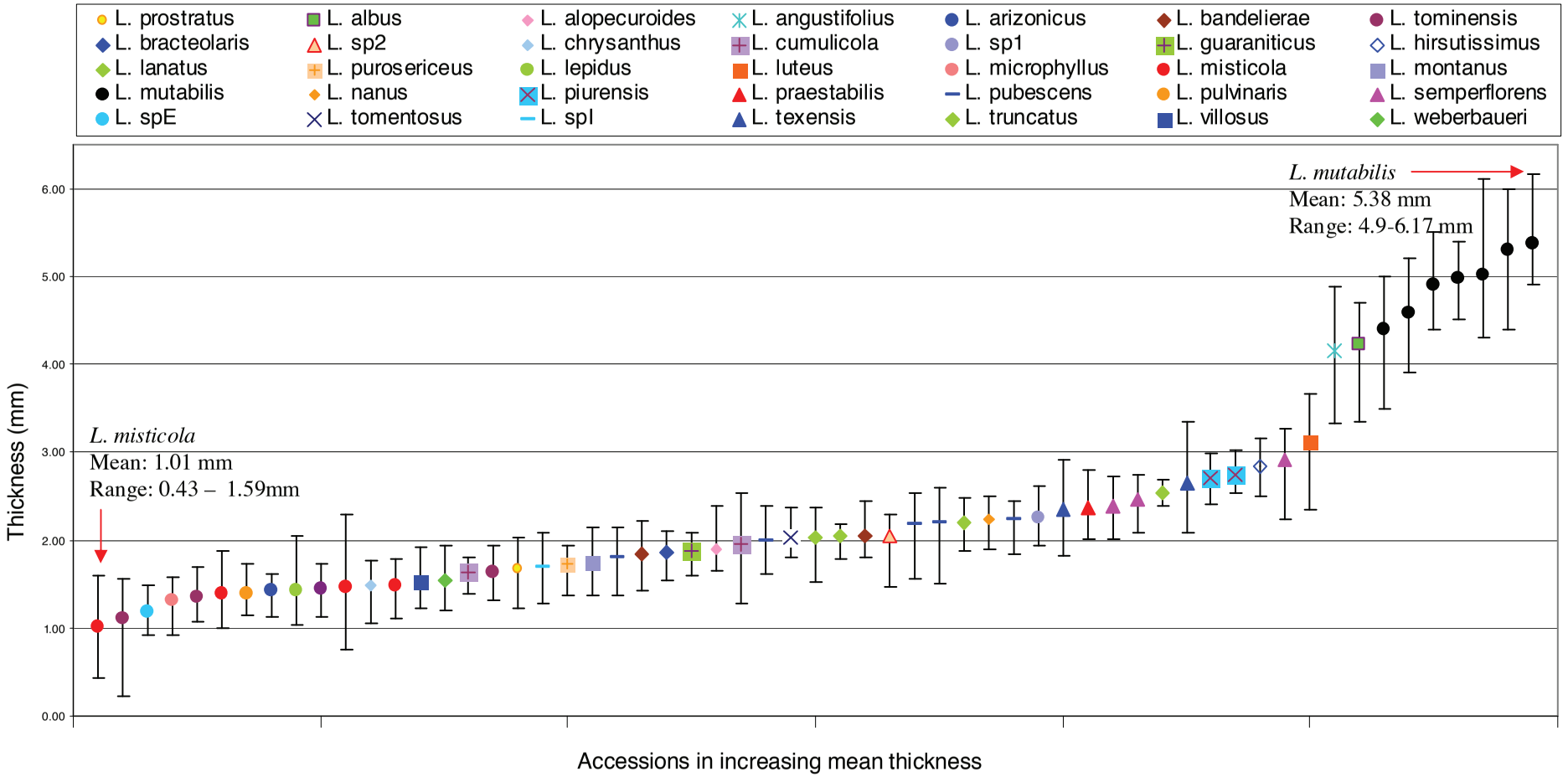


Fig. 4.12 Variation in seed thickness within *Lupinus*. Data points represent mean thickness from a sample of 25 seeds. Range of depths is indicated by the y-bars. Where multiple accessions of the species were available these are represented by the same shape and colour data point. Data available in Appendix 13.

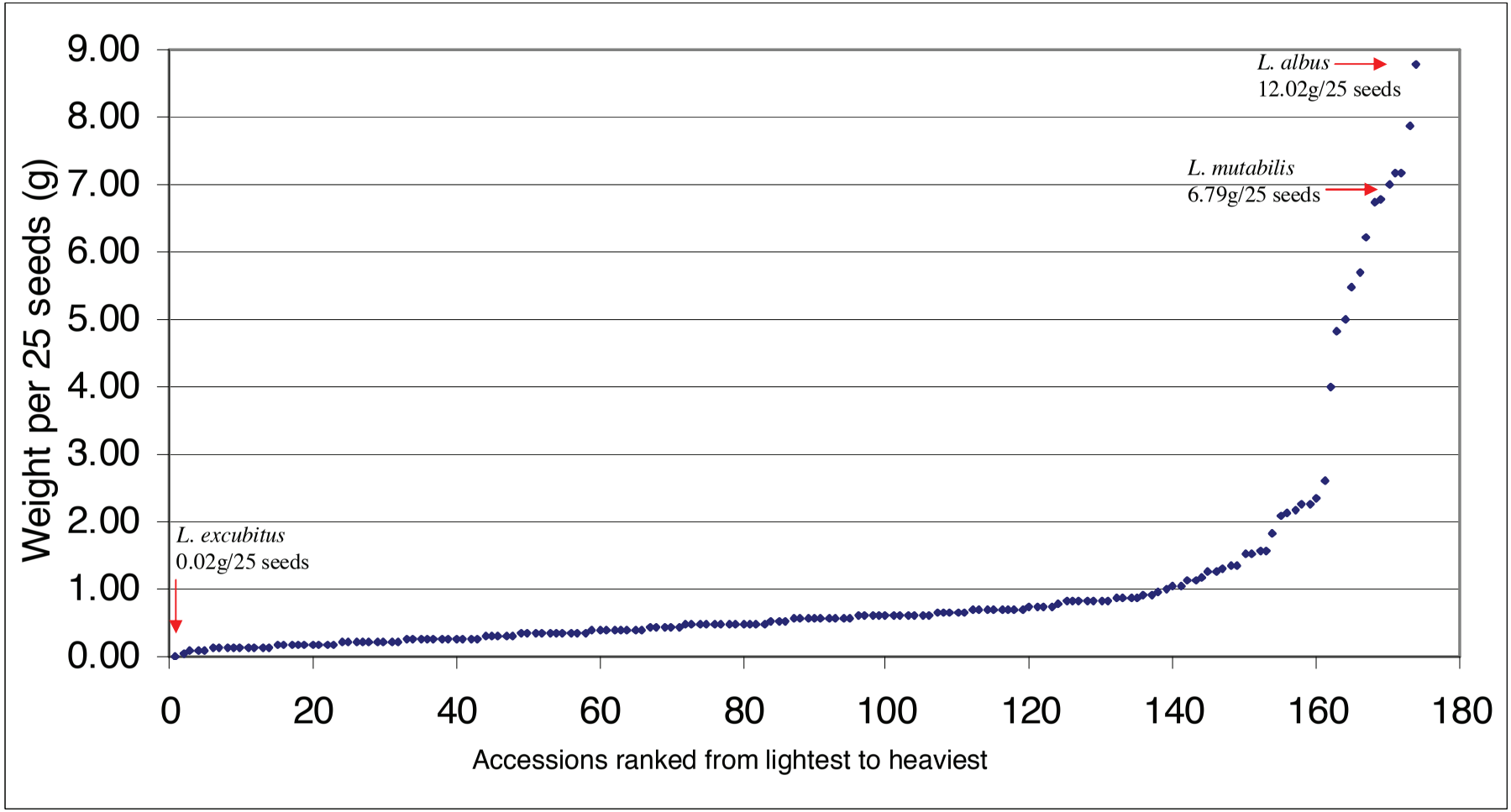


Fig. 4.13 Variation in seed weight within *Lupinus*. Data available in Appendix 13. Mean weights for 175 accessions representing (at least) 82 species. Sampling includes all sections of *Lupinus*.

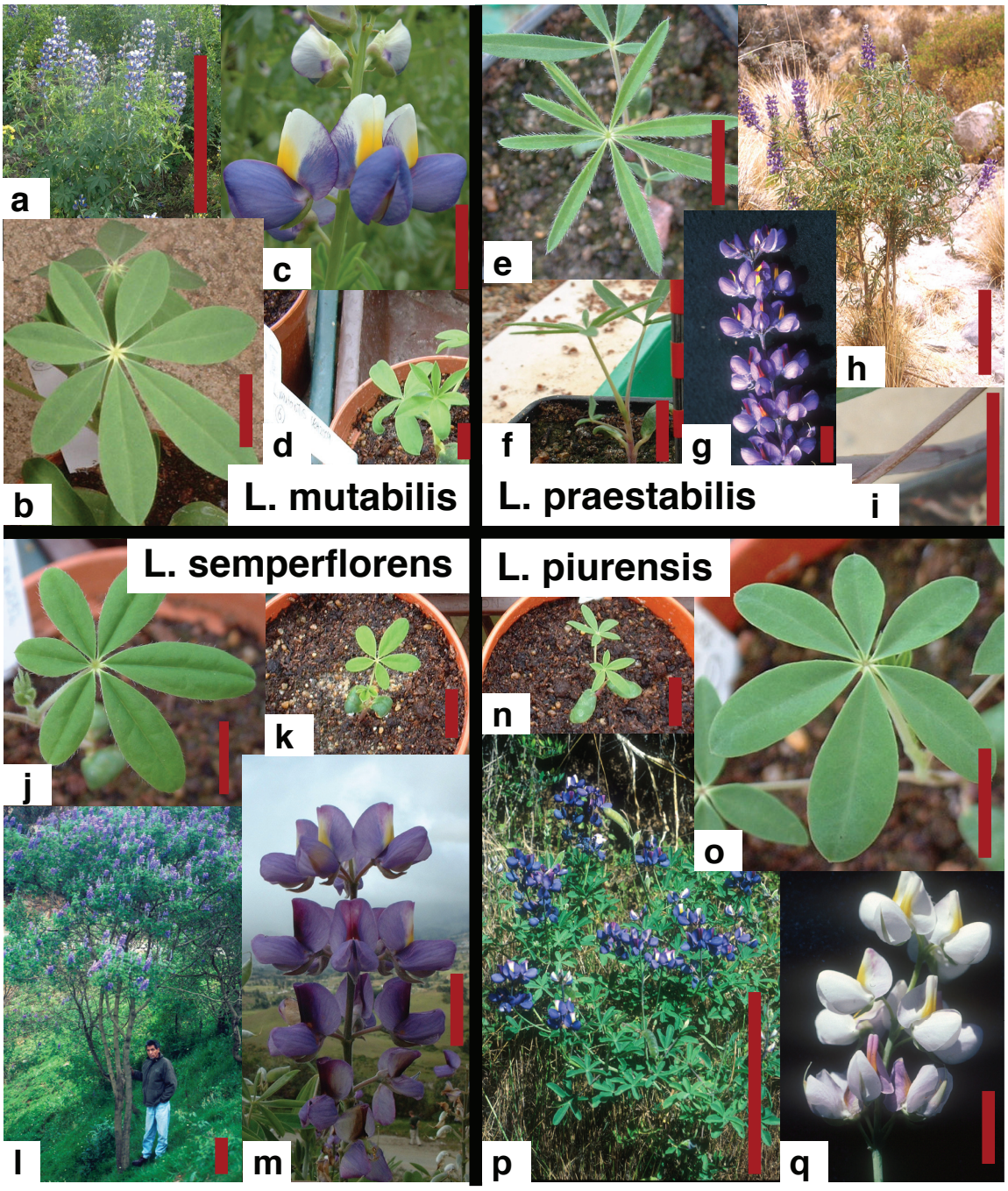


Fig. 4.14 *Lupinus mutabilis* and putative progenitors. a-d. *L. mutabilis,* e-i. *L. praestabilis,* j-m. *L. semperflorens,* n-q. *L. piurensis*. A. habit RJE56, b. leaf CEH2009, c. inflorescence Oxford Botanic Garden, d. seedling CEH2009, e. leaf RJE142, f. seedling RJE142, g. inflorescence CEH2353, h. habit CEH2353, i. glaucous stem RJE142, j. leaf CEH2032, k. seedling CEH2032, l. habit CEH2012, m. inflorescence RJE219, n. seedling CEH1997, o. leaf CEH1997, p. habit CEH1997, q. inflorescence RJE110. Scale bars 2 cm except 50 cm for a., h., l. and p. Photographs a-f., i-k., m-o. RJE and g., h., l., p. and q. CEH.

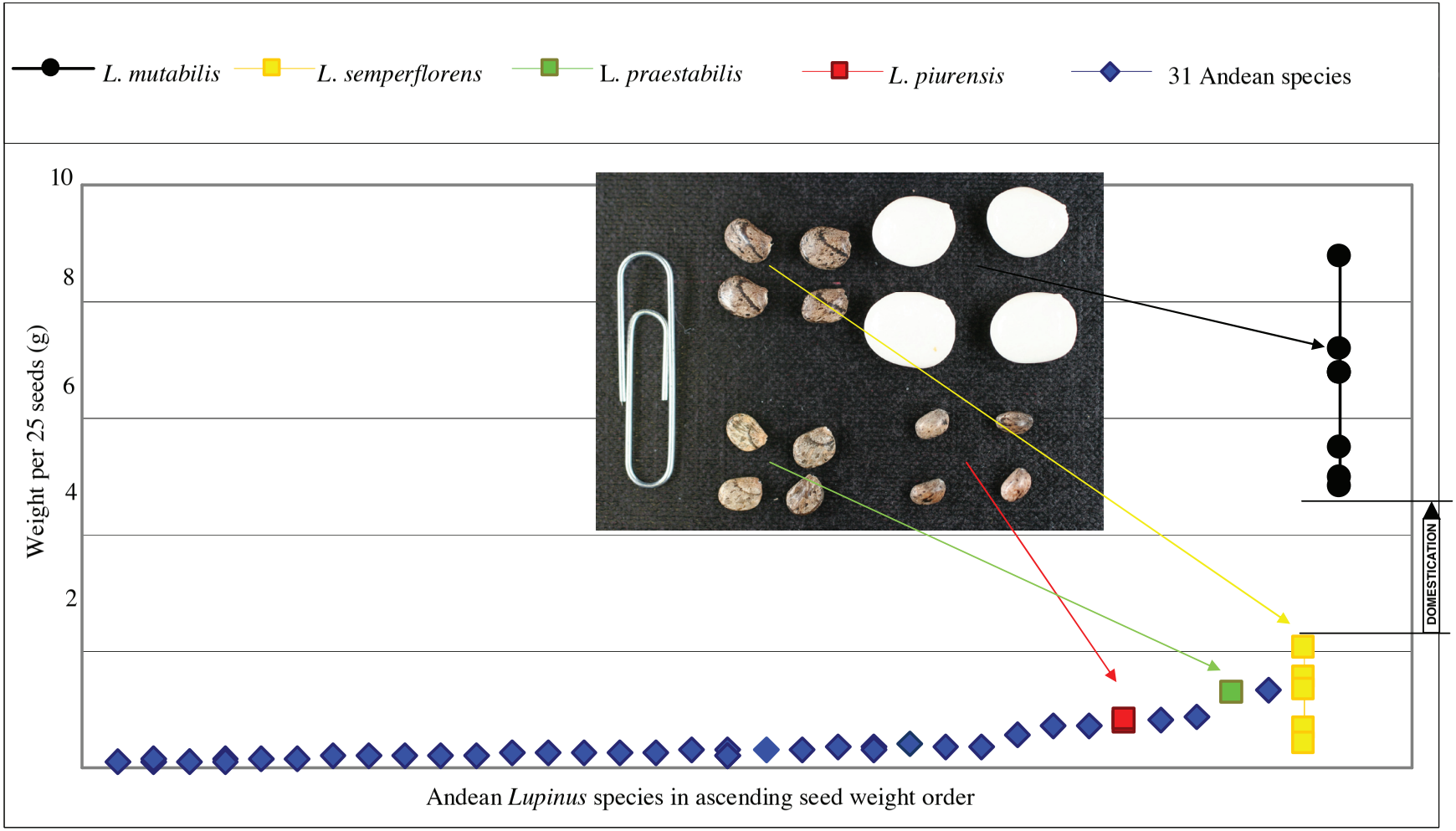


Fig. 4.15 Seed weight of 32 Andean species. *L. mutabilis* and allies are highlighted from other Andean species. Seed weight increase associated with domestication has been dramatic within the Andean domesticate *L, mutabilis*. The lightest seeds of *L. mutabilis* are 2.2 times heavier than the largest seeded wild species (*L. semperflorens*) and 6.6 times heavier than *L. piurensis,* a putative progenitor. Other heavy seeded species are *L. pubescens* and *L. tarus*.

Fig. 4.16 Morphological change through domestication. Comparison of the morphology of the domesticate *L. mutabilis* with wild Andean species**.** A. Indehiscent pods, *L. mutabilis* RJE166, b. dehiscent pods *L. semperflorens* CEH2032, c. Uniform and rapid germination, five days after planting, on the left hand side *L. pubescens* CEH1989, *L. bandelierae* CEH2321 and *L. sp* CEH2325 on the right handside *L. mutabilis* CEH2009 and CEH2326, d. large robust seedling five days after planting *L. mutabilis* CEH2009, e. rapid growth rate, 7 days after planting, *L. mutabilis* CEH2009, f. slow growth rate, seven days after planting, *L. piurensis* CEH1997, g. Increased seed size left hand side *L. piurensis* RJE113 right hand side *L. mutabilis* RJE151. Scale bars 20 mm.

5 Comparative morphology of *Lupinus*

5.1 Morphology versus molecules

Morphology is fundamental in systematics because for the great majority of organisms comparative morphological data have, at least up to now, provided the primary basis for delimitation, diagnosis and classification of species as well as higher taxa. Comparative morphological data continue to play a central role across many facets of systematics (Endress, 2002) including evolutionary studies, development, classification, identification, hybridisation and domestication.

The rapidly increasing availability of large molecular datasets which can provide much larger numbers of characters than morphology, the relative ease of homology assessment for molecular data (in the form of DNA sequence alignments) in comparison to morphology, the less subjective nature of character state definitions and the suggestion that molecular data are less prone to homoplasy than morphological data (but see Hillis & Wiens, 2000), have all fuelled recent debate over the role of morphological data in phylogeny reconstruction (Scotland *et al.*, 2003; Sytsma, 1990). Simulation studies have shown that the likelihood of reconstructing the correct tree increases with amount of data (Hillis, 1996; 1998). Scotland *et al.* (2003) argued that homologous morphological characters, which are unproblematic to partition and code, are just not available in the quantity needed for the reconstruction of well-resolved and robustly supported phylogenies. The problems surrounding use of morphological data in phylogeny reconstruction focus on disagreement about how morphological characters are defined and partitioned into character states and which types of characters are useful (Hawkins, 2000; Poe & Wiens, 2000). Qualitative characters are generally considered to be better than quantitative characters because character state partitions are less subjective (but see Poe & Wiens, 2000). Morphological data do have some advantages over molecular data (Hillis & Wiens, 2000) including: ease of broader taxon sampling, applicability to fossils and avoidance of the gene tree, species tree problem (Doyle, 1992) encountered with single molecular data sets. Clearly, both molecular and

morphological data are important. Phylogeny reconstruction is increasingly dependent on DNA sequences data yet the role of morphology in delimitation and classification remains crucial and central to systematics (Hillis & Wiens, 2000).

5.2 Homology Assessment

In order to use morphological data in a phylogenetic context, characters must be identified and partitioned into character states. Although Patterson (1982b; 1988) did not find the distinction between characters and character states useful, Hawkins *et al.* (1997) argued that this is vitally important, particularly for primary homology assessment. Both character and character states are implemented in this study. I am referring to characters as homologous attributes of organisms identified through a process of homology assessment. Defining morphological characters begins with a comparative morphological survey i.e. generating statements of primary homology, followed by legitimisation. Discovery of homology is critical because only homologous characters are meaningful to distinguish natural groups. Only homologous characters can form synapomorphies (de Pinna, 1991; Eldredge & Cracraft, 1980; Patterson, 1982b). There are numerous definitions of homology but most authors refer to homology as equivalence of parts due to shared evolutionary history (de Pinna, 1991).

Three tests or criteria must be met in order for a character or attribute to be considered homologous: congruence, conjunction and similarity (Patterson, 1982b; 1988). Rigorous homology assessment involves multiple steps (Brower & Schawaroch, 1996; de Pinna, 1991; Hawkins *et al.*, 1997). De Pinna (1991) distinguished primary and secondary homology, depending on the stage, generation or legitimation, of testing completed. Brower and Schawaroch (1996) break down de Pinna's primary homology assessment further into identifying similarity and identifying character states. The similarity and conjunction criteria are central in the formulation of primary homology statements. These tests form the basis of character definition and verification. As they are somewhat intuitive, particularly the

similarity criterion, and routinely applied by taxonomists, these steps are often overlooked in methodological discussions (Rieppel & Kearney, 2002).

Similarity was first viewed in terms of topographic correspondence, the structural correspondence under every variety of form and function of Owen (1843), but the concept has widened over time to include similarity in structure and transformation (ontogeny, fossil intermediates, morphoclines) (Hawkins *et al.*, 1997; Rieppel & Kearney, 2002). Topographic correspondence is often reported to be the most important of these because morphological similarities in terms of size, shape or function may be non-homologous and likewise morphological dissimilarities in terms of size, shape or function may be homologous (Jardine, 1969b; Rieppel & Kearney, 2002). While similarity has been described by some as a test of homology, it is better views as a criterion for proposing homology (Brower & Schawaroch, 1996; de Pinna, 1991).

The conjunction test is based on the logic that if two structures are found in the same organism then they cannot be homologous (de Pinna, 1991; Patterson, 1982b). This is to some extent linked with similarity and the topographic criterion but it is comparative within individuals rather than between them.

The congruence test is regarded as the strongest test of homology and under de Pinna's (1991) scheme is necessary to produce a statement of secondary homology. This test evaluates primary homologies under parsimony to assess their congruence with other characters i.e. the current phylogeny estimate. If a character is placed once on the tree providing a synapomorphy for a group, at the level of the group under investigation, it is congruent with the phylogeny and accepted as a secondary homology.

A special case of homology which is particularly relevant to plants is serial homology (homonomy sensu Patterson, 1982b). This accounts for repeated units within an individual e.g. flowers and leaves. These are rarely identical (Weston, 2000) and can be very different

e.g. male and female flowers or heteroblastic leaves. Plants are modular in their vegetative growth with each module comprising of stem, leaf and auxillary bud. One way to deal with serial homology is to take one module as the unit of comparison for the conjunction test and code alternative character states observed in different modules as polymorphic. This approach is adopted here.

There are a number of different types of non-homologous characters (Patterson, 1982b) which are of interest in studies of character evolution and adaptation. The most commonly encountered are parallelisms and convergences. Parallelisms pass the similarity test but fail the congruence test i.e. they appear more than once on the tree. Convergences also appear more than once on the tree but the characters have mistakenly been assumed similar. Often only re-examination highlights their non-similarity. Some patterns of character state distribution on a tree can be interpreted in multiple ways e.g. the possession of a homoplasious character state defining a nested taxon or clade could be equally explained by either one reversal or multiple independent origins of that state (Fig. 6.1, Chapter 6). Parsimony and homology assessment are used to infer whether the character state has been secondarily lost (reversal) or independently derived (convergence or parallelism).

5.3 Characters, character states and coding

It is recognised that not all characters are useful in phylogeny reconstruction or studies of character evolution. Reasons for excluding characters include: variation within terminal taxa, missing data, continuous/quantitatively coded variation, unknown polarity and high levels of homoplasy (Poe & Wiens, 2000). Different datasets can potentially lead to different conclusions which can not be compared unless character selection and coding methods are reported (Poe & Wiens, 2000). However, there are no guidelines or criteria for character selection/rejection and authors very often take different approaches (Hawkins, 2000; Poe & Wiens, 2000) or fail to detail their methods altogether (Hillis & Wiens, 2000). This discrepancy has led to calls for at best some guidelines, or at worst, explicit explanations of

methods and criteria used (Pimental & Riggins, 1987; Stevens, 1991) to allow comparisons of results, and make character assessment more rigorous and objective. Absence of explicit justification and discussion of decisions about characters seems to be a general failing in taxonomy that similarly pervades issues surrounding species delimitation.

Taking into consideration the kinds of characters that are useful for diagnosability of infrageneric groups, of interest in studies of character evolution, and which can be unambiguously partitioned into character states, a comparative morphological survey of *Lupinus* species was carried out. While it is acknowledged that many of the characters rejected for this study may be useful and have indeed been used for species delimitation where any characters showing fixed differences between species can be used the selection of characters for detailed study has been limited to those showing discrete, qualitative patterns of variation across species that are amenable to objective partitioning into character states. A number of characters such as banner shape and indumentum length which initially appeared to be discrete, were found to have an underlying continuous quantitative basis. Most of these were rejected because they showed continuous variation and could not be unambiguously partitioned into discrete character states. Leaf type was included because it can be unambiguously partitioned into discrete qualitative states, non-digitate and digitate. Characters such as leaf indumentum which were not fixed within individuals or species were also rejected with the exception of metaphyll type which, although polymorphic within individuals of *L. paraguariensis*, was considered for reasons of historical and classificatory interest. Flower colour was rejected because the same colour combinations are sporadically seen in many different species and the mutability of the banner spot post-pollination is difficult to record consistently for all species (Drummond, 2004). Characters thought to be unduly influenced by environment (e.g. indumentum), habit or age (e.g. persistence of floral bracts and pedicel length) were also rejected. Adoption of such a rigorous approach reduced a large pool of potential diversity of morphological characters to just four: cotyledon type, eophyll type, metaphyll type and seed coat surface pattern. Additional to these, chromosome number was also assessed.

Different phylogenetic results can result from differences in character coding (Hawkins, 2000; Hawkins *et al.*, 1997). Hawkins (2000) lists 17 non-conventional coding methods found within matrices published between 1995 and 1996 in two major systematic journals and two conference proceedings. Partitioning and coding characters into character states is not always straightforward, as there may be alternative coding strategies. Conventional coding follows the premise that character states are attributes of the same character which can be considered alternative forms (Platnick, 1979) which can be proposed as transformations of each other (de Pinna, 1991; 2000; Hawkins *et al.*, 1997). Within conventional coding, characters are independent and have multiple exclusive character states. Conventional coding has been followed in this study.

5.4 Morphological studies within *Lupinus*

No systematic genus-wide survey of the comparative morphology of *Lupinus* species has previously been carried out. This is primarily due to the large size of the genus but has additionally been hampered by the taxonomic chaos surrounding the New World species, the quantitative nature of most characters and their phenotypic plasticity, the fact that most species are delimited by unique combinations of characters and the lack of a thoroughly sampled, well-resolved and robustly supported phylogeny. A broad description of the morphology of the South American lupins was presented by Planchuelo (1984) but with no explicit definition of characters and character states. Although attempts to reconstruct the phylogeny of *Lupinus* through morphological data alone would be futile given lack of objectively definable discrete characters, morphological data can still be used in conjunction with a molecular phylogeny to investigate character evolution and diagnosability as a basis for infrageneric classification.

5.5 Methods

5.5.1 Herbarium material, field observations and greenhouse material

Data were collected by observation of herbarium specimens and plants in the field and greenhouse, or gathered from the literature. Every attempt was made to match taxon sampling of the morphological dataset with the species utilised in the 3-locus molecular phylogeny reconstruction.

Over 8,000 herbarium sheets of *Lupinus* from B, BM, CAS, CGE, CGE-B, E, F, FHO, G, G-DC, GH, ISC, K, M, MO, MOL, MSB, NMC, NY, OXF, QCA, QCNE, UB, UC, US, USM and USZ were examined. The majority of this material is South American but taxa from all Sections/Series within the genus have been observed. Detailed observations and measurements were made from a subset of the herbarium specimens.

A two-month collecting trip to the north and central Peruvian and Ecuadorian Andes was carried out in the summer of 2004. One hundred and twenty-two *Lupinus* collections were made. Duplicates have been deposited at FHO, K, MOL, USM and QCN. Over this period observations of plants in their native habitats and morphological variation in the field were made, providing valuable insights into field characters such as flower colour, stem thickness and habit, patterns of variation within and among populations and the likely extent of phenotypic plasticity.

Fifty-nine accessions of 42 species were grown in greenhouses in Oxford. Seed was collected by Chris Drummond, Colin Hughes, Silvia Miotto and Maria Pinheiro, the United States Department of Agriculture (USDA) and myself. Seeds were scarified with a scalpel and soaked overnight in cold water or for several hours in warm water, germinated on filter paper in a petri-dish and planted into a 50:50 John Innes Number one compost and sand mix when the radical had initiated. Plants were regularly observed and photographed. Sampling represents all the New World Sections/Series. Appendix 13 lists material grown in this study along with provenance and voucher information.

5.5.2 Scanning electron microscopy

A survey of the micromorphology of *Lupinus* seed coat surface patterns was carried out. Forty-six accessions (Appendix 13) spanning the major clades of the genus were chosen with reference to previous SEM studies (Ainouche *et al.*, 2004; Fuentes *et al.*, 1993; Heyn & Herrnstadt, 1977; Plitmann & Heyn, 1984) and the 3-locus gene tree (Fig. 3.10, Chapter 3).

Fragments of dry seed coats were removed and cut using a scalpel, fixed onto SEM stubs and coated with gold-palladium using the Sputter Coater E5000. The scanning electron microscope JSM-5510 linked to a PC was used to view and photograph samples at magnifications from x 200 to x 8,000. Critical point drying was tested for three samples but there was no noticeable increase in sample stability so this method was not used further.

5.5.3 Chromosome data

Chromosome count data was assembled from the literature (see Appendix 16). New counts for South American species collected by Colin Hughes and myself were included (Conterato & Schifino-Wittmann, 2006, and unpubl. data).

5.5.4 Characters and character states

Primary homology was based on published data and new observations by careful application of the similarity criterion in terms of topographic identity and structural, compositional and ontogenetic composition. The conjunction test is also considered notably in relation to leaves. Secondary homology of the character states was tested via character congruence of the combined morphological and molecular data matrix (sensu de Pinna, 1991). The results are discussed in Chapter 6 via optimisation of non-molecular character states onto the 3-locus gene tree.

5.6 Results and discussion

5.6.1 General

Five primary homologies were identified. Detailed descriptions of characters and character states and the full range of comparative morphological variation for these characters are presented below. The complete non-molecular data for those taxa sampled in the 3-locus DNA sequence-based phylogeny is presented in Appendix 17. Characters and character states are shown in Table 5.1.

Table 5.1 Morphological and cytological characters and character states

Character → ↓ Character state	Metaphyll type	Eophyll type	Seed coat surface pattern	Cotyledon	Chromosome number
State 0	Digitate	Digitate	Granular	Petiolate	2n=24
State 1	Non-digitate	Non-digitate	Papillose	Connate	2n=30
State 2			Fascicular		2n=32
State 3			Simple-foveolate		2n=34
State 4			Rugulate		2n=36
State 5			Rugulate and Protuberant		2n=40
State 6			Substriate simple-foveolate		2n=42
State 7			Ocellate		2n=48
State 8					2n=50
State 9					2n=52

5.6.2 Metaphylls

There is huge diversity of leaf form within the angiosperms (Prance & Sandved, 1985). The most prominent and fundamental division of leaf types is between simple and compound leaves. These are defined by position of the blade(s) relative to the axillary bud (Fig 5.1) (Bentham, 1859). However, distinguishing between simple leaves and 'compound' leaves with a single leaflet (unifoliolate) may not always be straightforward. In legumes simple and unifoliolate leaves have generally been distinguished morphologically or anatomically by the presence of a pulvinus (often referred to as a pulvinule in this position) at the base of a single blade in unifoliolate leaves and absence of a pulvinus at the base of the blade in simple leaves (Fig. 5.1) (Bentham, 1859; Cronquist, 1988; Polhill, 1976). Pulvini are cushion-shaped swellings at the base of the petiole/petiolule (Jackson, 1928). In legumes

the pulvinus provides a point of articulation allowing movement of the leaf or leaflet (Fahn, 1982), and it is also associated with the formation of an abscission layer. Digitate leaves are composed of a petiole and individual leaflets (Fig. 5.1c). Each leaflet in theory comprises a leaflet blade and petiolule however, in practice the petiolule is rarely visible in lupin leaflets and may be completely absent in sessile leafleted species. More superficial definitions of simple leaves, as those with one blade subtending an axillary bud and compound leaves as those with more than one blade subtending an axillary bud (Champagne & Sinha, 2004; Jackson, 1928) fail to distinguish between simple leaves and unifoliolate leaves, i.e. those derived via reduction from multifoliate leaves, and which are therefore fundamentally compound (Fig. 5.1). This difference is important because the evolutionary histories of simple and unifoliolate may be quite distinct. Furthermore, the homology of the compound leaf to simple leaves or leafy shoots has been questioned (Champagne & Sinha, 2004; Kaplan, 1975; Sattler & Rutishauser, 1992). These fuzzy boundaries between leaf types and how they are defined are not widely acknowledged.

The absence of clear anatomical evidence to distinguish between simple and unifoliolate leaves in some species and the fundamentally compound digitate nature of the leaf in the Leguminosae more generally has commonly led to the use of the terms simple and unifoliolate variably and interchangeably. Polhill (1982) distinguished simple, 1-foliolate and digitately-foliolate leaved-species of *Crotalaria*. However, Cronquist (1988, p 179) stated that "In the legumes the compound leaf has repeatedly been reduced to a unifoliolate leaf, and the joint at the base of the remaining leaflet may then be suppressed, so that the leaf is truly simple". Two papers have discussed in detail whether non-digitate lupin metaphylls are simple or unifoliolate (Monteiro & Gibbs, 1986a; Planchuelo & Dunn, 1984). However, discussion has not been about the definition of simple versus unifoliolate leaves (although maybe it should have been) but rather how and whether the presence of the pulvinule can be detected.

Unifoliolate-leaved species are not uncommon in predominately compound-leaved genera, particularly within the Leguminosae e.g. *Andira, Baphia, Dalbergia, Griffonia, Poissonia* and *Pterocarpus*. Also it is not uncommon for this character to have arisen multiple times within a group e.g. *Aspalathus* L. according to Dahlgren's (1970) classification. Although members of the Genisteae are generally characterised by digitate leaves, non-digitate metaphylls are not uncommon in *Cytisus, Retama, Spartium* and *Ulex,* and predominate in *Genista* (Polhill, 1976). Often however, these non-digitate metaphylls are reduced to scales, narrow phyllodes or thorns (Bieniek & Millington, 1967; Polhill, 1976). As the *Pisum* mutant, UNIFOLIATA, shows, the production of unifoliolate rather than multifoliolate leaves can also occur spontaneously within species (Hofer *et al.*, 1997). Furthermore, it has been observed that, in contrast to floral development where aborted structures are commonly observed in early ontogeny of flowers (e.g. McMahon & Hufford, 2005; Tucker & Hodges, 2005) with reduced numbers of parts, no aborted leaflets are seen in early ontogeny of unifoliolate leaves (Galey and Hawkins, unpublished data). Frequent switches from simple to compound and vice versa at all taxonomic levels indicate this character is highly labile.

The presence of digitate and non-digitate leaved species within *Lupinus* (Figs. 5.2-5.4, 5.6) was first reported by Agardh (1835) and has been widely utilised in various infrageneric classifications (Bentham, 1859; Gross, 1986; Sholars, 1996; Taubert, 1894) as well as being the focus of much speculation about the evolution of leaf types within the genus (Chapter 6) (Dunn, 1971, 1984; Monteiro & Gibbs, 1986a; Planchuelo & Dunn, 1984). However, precise structural and topographic correspondence i.e. primary homology, of these two metaphyll character states has not been satisfactorily established (Monteiro & Gibbs, 1986a).

Three direct and indirect criteria can be used to identify the presence of a pulvinus: a) the visible presence of a pulvinus, b) the presence of an abscission layer indicating the position of a pulvinus and c) the presence of a point of articulation indicating the presence and location of a pulvinus. Although lupins are reported to have a pulvinus at the base of

each petiole these are rarely visible. Pulvini/pulvinules can sometimes be identified on herbarium specimens by either a slight bulge in the petiole/petiolule or a small region which has dried a different colour (Gwil Lewis personal communication). Digitate lupin leaves have multiple pulvinules, one at the base of each petiolule (Fig. 5.1). Again the presence of these can be difficult to observe but can be detected indirectly by leaflet movement. Lupins can exhibit diaheliotropic (sun-tracking) and paraheliotropic (cupping) leaflet movements (Figs. 5.5b and 5.6d) (Wainwright, 1977) and nyctinastic (night) movements in some species (Barneby, 1990; Kurlovich, 2002). On herbarium specimens the position of the blade at an angle to the petiole (Fig. 5.5c) may indicate the presence of a pulvinule and possible leaflet movement but this may be an artefact of pressing and drying.

Using the abscission layer criterion Planchuelo and Dunn (1984) concluded that the leaves of both the North and South American non-digitate species are simple. They observed that digitate-leaved species have abscission layers at the base of the blade and the base of the petiole. In simple-leaved species they found evidence of only one abscission layer near the base of the petiole. However, this conclusion is undermined by *L. paraguariensis,* a species which has both non-digitate and digitate metaphylls (Figs. 5.2o, 5.6c, 6.7c). The non-digitate metaphylls appear at the base of the plant and at the base of each lateral shoot (Planchuelo, 1984; Planchuelo & Dunn, 1984). *Lupinus paraguariensis* is reported also to only have one abscission layer in both digitate and non-digitate leaves (Planchuelo & Dunn, 1984). However, in contrast to this I observed evidence of abscission only at the base of the blade.

Lupinus demonstrates the fuzzy boundary between unifoliolate and simple leaves supporting the view of Cronquist (1988). A point of articulation is indicated in some species, and even in some non-digitate-leaved ones (Appendix 14), by presence of a pulvinule, movement of the leaflet relative to the petiole (Fig 5.5a-d) and/or leaflet abscission (Fig 5.5e), supporting a unifoliolate condition. However, in other species, no such evidence of compoundness is apparent, suggesting a simple-leaved condition.

The point of articulation in the form of the pulvinule, is a late developmental stage feature suggesting that its reduction or elimination is feasible (Galey and Hawkins, unpublished data). Thus, as suggested by Cronquist (1988) sole reliance on the point of articulation when describing leaf types can be misplaced. Rather than changing the definition of a simple leaf additional clarifying terms that highlight the evolutionary history of a leaf, where known, are needed. This could be achieved by utilising the terms 'primarily simple' for truly simple leaves and 'secondarily simple' for what are morphologically simple leaves derived from compound leaves. It is acknowledged that these can not be readily distinguished morphologically and that a phylogeny will be required to distinguish these types. For the purposes of character state definition I have circumvented this definition problem by lumping 'simple' and unifoliolate leaves into a single non-digitate character state, thereby recognising two metaphyll character states, digitate and non-digitate (Fig. 5.6a and b).

Leaflet number is rarely, if ever, constant within lupin species (Barneby, 1990) and shows continuous variation between species and is thus not a useful character for diagnosing infrageneric groups. There is also considerable diversity in leaflet morphology among and within both digitate and non-digitate species (Figs. 5.2 – 5.8).

In common with many Genisteae, most lupin metaphylls have intrapetiolar stipules (Fig. 5.1). Where present in lupins stipules are adnate to the petiole for at least some of their length. Stipule morphology can be a useful character for species delimitation and has been particularly useful to delimit species within Section Orientalis (See Chapter 7) (Monteiro & Gibbs, 1986a). Observation and measurement of Andean species revealed that within species and even within individuals stipule morphology varied quantitatively in terms of size (Fig. 5.9) and qualitatively in terms of pubescence. Due to the high levels of variation and the challenge that coding this objectively presents, I rejected stipules as a character to code for phylogenetic analysis.

5.6.3 Eophylls

Adult leaves are referred to as metaphylls (Jackson, 1928). The distinction from eophylls (juvenile leaves sensu Tomlinson, 1960) is important in the discussion of *Lupinus* morphology because the two show some striking differences as well as similarities. It is common in many angiosperm groups for compound metaphylls to follow simple or reduced compound eophylls (Ashby, 1948; de Vogel, 1980; Ricardi, 1996b, 1999; Tomlinson, 1960). This is a pattern frequently seen within the Leguminosae where species with compound metaphylls have eophylls which are non-compound or have reduced numbers of pinnae or leaflets (Akiyama, 1988; Duke, 1969; Hopkins & Silva, 1986; Veasey *et al.*, 1999). Specific eophyll characters, include arrangement of eophylls (opposite/alternate), number of eophylls (none/one/two) eophyll type (pinnate/scale) and number of leaflets in the first eophyll (one/more than one), have been utilised in phylogenetic analyses of several legume groups (e.g. Lavin *et al.*, 2001a; 2003; Luckow & Hopkins, 1995).

Eophyll morphology varies within *Lupinus*. Eophylls are recognised by their position (i.e. first leaves), reduced size and reduced number of leaflets relative to the metaphyll. Both digitate and non-digitate eophylls have been reported in *Lupinus* (Acuña & Planchuelo, 1996; Planchuelo & Dunn, 1984) and were observed in growing seedlings (Fig. 5.10 and Appendix 13) but are rarely present on herbarium specimens.

Eophylls are serially homologous to metaphylls within individuals but can be separated on topographic and morphological grounds as a separate character although there is often a gradation in number and size of leaflets from first eophyll to adult leaves. The distinction between eophylls and metaphylls is less marked in non-digitate species although, eophylls are generally smaller than adult leaves. However, because of these potential continuities, only first eophyll rather than eophyll number was defined as a character. Furthermore, the number of eophylls is not constant within species, ranging from two to many. Other studies have utilised eophyll position as a character in morphological matrices. However, it was difficult to observe whether eophylls were alternate (similar to the metaphylls) or opposite

due to crowding of young metaphylls. Further investigations are needed to dissect this thoroughly. Although the precise number of leaflets varies, the digitate nature of eophylls when present is fixed within species facilitating recognition of two character states, digitate and non-digitate (Fig. 5.10).

5.6.4 Seed coat surface pattern

Micromorphological features of seed coat surfaces can provide valuable taxonomic characters across many groups of angiosperms (Barthlott, 1984). Seeds of most papilionoid legumes appear smooth to the naked eye and magnifications of x1000 or more are necessary to observe micromorphological patterns. This level of magnification can only be observed using the scanning electron microscope (SEM). Barthlott (1984) listed four ways in which seed coats vary: cellular arrangement, shape of cells (primary sculpture), relief of outer walls (secondary sculpture) and epicuticular secretions (tertiary sculpture). Cellular arrangement, primary and secondary sculpture provide most of the variation in seed coat patterns whilst tertiary sculpture is rarely seen (Barthlott, 1984).

In some genera seed coat characters are variable among species and have been used in infrageneric classification, species delimitation and identification (Chernoff *et al.*, 1992; Gil-ad, 1998; Gupta, 1991; Isaacs *et al.*, 1993). Additionally, seed coat characters can be diagnostic of higher level clades revealed by molecular phylogenetic reconstruction (Buss *et al.*, 2001). A broad survey of the Papilionoideae identified a range of seed coat types (Lersten, 1981; Saint-Martin, 1986). While most tribes tend to exhibit one or two seed coat types, the tribe Genisteae displays an unusually wide range of seed coat surface patterns. The diversity of seed coat surface patterns observed within *Lupinus* is higher than most other Genisteae genera. Other genera such as *Adenocarpus, Chamaespartium, Cytisus, Genista, Goodia, Spartium* and *Ulex* exhibited only rugulate or multi-reticulate seed coat surface patterns (Lersten, 1981; Saint-Martin, 1978). Polhill (1976) also illustrated seed coat surface patterns from many Genisteae genera but mostly at a magnification of

x 800 which is insufficient to clearly identify micromorphological surface patterns. In Polhill (1976) there is a discrepancy between the illustrations and some of the electron micrographs presented.

Old World species of *Lupinus* have long been divided into two groups, rough- and smooth-seeded, based on macromorphological seed coat structure (Fig. 5.11) (Gladstones, 1974). The uniformity of the rough-seeded Old World species seed coat patterns has been confirmed by micromorphological investigations (Heyn & Herrnstadt, 1977) and this character state provides a unique synapomorphy supporting Section Scabrispermae (Plitmann, 1981). However, the heterogeneous nature of the smooth-seeded group was only recognised by micromorphological investigations (Heyn & Herrnstadt, 1977). These demonstrate that seed coat surface patterns are variable between and constant within the major groups of Old World lupins (Heyn & Herrnstadt, 1977). Additionally, the distribution of seed coat patterns among Old World *Lupinus* is congruent with relationships suggested by recent phylogenetic studies supporting the infrageneric classification of the Old World taxa into five sections, *Angustifoli, Lutei, Micranthi, Albi* and *Scabrispermae*, as currently recognised (Chapter 7, Ainouche *et al.*, 2004).

Whilst seed coat patterns of Old World species have been surveyed in detail, less attention has been paid to the New World species. Heyn and Herrnstadt (1977) reported relatively uniform testas from a sample of six New World species. Bragg (1983) specifically compared three species from Texas and found variation on the simple-foveolate pattern. Montiero (1987) survyed the seed coat patterns of North and South American unifoliolate-leaved species. He also sampled a small number of multifoliolate leaved-species. Plitmann and Heyn (1984) and Fuentes *et al.* (1993) both sampled three New World taxa. There is disagreement in the descriptions of seed coat patterns between Heyn and Herrnstadt (1977) and Fuentes *et al.* (1993). Together these studies showed that the New World species exhibit a wide range of microscopic variation in their seed coats.

Thus, it can be seen that while previous studies suggest that microscopic seed coat patterns of the New World taxa are variable and could provide useful taxonomic characters, this potential has not been realised because of sparse taxon sampling and failure to assess variation in a taxonomic context. Taxon sampling in seed coat studies has been heavily biased towards Old World species where almost all species have been surveyed. In contrast only 14 of the ca. 260 New World species have been examined. Lack of integration of phylogenetic knowledge with seed coat surface patterns has also meant that the congruence of seed coat pattern with major clades has not been tested. Further confusion has arisen because character states have been inconsistently partitioned and defined. The aim of my study is to overcome these problems via a detailed SEM survey embracing much denser sampling of New World species, to provide the basis for a genus-wide comparative survey of seed coat patterns for analysis in a phylogenetic context.

Variation in the seed surface patterns as observed at magnifications of x2700 – x8000 in *Lupinus* can be partitioned into eight character states (Fig. 5.12 and Table 5.2). These are described in part using four of the categories established by Lersten (1981). Five seed coat types have been previously reported for *Lupinus* while three (granular, rugulate & protuberant and ocellate) are newly described. A description of each character state is given below. Appendix 15 shows the character states of each accession observed alongside previously published data.

Table 5.2 Description of *Lupinus* seed coat surface patterns
Granular (Fig. 5.12a) – Densely granular with raised ridges, sometimes arranged into reticulate patterns. Occasionally granular protuberances may also be present. Only *L. albus* shows this type of seed surface. Independent studies (Fuentes *et al.*, 1993 - Fig. 1a, b, & c.; Heyn & Herrnstadt, 1977- Fig. 1. ; Plitmann & Heyn, 1984 - Fig. A2 & B) show some variation in the pattern of ridges within the species.
Papillose (Fig. 5.12b) – Rounded or columnar papillae (protrusions) of individual epidermal cells (5-8 µm diameter at base). Threads or/ and a dense cover of small bran-like scales or granules may be present particularly between papillae (Heyn & Herrnstadt, 1977- Fig. 3). Heyn and Herrnstadt (1977) describe this as "strewn with some amorphous substance". Narrow apical furrows are also sometimes observed.

Fascicular (Fig. 5.12c) – Fascicular cones (20-150 µm diameter at base), which comprise many elongate palisade cells that are considerably larger than protruding surface structures seen in any other *Lupinus* seed coat pattern. The cones are ribbed and sometimes twisted. Heyn and Herrnstadt (1977) report that fascicule height ranges from 100 – 300 µm. The fascicules can be exposed or covered with a delicate membrane.

Simple-foveolate (Fig. 5.12d) – Individual epidermal/palisade cells (ca. 6 µm diameter at base) are visible and isolated by grooves. The cells are cylindrical or conical in shape each with a centre shallow crateriform hollow (ca. 1-3 µm inside diameter). The width of the groove is highly variable in this category ranging from wide, ~3 µm apart, (Figs. 5.13a and 5.13e) to barely visible with some cells abutting (Fig. 5.13b & d). Cell wall thickness, surface and shape also vary considerably. Heyn and Herrnstadt (1977) described the upper margin as surrounded by a thick amorphous layer which they assumed to be cutinous. My observations suggest that the apices shown in their micrographs are just inflated. Most are inflated and rounded but *L. luteus* seed coats have thin cell walls which are wrinkled. Cell depth also varies considerably. Previously the shape of these cells has been referred to as volcano- or crater-like (Heyn & Herrnstadt, 1977).

Rugulate (Fig. 5.12e) – A uniformly flat irregular or weakly rugose surface. A shallow groove surrounding single cells is sometimes visible giving a reticulate pattern. There is some variation in the thickness and depth of the rugulae (for example compare Figs. 5.12e and 5.13g). Damaged testas may exhibit smooth plate-like tessellation in parts (see later discussion on damage and seed quality). Some accessions show a sparse scattering of small hairs or granules (Figs. 5.13g and 5.13h).

Rugulate and protuberant (Fig. 5.12f) – The two species sampled that show this character state have a rugulate ground plan with additional ornamentations in the form of small (2-5 µm diameter at base) spherical or columnar protuberances, the type and surface of which vary and may be species-specific. *L. microcarpus* has granular cylindrical protuberances (Fig. 5.14a) whereas *L. odoratus* has smooth spherical protuberances (Fig. 5.14b). Further sampling is needed to confirm these species differences.

Substriate simple-foveolate (Fig. 5.12g) – Single cells with wide grooves characterise this pattern. The cell walls are pressed tightly together at the apex and a prominent striate pattern can be seen on the cell surface. Although the two species placed in this group show some variation in cell shape (elliptic or cylindrical at the apex) they are sufficiently similar to be placed together. *L. hirsutissimus* has elliptical cells (Fig. 5.14c) while *L. truncatus* has elongated cells (Fig. 5.14d). It can be difficult to distinguish this pattern from simple-foveolate particularly at low magnification. However, the papillae in this group do not have a prominent central hollow and are distinctly striate rather than rugulate.

Ocellate (Fig. 5.12h) – Prominent ridges appearing like inflated rings are superimposed upon a rugulate ground plan in this seed coat type. It is unclear whether the ridges surround one or more cells.

Observations and micrographs presented here show close agreement with previously published data (Appendix 15). The broad taxon sampling of this survey compared with previous studies forms a solid basis to compare seed coat patterns and partition character states. Additionally, the terminology used previously has been standardised following Lersten (1981). Partitioning of character states is broad allowing significant variation within character states. Whilst others could recognise further partitions, this approach identifies clearly diagnosable and qualitative characters states. The only boundaries which are weak are those between simple-foveolate and substriate simple-foveolate and papillose and simple-foveolate (Fig. 5.12b, d & g).

Comparisons of multiple accessions within species show that seed coat patterns are generally constant providing fixed differences appropriate to delimit taxa. Exceptions to this are *L. albus*, *L. angustifolius* and *L. texensis.* The ridges of *L. albus* exhibit variation in their regularity but this granular character state cannot easily be confused with any other. The discrepancy between the character states reported for *L. angustifolius* is puzzling. Some studies (Heyn & Herrnstadt, 1977), including this one, find the papillose seed coat pattern whilst others (Ainouche *et al.*, 2004; Saint-Martin, 1978) reported a simple-foveolate pattern. In my observations a few papillae were not entire and showed a small apical groove but not to the extent seen by Ainouche *et al.* (2004) and I was in no doubt about their classification as papillose. Mis-identification is a possible cause for this disagreement. Similarly this may explain the report of both simple-foveolate and rugulate seed coat patterns in *L. texensis* (Bragg, 1983). My survey suggests that *L. texensis* seed coats (Fig. 5.13b) are indistinguishable from those of *L. havardii* (Fig. 5.13d), both being simple-foveolate. The observation of two seed coat surfaces patterns within a species could be explained by mis-identification, natural variation or hybridisation. Bragg (1983) discussed this and rejected mis-identification or other human error. The source material for the seed used in this study was the United States Agriculture Department. Identification of accessions of *L. texensis* used in this study were verified through observation of seedling material and is supported by the placement of this accession on gene trees alongside other *L. texensis* accessions.

Although no natural hybrids involving *L. texensis* have been reported it grows sympatrically with several other species. Extensive natural variation cannot be excluded and further sampling will help to clarify just how labile seed coat surface patterns are.

This survey has revealed discrepancies and inconsistencies between previous seed coat descriptions (Fuentes *et al.*, 1993; Heyn & Herrnstadt, 1977; Plitmann & Heyn, 1984). These can result from both sampling and magnification. Heyn and Herrnstadt (1977) and Plitmann and Heyn (1984) observed only New World species from the western New World clade at x 2,600. Fuentes *et al.* (1993) however observed one taxon from the western New World clade and two taxa from the eastern New World clade at a maximum magnification of x 1,250.

The SEM investigation undertaken here has shown the importance of good quality seed and careful handling and preparation of samples. The very smooth regions seen on some seed coats are attributable to mechanical damage which can obscure underlying patterns (Fig. 5.14g). The importance of looking at more than one piece of seed coat and observing patterns over the entire fragment are thus reinforced. In contrast to Lersten (1981), I did not find seed coat patterns to be more prominent near the hilum. The age of the seed appears to be relatively unimportant as similar patterns were observed using both herbarium and fresh seed. In most cases little deterioration was observed on herbarium seeds. Immature seeds were avoided where possible but did still provide identifiable seed coat patterns (Fig. 5.14h).

Limited seed coat surface pattern variation is seen within the western New World clade which is estimated to contain ca. 240 species. Three seed coat patterns found in this clade account for an estimated 16 species, whereas the vast majority show a rugulate pattern. Seed coat character is thus of very limited use within this homogeneous group. However, seed coat patterns can be used to distinguish between the two major New World clades and thus is useful as a diagnostic character in infra-generic classification (see Chapter 7). There

is overlap in the distribution of these two lineages in the southern Andes and south-central North America. Although these sections form well-supported monophyletic groups in all gene trees, they are apparently distinguished by few diagnostic morphological characters. The hypothesis that seed coat pattern provides a diagnostic character for these lineages was tested with *L. bandelierae* CEH2301 which is placed in the eastern New World clade in all gene trees and *L. sp* Wood s.n. (Fig. 10, Chapter 3) which grow in the Bolivian Andes alongside known members of the western New World clade. These accessions clearly showed a simple-foveolate pattern suggesting a close relationship with the eastern New World taxa.

Additional sampling, particularly of North American, Mexican and Central American, taxa would be desirable to ascertain whether further seed coat patterns exist. Several distinct and rare seed coat patterns were discovered within Series Occidentalis by phylogeny-guided sampling. Use of phylogeny as a guide for further sampling could be a productive approach particularly as there is some resolution between unsampled North American species.

5.6.5 Cotyledons

Cotyledon structure varies greatly across the angiosperms (Duke, 1969; Ricardi, 1996a, 1999). Cotyledon type has been used as a character for infrageneric classification of *Lupinus* (Chapter 7). *Lupinus* cotyledons are phanerocotylar and two types of cotyledon, petiolate and connate, have been observed (Acuña & Planchuelo, 1996; Ainouche *et al.*, 2004). In connate cotyledons the two cotyledons are sessile and fused to surround the hypocotyl. Connate cotyledons are infrequent amongst angiosperms being present in some members of the Casurinaceae (Flores, 1980), Myristicaceae (Mabberley, 1997) and Nymphaceae (Stevens, 2001 onwards).

Species and individuals of *Lupinus* only have one type of cotyledon (see non-molecular data matrix – Appendix 17). Only 5% of species for which data are available, show the connate

cotyledon type. This character is discrete with two character states, petiolate and connate (Fig 5.15).

5.6.6 Chromosome number

Lupinus, as with the Genisteae as a whole, is presumed to be derived from an ancient polyploid due to the prevalence of elevated chromosome numbers (2n=24 – 104) (Atkins *et al.*, 1998; Goldblatt, 1981; Wolko & Weeden, 1989). However, physical RNA mapping shows only low numbers of loci (Naganowska & Zielinska, 2002) but this could be due to a high degree of diploidization. Chromosome counts are available for 85 species representing all the sections and series recognised within *Lupinus* (Appendix 16 including Conterato & Schifino-Wittmann, 2006; Gladstones, 1980; Kazimierski, 1984; Maciel & Schifino-Wittmann, 2002; Phillips, 1957). However, these data have not been analysed in a phylogenetic context.

Individual *Lupinus* chromosomes are difficult to identify and distinguish to karyotype (Maisurian & Atabekova, 1974). Thus, until improved methods of identifying individual chromosomes are developed it is impossible to know whether two species sharing the same chromosome number have homologous chromosome sets.

Chromosome numbers range from 2n=24 to 2n=104. There is some within-species variation especially among the Old World species (Gladstones, 1974; Naganowska *et al.*, 2003) but generally one number is more prevalent. Until recently few counts of South American species had been made. Teresa Schifino-Wittmann and students have been addressing this gap (Conterato & Schifino-Wittmann, 2006; Maciel & Schifino-Wittmann, 2002). Chromosome numbers are much less variable both between and within New World species compared to Old World species. Polyploids, which have a doubling of the common chromosome number, are recorded in several species. Evolution of chromosome number is discussed in Chapter 6. I have chosen to define only ten character states and exclude rare polyploids or haploids.

5.7 Conclusion

This study presents the first genus-wide survey of *Lupinus* morphology with representative taxon sampling. It is a synthesis of both published and new data on leaves, cotyledons and seed coat surface patterns. Guidance from phylogenetic trees ensured New World sampling was phylogenetically as well as geographically and morphologically, representative. This study focused on a subset of potential characters within lupins, reports the morphological diversity of these and partitioned these into character states. The product of this study is a morphological matrix which is integrated with the phylogeny developed in Chapter 3 to investigate character evolution (Chapter 6) and diagnosibility for a new infrageneric classification (Chapter 7).

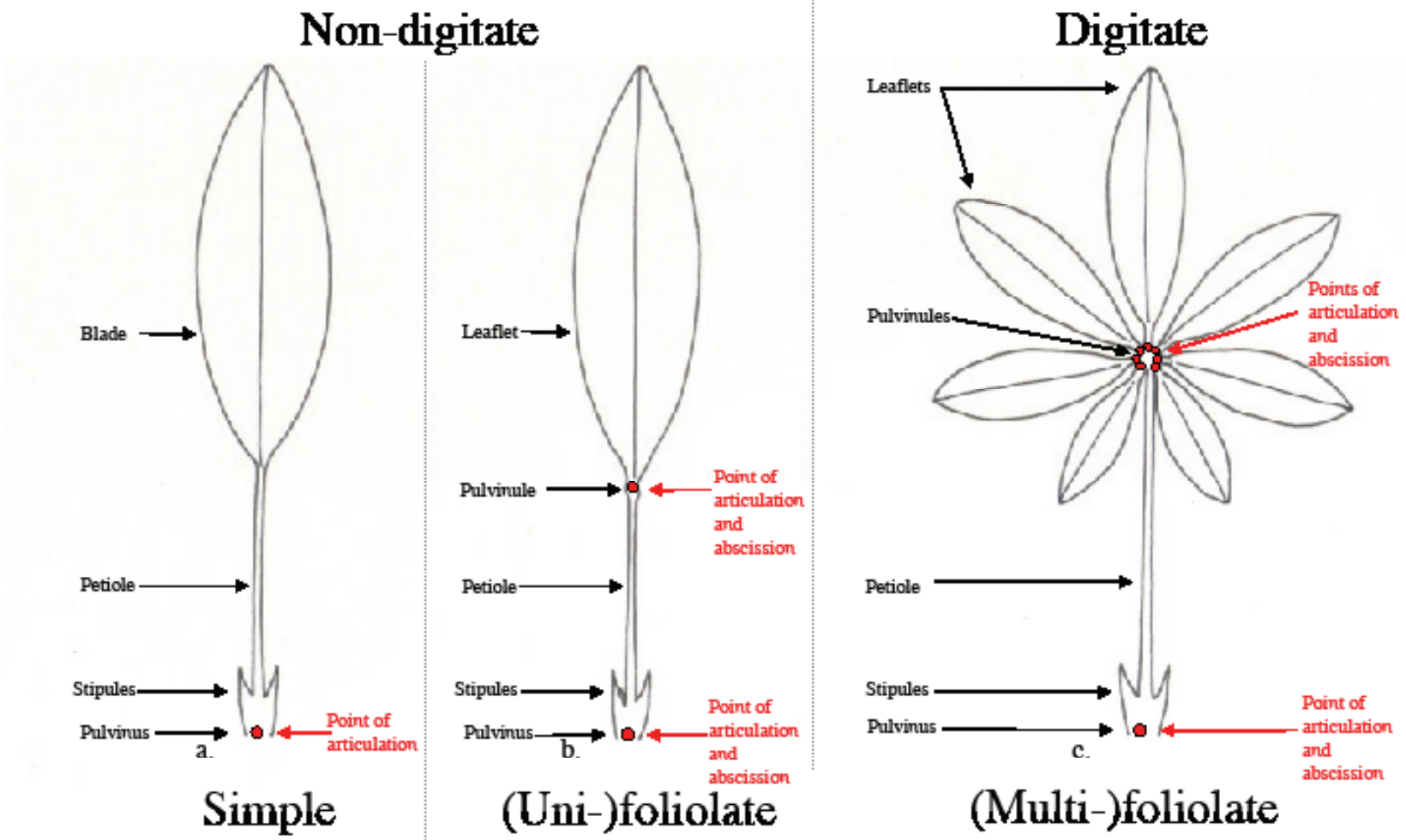


Fig. 5.1 Simple vs. foliolate *Lupinus* metaphyllls. This illustrations follow Bentham (1859) and Planchonia & Dunn (1984). a. Representation of a simple leaf, single point of articulation at base of petiole, pulvinus at base of petiole, b. Representation of a unifoliolate leaf, points of articulation at the apex and base of petiole, pulvinus at petiole base and pulvinule at leaflet base, c. Representation of a multifoliolate leaf, points of articulation at the base and apex of petiole, pulvinus at petiole base and pulvinules at base of leaflets. Illustration by Rosemary Wise.

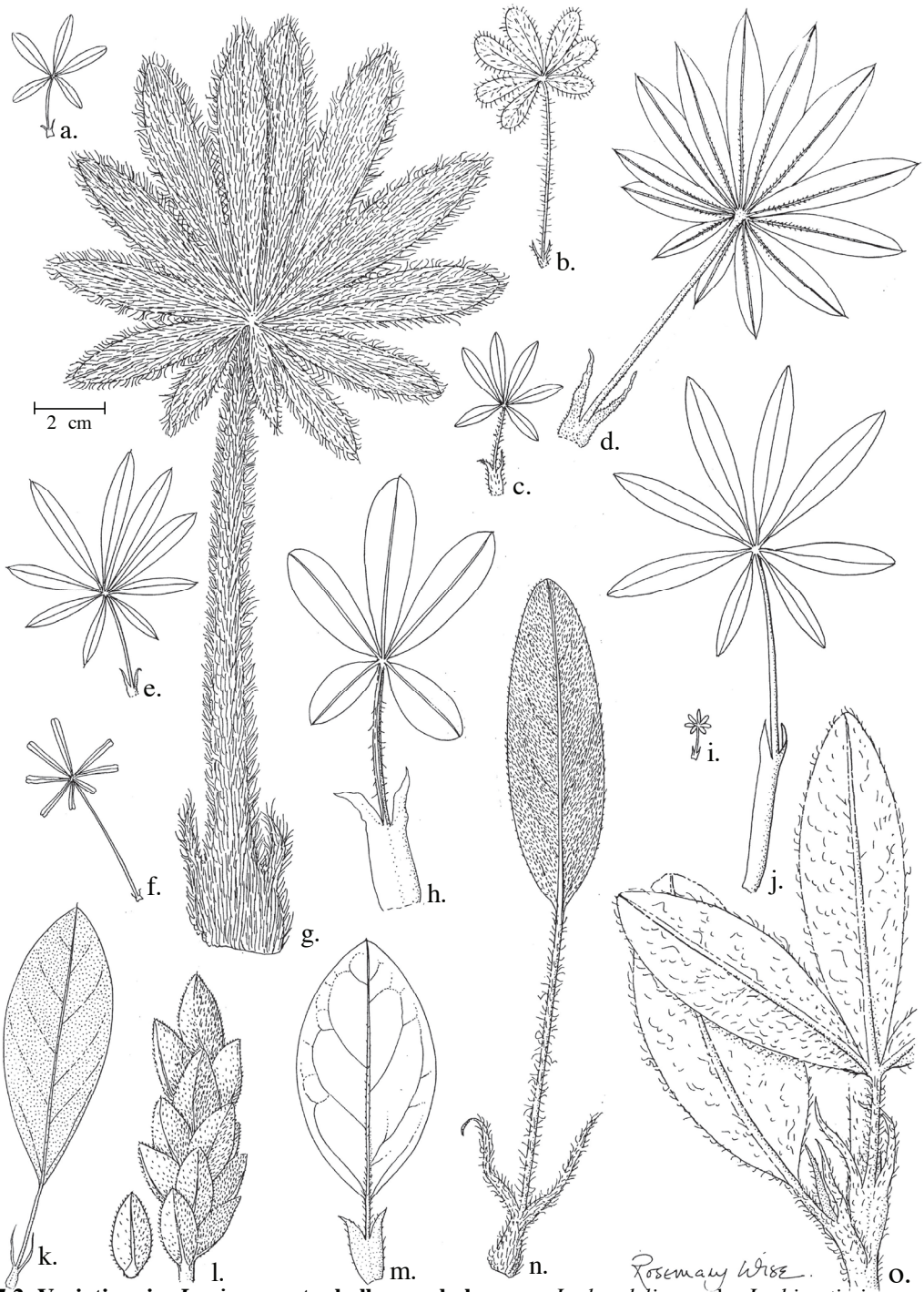


Fig. 5.2 Variation in *Lupinus* metaphyll morphology. a. *L. bandelierae*, b. *L. hirsutissimus*, c. *L. bracteolaris*, d. *L. sp3,* e. *L. arboreus,* f. *L. truncatus*, g. *L. alopecuroides*, h. *L. magnistipulatus*, i. *L. microphyllus*, j. *L. praestabilis*, k. *L. cumulicola*, l. *L. parvifolius*, m. *L. crotalarioides*, n. *L. villosus*, o. *L. paraguariensis*. Illustration by Rosemary Wise.

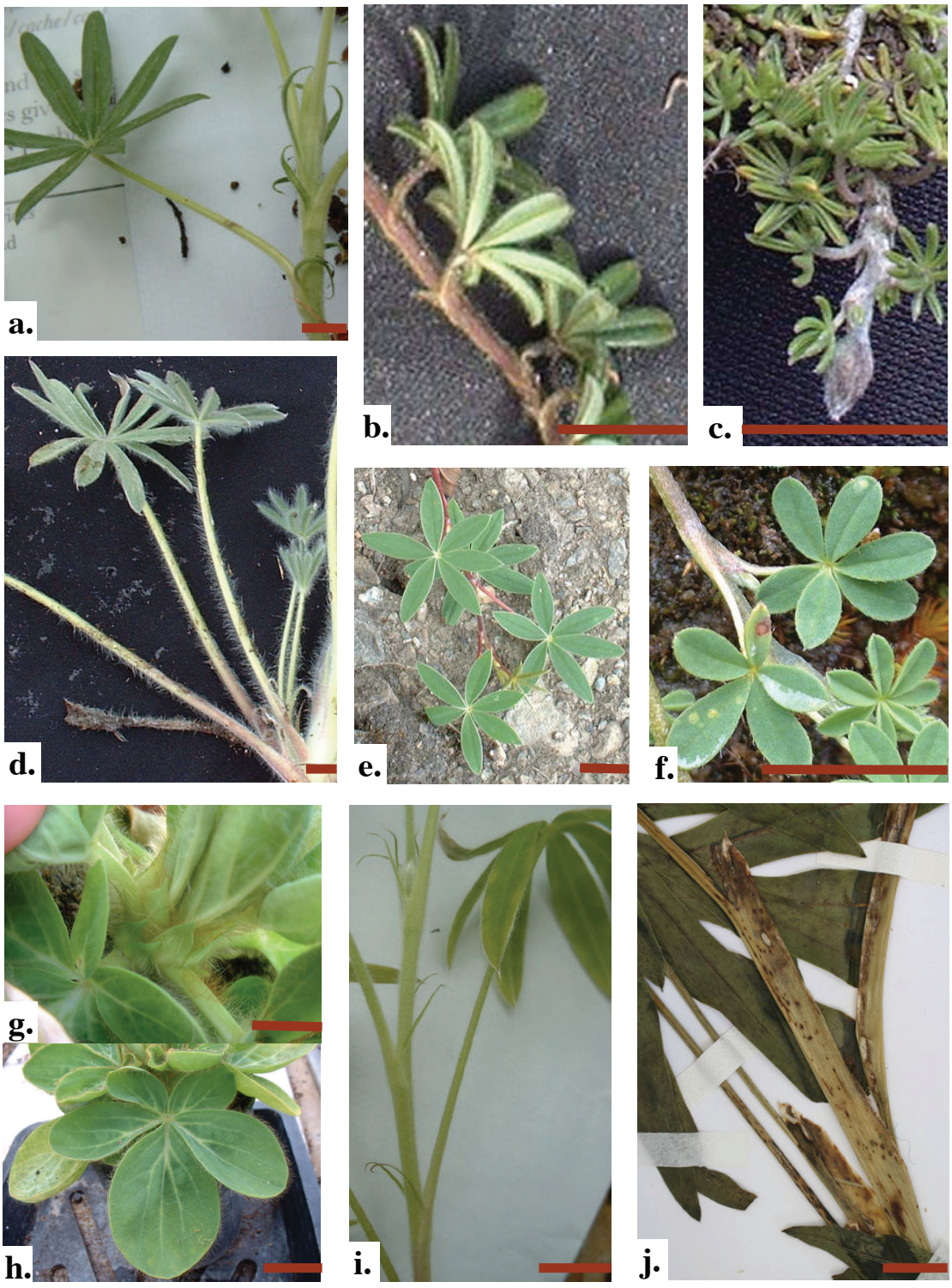


Fig. 5.3 Variation in digitate metaphyll morphology. a. Seedling of *L. sp*2 CEH2004, b. *L. sarmentosus* RJE174, c. *L. microphyllus* RJE180, d. *L. nubigenus* RJE200, e. *L. piurensis* var. *jamesonianus* RJE193, f. *L. sp*8 RJE168, g. and h. Seedling of *L. reitzii* Miotto & Pinheiro 1866, i. Seedling of *L sp*1 CEH2002, j. *L. praestabilis* RJE143. Scale bars = 1cm. Photographs RJE.

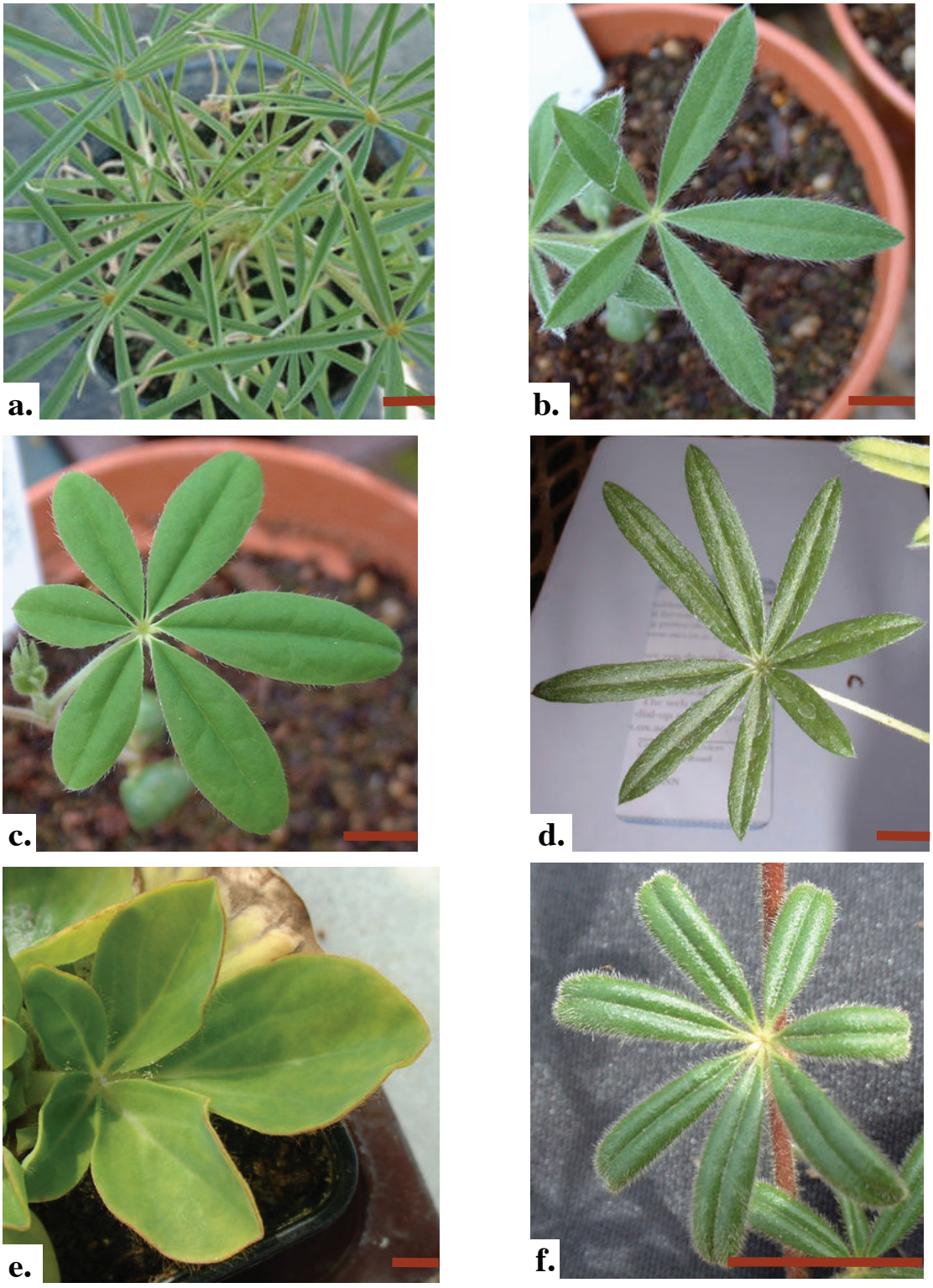


Fig. 5.4 Variation in leaflet morphology of digitate metaphylls. a. Linear *L. linearis* Miotto & Pinheiro 1798, b. Narrow elliptic Seedling of *L. sp2* CEH2003, c. Oblanceolate Seeding of *L. semperflorens* CEH2032, d. Narrow oblong Seedling of *L. sp3* CEH2004, e. Wide obovate *L. magnistipulatus* Miotto & Pinheiro 1840, f. Narrow oblong and revolute *L. sarmentosus* RJE174. Scale bars = 1cm. Photographs RJE.

Fig. 5.5 Indications of articulation. a. Leaflets exhibiting a drooping position *L. semperflorens* RJE219, b. Seedlings exhibiting heliotropism, c. Pulvinule of a unifoliolate metaphyll *L. cumulicola* CEH1983, d. Leaflets in the cupped position *L. sp* CEH2258, e. Petioles showing points of abscission at the apex where the leaflets have abscised *L. sp25* RJE91 All scale bars = 1cm. Photographs a. b. and e. RJE, c. and d. CEH.

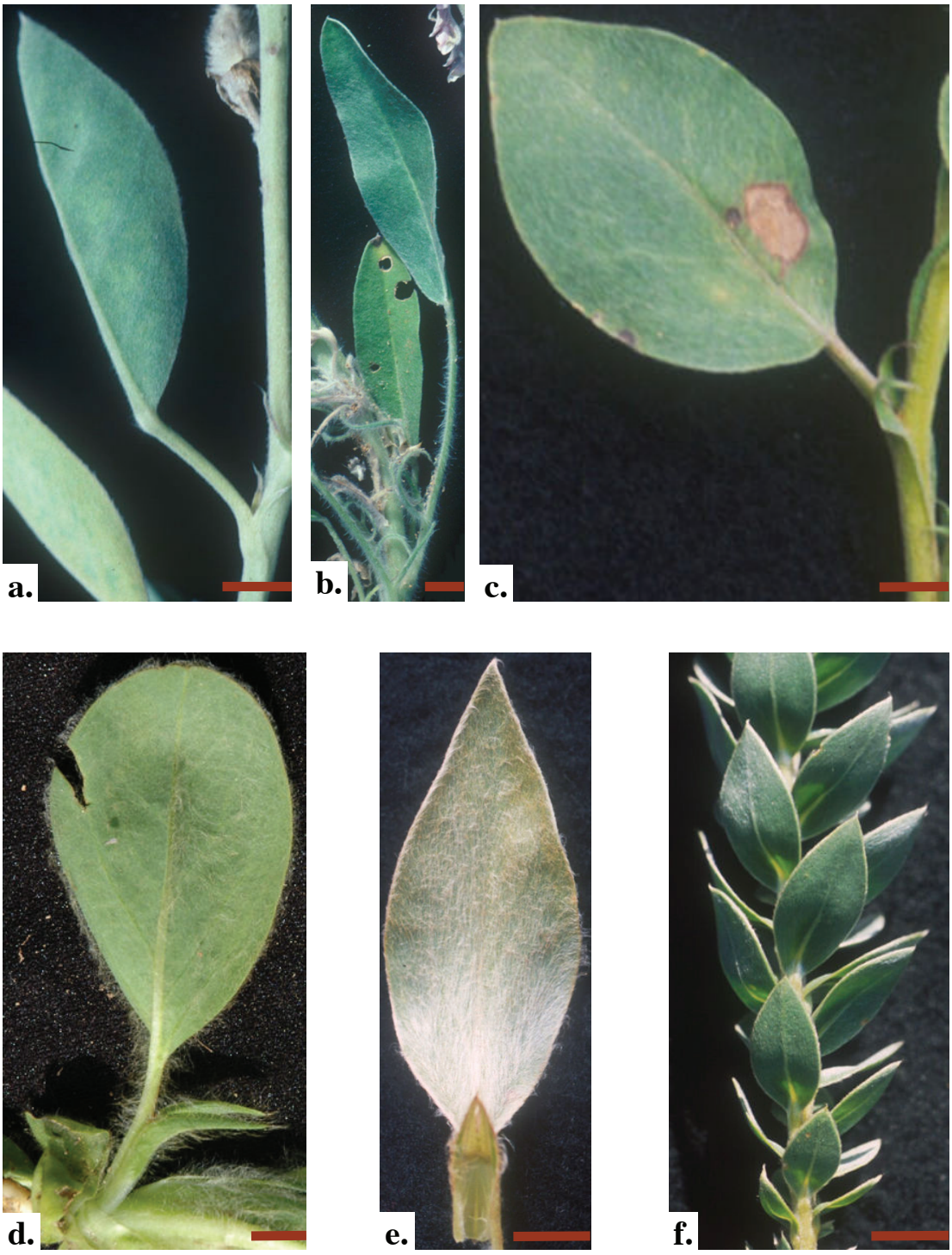


Fig. 5.6 Variation in unifoliolate metaphyll morphology. a. *L. cumulicola* CEH1983, b. *L. villous* CEH1987, c. *L. crotalarioides* CEH2481, d. *L. paraguariensis* CEH2493, e. *L. velutinus* CEH2477, f. *L. parvifolius* CEH2480. Scale bars = 1cm. Photographs CEH.

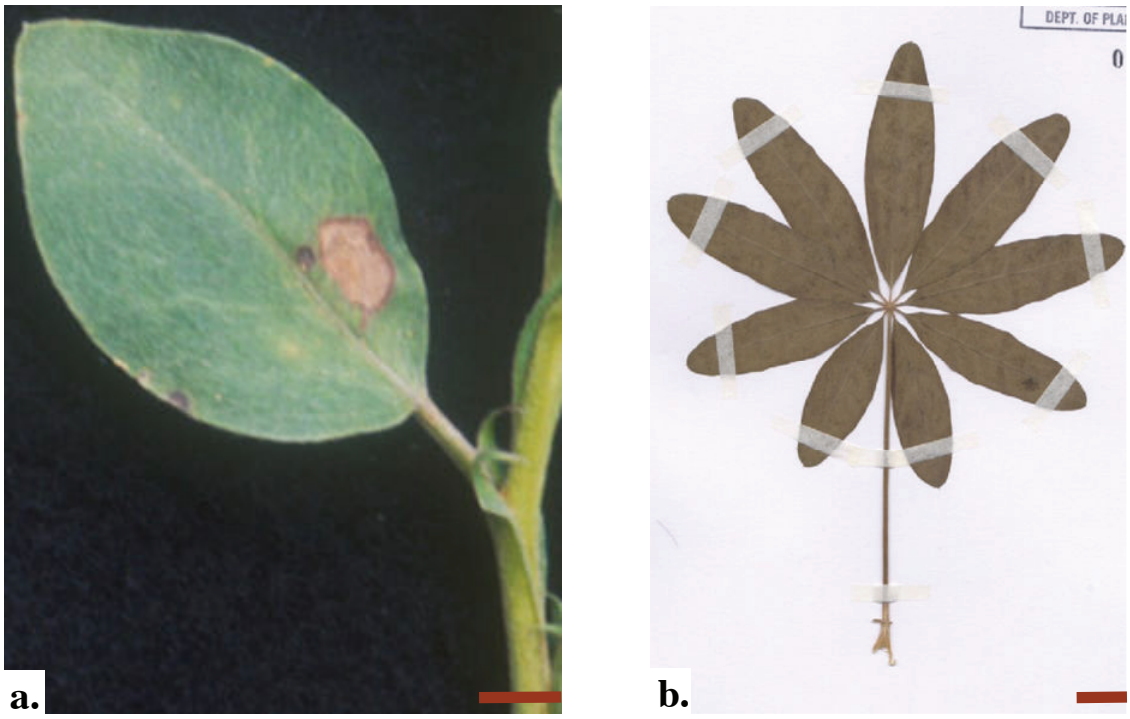


Fig. 5.7 Leaf character states. a. Unifoliolate *L. crotalarioides* CEH2481, b. Digitate *L. semperflorens* RJE15, c. Both unifoliolate and digitate *L. paraguariensis* CEH2496. Scale bars = 1cm. Photographs CEH.

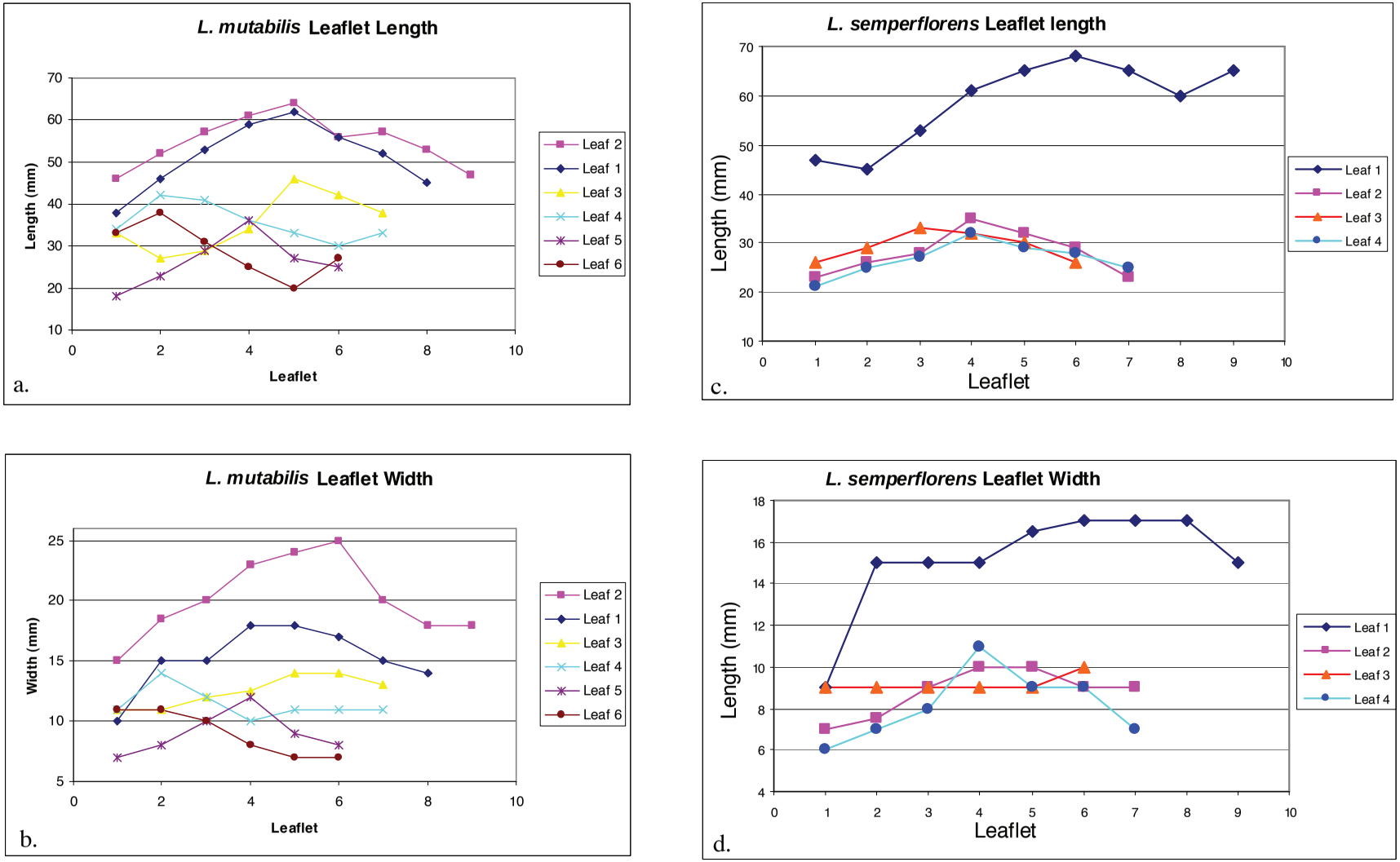


Fig. 5.8 Leaf and leaflet measurements for *L. mutabilis* and *L. semperflorens*. a. *L. mutabilis* leaflet length, b. *L. mutabilis* leaflet width, c. *L. semperflorens* leaflet length, d. *L. semperflorens* leaflet width. Leaves were measured on the same plant. Both species exhibit size variation between and within leaflets.

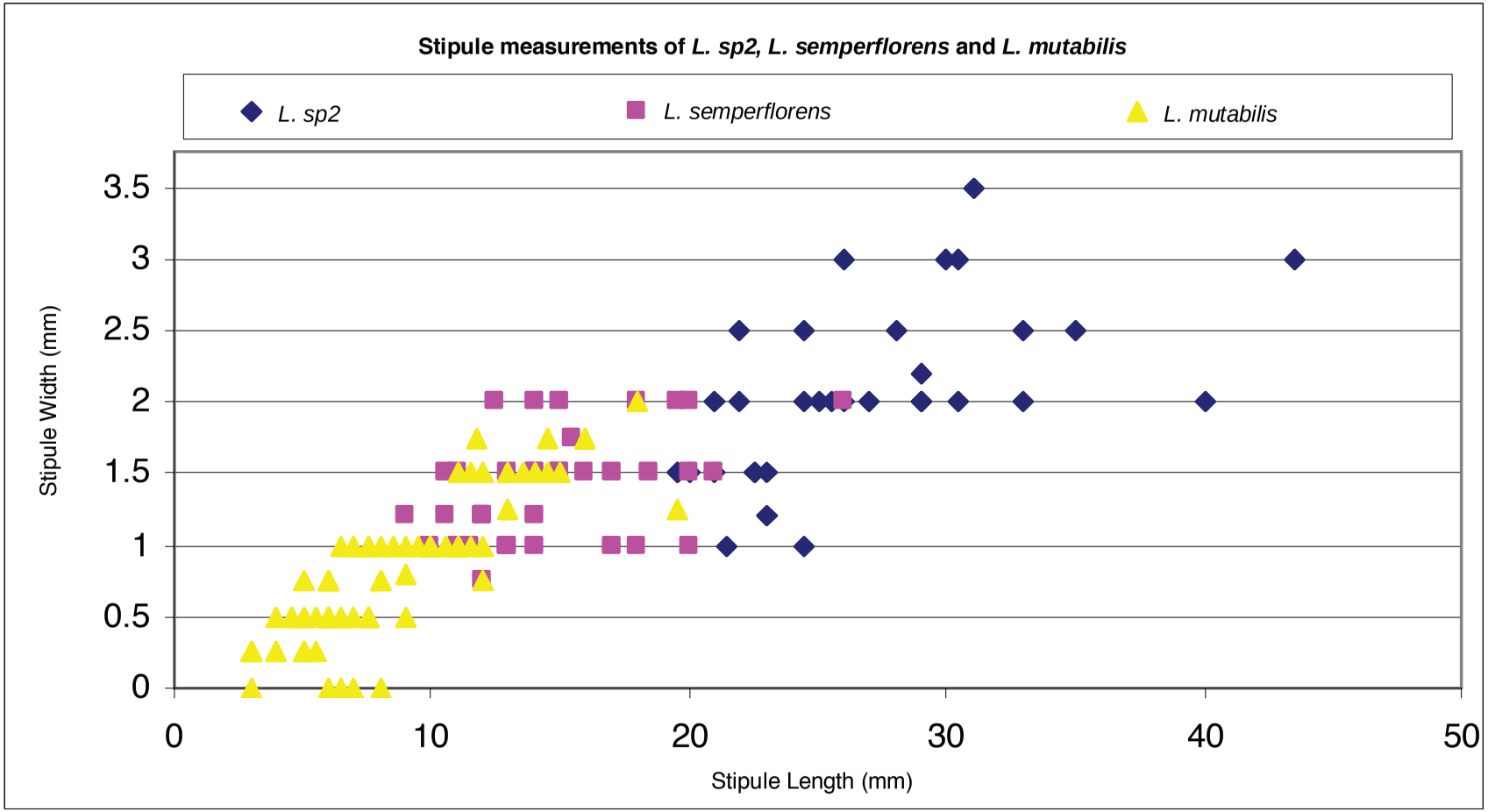


Fig. 5.9 Stipule measurements for *L. sp2, L. semperflorens* and *L. mutabilis*. Stipule size varies continuously and there is no overlap between species.

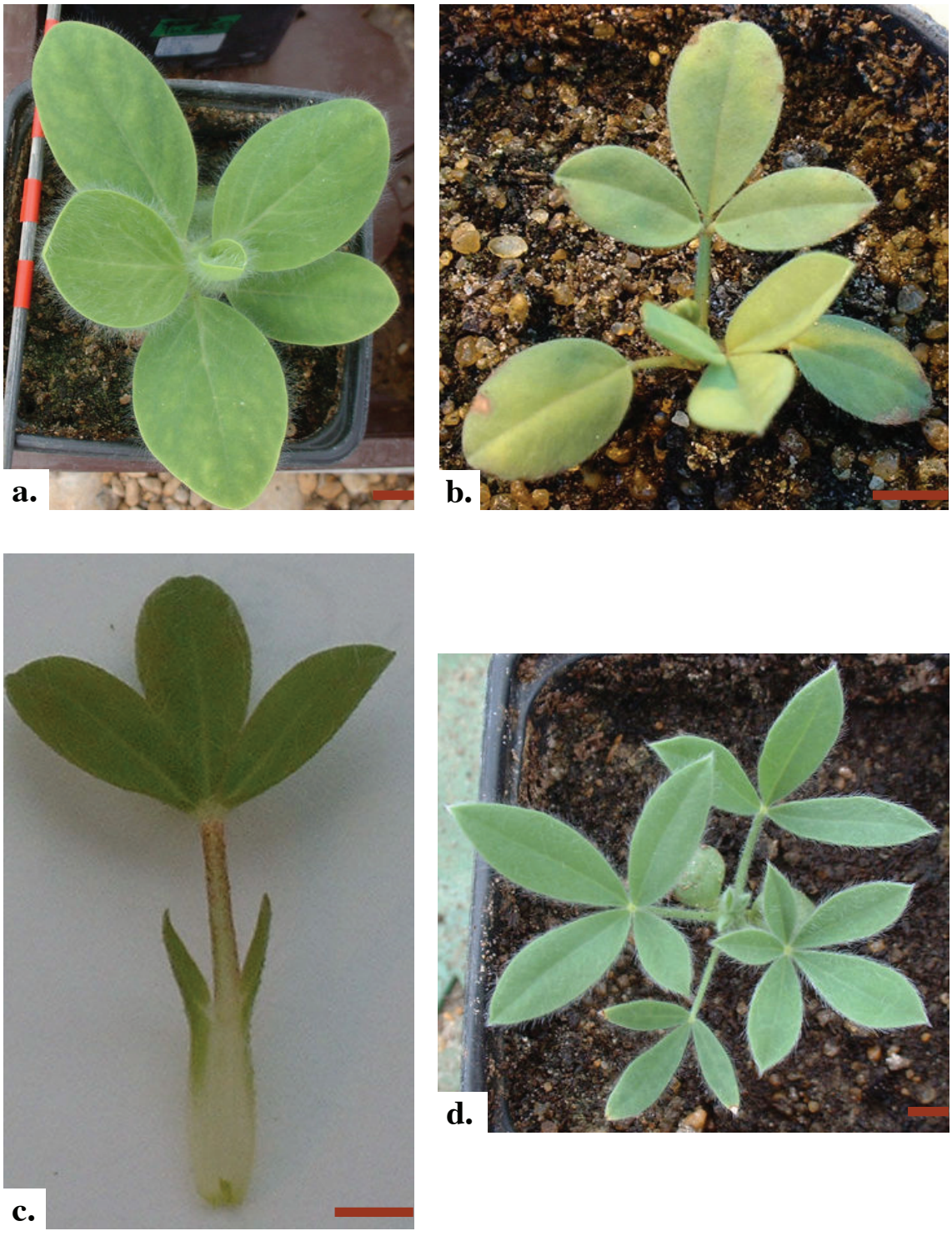


Fig. 5.10 Eophyll character states. a. Unifoliolate *L. lanatus* Kenicer s.n. 64 days, b. Unifoliolate and digitate (trifoliolate) *L. paranensis* ind2,18 33 days, c. Digitate (trifoliolate) *L. bracteolaris* s.n. 36 days, d. Digitate (tri- & multifoliolate) *L. texensis* Fothergills 41 days. Scale bars = 1cm. Photographs RJE.

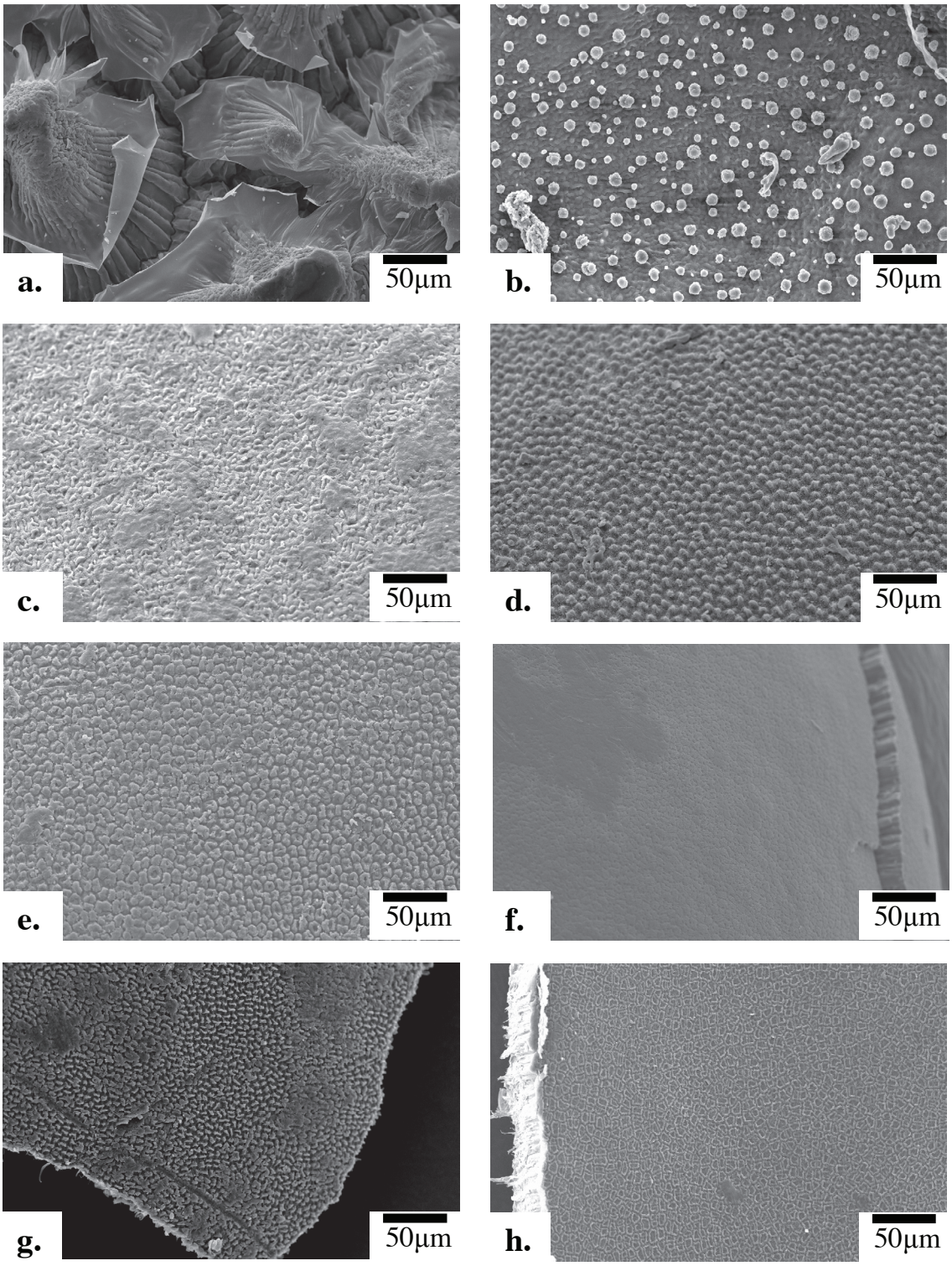


Fig. 5.11 Seed coat surface patterns at low magnification. The traditional division of macromorphologically rough- (a.) and smooth-seeded (c. - e.) Old World species can be seen at low magnifications. However, detailed patterns and character states of these and macromorphological rough (b.) and smooth-seeded (f. – h.) New World species require higher magnification and are illustrated in Fig. 6.22. a. fascicular, *L. cosentinii* USDA s.n., b. rugulate & protuberant, *L. microcarpus* CD012109, c. granular, *L. albus* Lopez and Guerro 26/2/1992, d. papillose, *L. micranthus* Anon. 29/5/2001, e. simple-foveolate, *L. villosus* CEH1988, f. rugulate, *L. honoratus* Montiel 57, g. substriate simple-foveolate, *L. truncatus* S & S Seeds Ltd, h. ocellate, *L. sp* CEH1664. Micrographs x 400. Micrographs RJE.

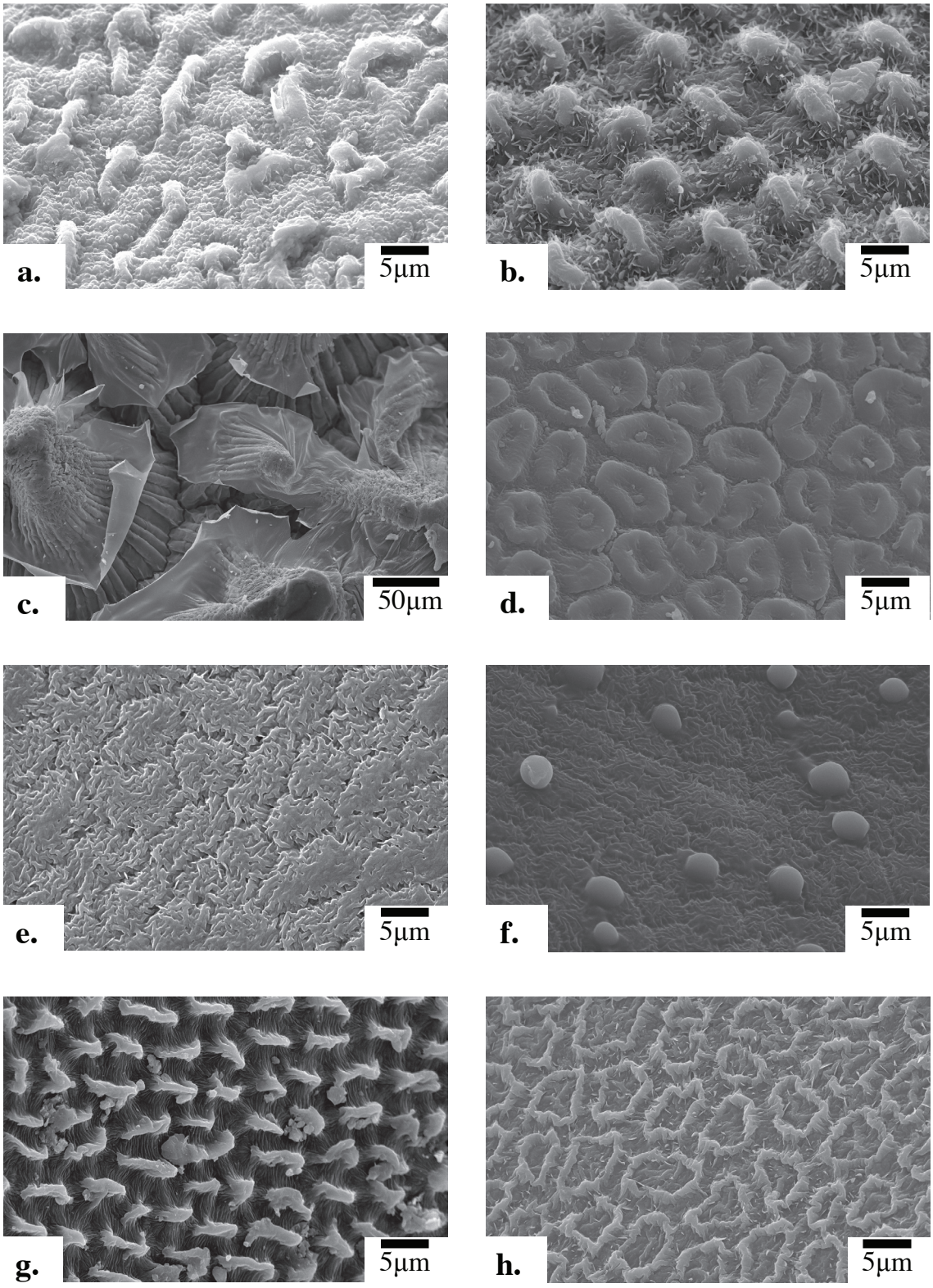


Fig. 5.12 Seed coat surface pattern character states. a. granular, *L. albus* Lopez and Guerro 26/2/1992, b. papillose, *L. micranthus* Anon. 29/5/2001, c. fascicular, *L. consentinii* USDA s.n., d. simple-foveolate, *L. bandelierae* CEH2318, e. rugulate, *L. mutabilis* CEH2326, f. rugulate & protuberant, *L. odoratus* CD0331A01, g. substriate simple-foveolate, *L. truncatus* S & S Seeds Ltd, h. ocellate, *L. sp* CEH2099. Magnification; a., b., c. - h. x 2,700, c. x 400. Micrographs RJE.

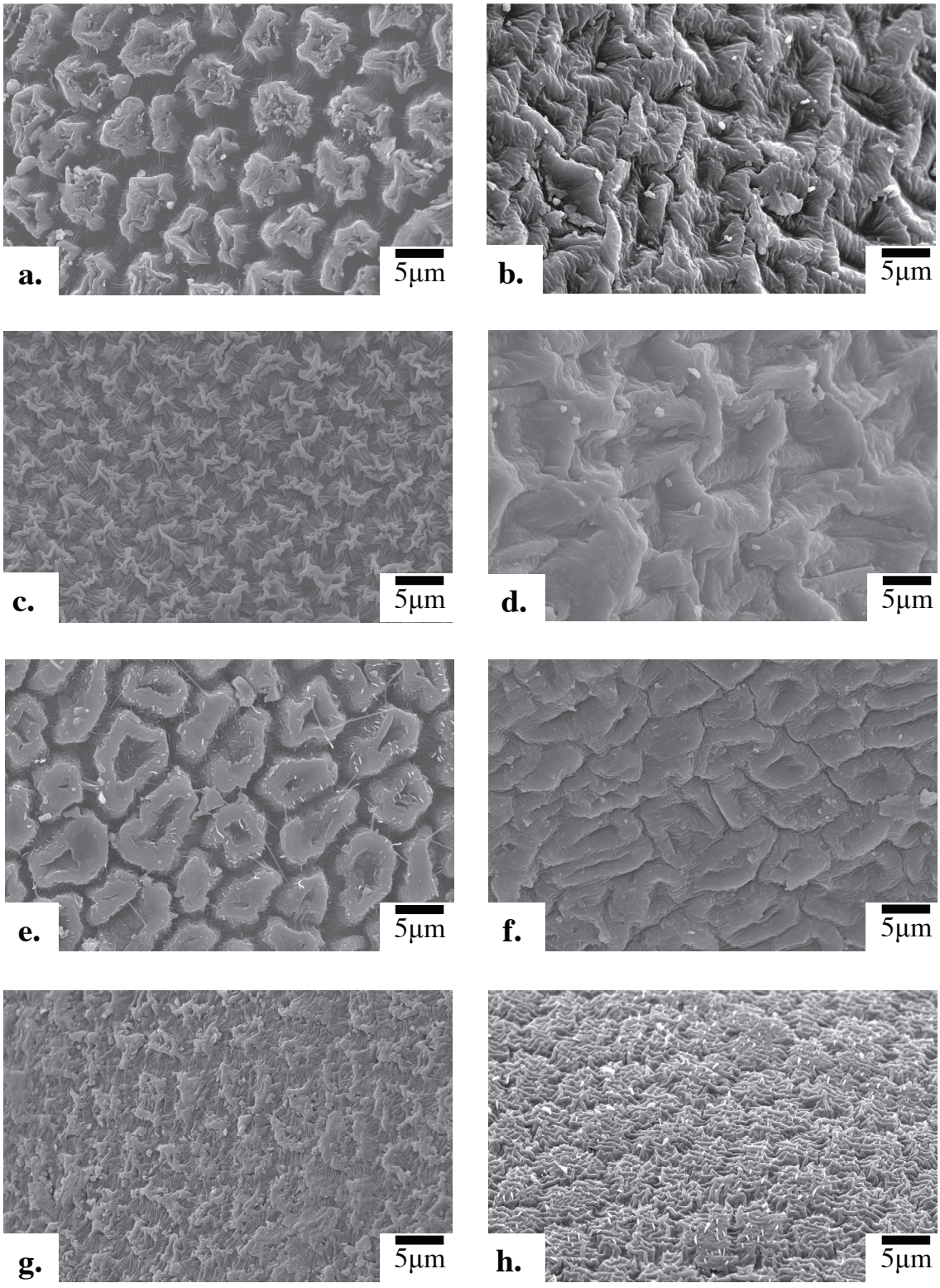


Fig. 5.13 Variation within *Lupinus* seed coat surface pattern character states. Variation in simple-foveolate character state (a.-d. and f.) and variation in rugulate character state (e., g. and h.) a. *L. luteus* Lopez and Guerro 23/3/91, b. *L. texensis* USDA 577291, c. *L. hispanicus* Castroviejo 21/6/96, d. *L. havardii* USDA 930123, e. *L. cumulicola* CEH1983, f. *L. guaraniticus* Miotto and Pinheiro 1841, g. *L. arizonicus* USDA 577285, h. *L. piurensis* RJE113. Magnification x 2,700. Micrographs RJE.

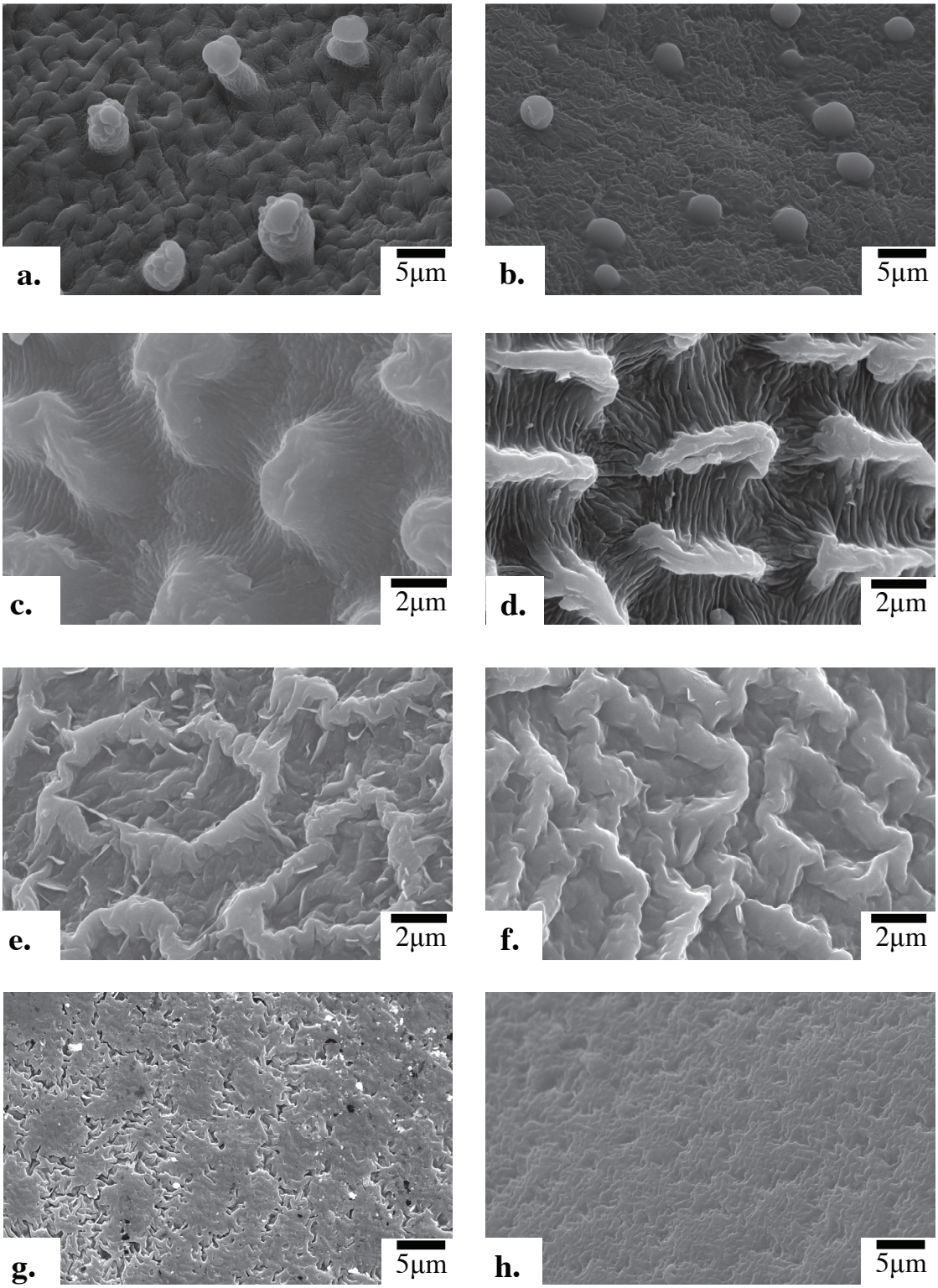


Fig. 5.14 Variation within *Lupinus* seed coat surface pattern character states. Variation within the rugulate and protuberant character state (a. and b.), variation within the substriate simple-foveolate character state (c. and d.) and variation within the ocellate character state (e. and f.). a. *L. microcarpus* CD012109, b. *L. odoratus* CD0311A01, c. *L. hirsutissimus* USDA 920120, d. *L. truncatus* S and S Seeds Ltd, e. *L. sp* CEH2099 f. *L. sp.* CEH1644, g. Example of damaged seed coat, *L. piurensis* RJE113, h. Immature seed coat surface pattern, *L. verbasciformis* Cuatrecasas 28020. Magnification; a., b., g. and h. at x 2,700, c. – f. at x 8,000. Micrographs RJE.

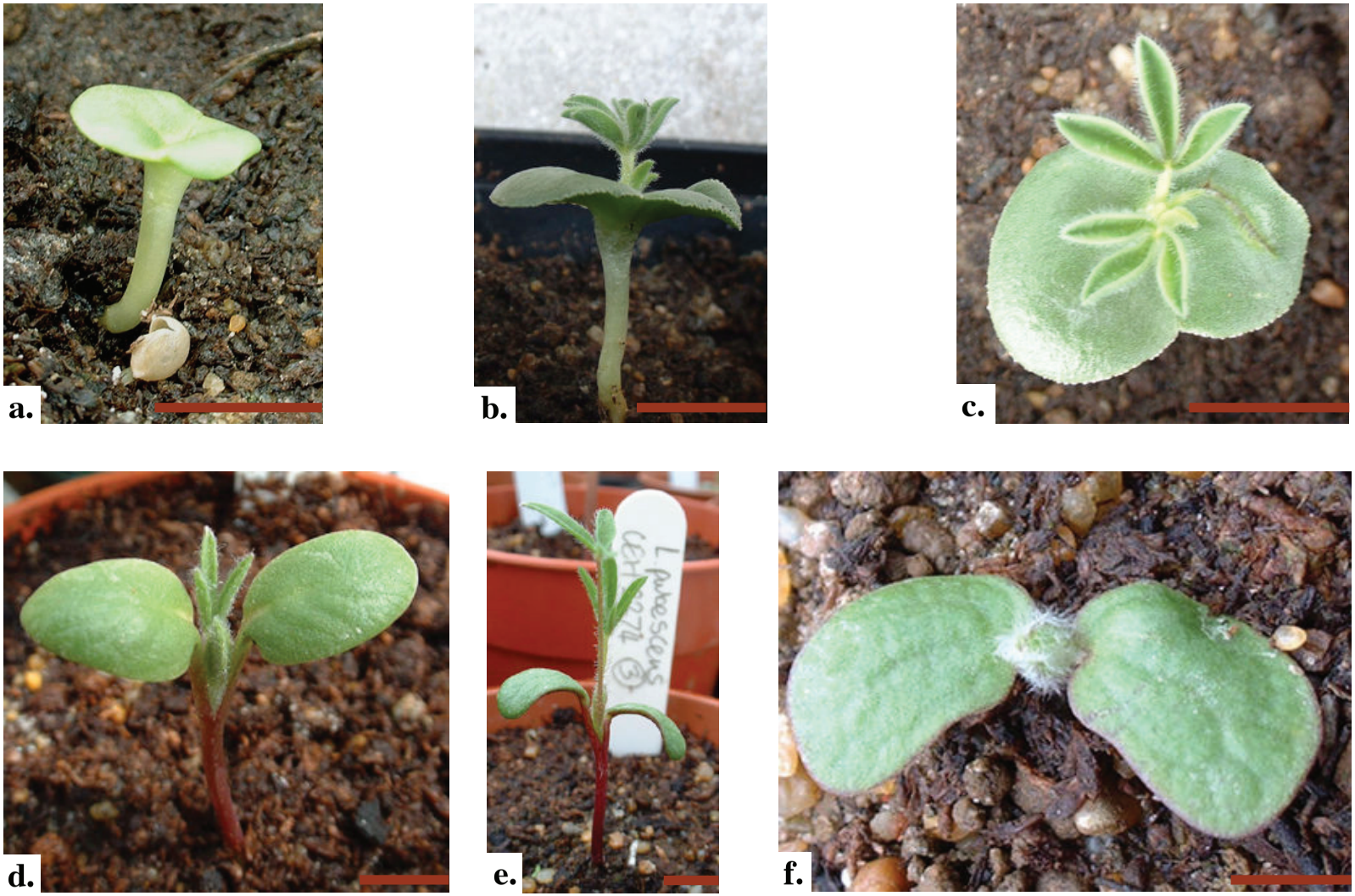


Fig. 5.15 Cotyledon character states. Connate and sessile (a., b. and c.) and petiolate cotyledons (d., e. and f.) . a. *L. odoratus* CD0311A01 7 days b. and c. *L. densiflorus* CD0331A02 17 days, d. *L.* sp1 CEH2002 14 days, e. *L. sp*. CEH2274 21 days, f. *L. villosus* CEH1986 21 days. Scale bars = 1 cm. Photographs RJE.

6 Character evolution within *Lupinus*

6.1 Using phylogeny to assess character evolution

Systematists as well as evolutionary and developmental biologists are interested in morphological evolution. Hypotheses of morphological evolution can be made on the basis of comparative morphological studies between extant and extinct taxa. The advent of robust tree-building methods and associated analyses have provided rigorous methods to assess tree support and undertake studies of character evolution (Endress, 2002). This is achieved by optimising characters onto phylogenies (Stevens, 2006). While in the past phylogenies were constructed from solely morphological data, phylogenetic estimates based on molecular data provide independent and robustly supported hypotheses for character evolution studies. Indeed, some authors have argued that tree-based approaches to analyse character evolution should utilise trees built from a dataset independent to the character(s) under investigation (Hedges & Maxson, 1996; Olmstead, 1989; Sytsma, 1990).

Tree-based studies of character evolution are now routine and very common (e.g. Borchiellini *et al.*, 2004; Carlsward *et al.*, 2006; Donoghue *et al.*, 2003; Goldblatt *et al.*, 2002; Hall *et al.*, 2004; Sauquet *et al.*, 2003; Smith *et al.*, 2005; Soltis *et al.*, 2005) and are carried out at various taxonomic levels from orders to within genera e.g. (Ainouche *et al.*, 2004; Archibald *et al.*, 2005; Crawford & Mort, 2005; Simpson *et al.*, 2004; Wilson & Calvin, 2006). A wide range of characters have been investigated in this way including number and shape of flower parts (Archibald *et al.*, 2005; Guzman & Vargas, 2005; Williams *et al.*, 2004), life history (Archibald *et al.*, 2005), fruit and root characters (Horandl *et al.*, 2005) as well as more complex functional characters such as pollination syndromes (Gravendeel *et al.*, 2004; Kay *et al.*, 2005) and breeding systems (Zhang *et al.*, 2006), or physiological characters such as plant metabolic pathways (Crayn *et al.*, 2004; Kadereit, 2003).

Various aspects of tree-based methods for investigating character evolution remain contentious and the focus of continued debate. An obvious but important assumption

of this method is that the gene trees are an accurate representation of phylogeny. If not then inferences of character evolution may inevitably be spurious. Furthermore, different character coding schemes and methods of optimisation can affect the position of character states changes on a tree (Fig. 6.1). Ambiguous, accelerated and delayed optimisations can be implemented. Accelerated optimisation (ACCTRAN) favours reversals over parallel evolution and tends to place characters state changes nearer the root of the tree (Fig. 6.1) (Farris, 1970; Swofford & Maddison, 1987). Delayed optimisation (DELTRAN) on the other hand, favours parallel evolution over reversals and thus tends to place character state changes nearer the tips of the tree (Fig. 6.1) (Swofford & Maddison, 1987). Additionally, weighting can be employed to impart polarity between character states and character ordering to model a transition series.

Parsimony analysis implements the congruence test of Patterson (1982b; 1988; 1993) by which primary homologies are evaluated under parsimony to assess their congruence with the accepted phylogeny estimate i.e. other characters. If a character or character state forms a synapomorphy, at the level of the group under investigation, it is congruent with the phylogeny and accepted as homologous. Conversely, the congruence test can highlight parallelisms which have been overlooked in the past (Crayn *et al.*, 2004).

6.2 Leaf evolution

The first angiosperm leaves are thought to have been entire and simple (Cronquist, 1988; Doyle & Hickey, 1976; Eames, 1961; Takhtajan, 1991). This hypothesis is consistent with current angiosperm phylogeny estimates where the majority of the basal groups have simple entire leaves (The Angiosperm Phylogeny Group, 2003; Hilu *et al.*, 2003; Stevens, 2001 onwards). The occurrence of compound leaves scattered in distantly related groups across the angiosperm tree strongly suggests multiple independent origins for this condition (Bharathan & Sinha, 2001). Reinforcing this idea is the discovery compound leaf development in the distantly related pea and tomato are controlled by different genes

(reviewed in Hofer *et al.*, 1997; Piazza *et al.*, 2005). Much historical debate surrounds the evolution of palmate and pinnately compound leaves (Parkin, 1953; Sinnott & Bailey, 1915; Takhtajan, 1991). Recent ontogenetic data suggest that these different leaf types can arise straightforwardly by differential growth and leaflet inception (Bharathan *et al.*, 2002; Kim *et al.*, 2003; Merrill, 1979).

Heterophylly, the production of different leaf types on one plant, is common across the angiosperms (Duke, 1969; Foster, 1951; Franck, 1976; Kaplan, 1975; Ricardi, 1996b, 1999). Such variation can occur among adult leaves (e.g. *Hedera helix* (Araliaceae), *Ranunculus aquatilis* (Ranunculaceae) (Briggs & Walters, 1984), *Cyanea solanacea* (Campanulaceae) (Givnish *et al.*, 1994) *Begonia* (Catherine Kidner, personal communication) and *Acer* (Sinnott & Bailey, 1915)).

However, more frequently, and of more relevance to *Lupinus*, heterophylly is commonly manifest between eophyll (juvenile) and metaphyll (adult) leaf morphologies (de Vogel, 1980) across the angiosperms including most legumes (Ricardi, 1996b). Eophylls are generally less complex than metaphylls and frequently there is a progression of increasing complexity from juvenile to full adult metaphylls (Duke, 1969; Tomlinson, 1960). However, there are exceptions where more complex leaves precede simpler ones e.g. *Ulex europeus* (Leguminosae) (Bieniek & Millington, 1967), *Dysoxylum densiflorum* (Meliaceae) (Burger, 1972) and *Lebeckia pungens* (Leguminosae) (Polhill, 1976). Few studies have investigated the ontogeny of heterophylly in detail. While some early studies e.g. Franck (1976) and Merrill (1986) concluded that disparate pathways operated in different leaf types and suggested that eophylls are produced by different mechanisms from metaphylls, other studies (Kim *et al.*, 2003; Merrill, 1979; Tomlinson, 1960) showed that differential development accounts for different leaf morphologies and that eophylls are fundamentally similar to metaphylls (see discussion in Chapter 5). Yet it is useful to retain the nomenclatural distinction when investigating leaf evolution in *Lupinus* since early and mature leaves often differ in morphology.

Heterochrony, a change in rate or timing of developmental processes, has been recognised as an important evolutionary process which can radically alter morphology (Li & Johnston, 2000). Whilst most studies of heterochrony in plants have focused on floral development, others have indicated that heterochrony can also be important in determining adult leaf morphology (Lammers, 1990; Tomlinson, 1960; Wiltshire *et al.*, 1994). Wiltshire *et al.*(1994) described the availability of a range of heterochronic pea mutants which demonstrate potential variation for selection to act upon. Tomlinson (1960) proposed that the arrest of a transition series from eophyll to metaphyll at different stages could account for some of the variation in leaf type seen between palm genera such that some species never produce true adult leaves. Similarly, Lammers (1990) highlighted a case of successive paedomorphosis in the *Cyanea solanaceae* complex (Campanulaceae), endemic to the Hawaiian archipelago.

Whilst the majority of lupins have digitate metaphylls, approximately 17 unifoliolate-leaved species are known, with ca. two species from the southeastern United States and ca. 15 species from lowland southeastern South America (1971; Dunn, 1984; Monteiro & Gibbs, 1986a; Planchuelo & Dunn, 1984). Dunn (1984) referred to the leaf type as simple rather than unifoliolate (see discussion in Chapter 5) and assumed this to be the plesiomorphic condition within *Lupinus*. One species, *L. paraguarensis,* has both unifoliolate and digitate metaphylls on the same individual (Fig. 5.7e, Chapter 5) (Planchuelo & Dunn, 1984). Plancheulo and Dunn viewed this intermediate species as representing an evolutionary transition between the unifoliolate-leaf and digitate leaf states. They also assumed the unifoliolate leaf type arose only once and explained the disjunct geographic distribution by long distance dispersal. An alternative hypothesis is that this character state arose more than once independently in North and South America. Previous molecular phylogenetic analyses (Ainouche & Bayer, 1996, 1999; Ainouche *et al.*, 2004; Kass & Wink, 1997b; Merino *et al.*, 1999) have not sampled North American unifoliolate species so these hypotheses have not been rigorously tested until now. Similarly, few South American unifoliolate species have been analysed (only *L. paraguarensis* and *L. subsessilis*) in previous studies. They were placed

these in a clade with S. American digitate-leaved species (Ainouche *et al.*, 2004; Merino *et al.*, 1999). Chromosome counts of North and South American unifoliolate species differ (2n=52 and 2n=36 respectively) suggesting that these groups may not be closely related (Conterato & Schifino-Wittmann, 2006). Little data are available describing the nature of eophylls within *Lupinus* but all previous reports indicated that North American and Old World species have digitate eophylls (Acuña & Planchuelo, 1996). Intriguingly it is known that a group of digitate southern South American lupin species have unifoliolate eophylls (Planchuelo & Dunn, 1984). There has been no work on the genetic basis of leaf evolution in *Lupinus.*

6.3 *Lupinus* cytological data and interpretations

The Genisteae is one of the most complex tribes in the Leguminosae in terms of chromosome evolution (Goldblatt, 1981). Chromosome number is highly variable even within genera and it is likely that the whole tribe is polyploid in origin (Goldblatt, 1981). Indeed, evidence is increasing to suggest that multiple polyploidization events may have occurred during papilionoid legume evolution (Pfeil *et al.*, 2005; Shoemaker *et al.*, 2006). *Lupinus* exhibits a range of chromosome numbers (see Chapter 5) and the generally high chromosome counts (24-48-104) indicate that it is an ancient polyploid genus (Atkins *et al.*, 1998). This is supported by Wolko and Weeden (1989) who interpreted the high instance (1/3) of duplicate isozyme locus expression to support the possible polyploid nature of *Lupinus*. However, physical RNA mapping contradicts the polyploid hypothesis as it identifies low numbers of loci (Naganowska & Zielinska, 2002) but this could be explained by subsequent diploidization. Although lupins appear to be mostly behaving in a regular way during meiosis, quadrivalents are occasionally reported (Conterato & Schifino-Wittmann, 2006; Plitmann & Pazy, 1984). Multivalents are common in polyploids (Jenkins & Rees, 1991) but their frequency decreases with decreasing chromosome size. The short length of lupin chromosomes decreases the likelihood of quadrivalents being observed so this can not be used as strong evidence against the polyploid hypothesis.

Despite the difficulties in preparing and counting lupin chromosomes due to their small size and the possibility of segmentation (Maisurian & Atabekova, 1974) chromosome counts are now available for 122 *Lupinus* species (Chapter 5, Appendix 16) (including Conterato & Schifino-Wittmann, 2006; Gladstones, 1980; Kazimierski, 1984; Maciel & Schifino-Wittmann, 2002; Phillips, 1957). However, these data have not been analysed in a phylogenetic context. The Old World species show considerable variation within and between species in chromosome number (Gladstones, 1974; Naganowska *et al.*, 2003). Plitmann and Pazy (1984) observed that the smooth-seeded species have high chromosome numbers, greater than 2n=50, whereas the rough-seeded species have chromosome numbers less than 42 and also exhibited greater levels of polymorphism. They attributed this to stochastic factors and an older age for the latter. Until recently only three counts were reported for South American species (Diers, 1961). New counts for South American species reveal chromosome numbers appear less variable both between and within species than the Old World species (Conterato & Schifino-Wittmann, 2006; Maciel & Schifino-Wittmann, 2002).

A wide range of monoploid numbers have been suggested for *Lupinus* including x=6 (Dunn, 1984; Gladstones *et al.*, 1998), x=5, x=10 (Maisurian & Atabekova, 1974) and x=12 (Darlington & Wylie, 1955). However, the most likely monoploid numbers are six or 12 given that the most common diploid numbers are 36 and 48. Not all the chromosome numbers reported will divide perfectly among any of these suggested base numbers implying at least some aneuploidy. The suggested base numbers are relatively low suggesting that extant species have high ploidy, a common occurrence within angiosperms (Briggs & Walters, 1984). There are several reported cases, particularly amongst the North American species, of doublings or halving of the common 2n=48 (Dunn, 1984) suggesting that sporadic neopolyploid events have occurred. Overall this suggests that the evolution of chromosome number in *Lupinus* is complex and has involved both duplication (polyploidy) and aneuploid reduction.

The exact types of genome rearrangements leading to changes in chromosome numbers within lupins are not known and there is no consensus as to whether it is more likely that whole or partial genome duplications occurred. Gladstones (1998) suggested increases through individual chromosome duplication or subdivision whereas Naganowska (2003) also saw a role for whole genome duplication. Conterato and Schifino-Wittmann (2006) suggested reductions could occur through the production of aneuploid gametes.

Genome size in *Lupinus* varies in a continuous manner from 0.97 pg to 2.68 pg (2c values) (Naganowska *et al.*, 2003; 2006). In comparison to other angiosperms and legumes in particular this is small e.g. *Arachis hypogea* 2n=40, 1c = 5.13-5.91 pg, *Glycine max* 2n=40 1c= 0.92 – 1.68 pg and *Vicia faba* 2n=12, 1c =10.94-14.73 (Ohri, 1998). There is no linear correlation between genome size and chromosome number in *Lupinus* (Naganowska *et al.*, 2003; 2006) but some genome size clusters support Old World clades recovered in molecular phylogeny reconstruction.

6.4 Character optimisation methods

In this chapter comparative morphological characters (Chapter 5) are optimised onto the 3-locus phylogeny estimate (Fig. 3.10, Chapter 3) to investigate the evolution of leaves, seed coats, cotyledon and chromosomes. The strict consensus 3-locus parsimony tree (of 82,910 equally parsimonious trees) was chosen because all the major nodes seen in the different analyses were present in all trees yet it still illustrates the uncertainty about specific areas of the reconstruction. The sister group of *Lupinus* is unknown and this introduces considerable uncertainty into the reconstruction character evolution across *Lupinus.* The outgroups used were sufficient for rooting the tree using molecular data but are not morphologically representative of the diverse Genisteae and thus have not been coded for each character. This makes outgroup comparisons difficult for some characters.

Optimisation of character states was carried out using WinClada (Nixon, 2002). Three modes of optimisation ambiguous, accelerated (ACCTRAN) and delayed (DELTRAN) were compared. Character states were treated as unordered and no weighting was implemented.

The only supported incongruence between the individual gene trees (Figs. 3.5-3.8, Chapter 3) which could affect metaphyll optimisation is the position of *L. parvifolius* Gardn. The topology of the 3-locus consensus is congruent in this respect with that of the individual consensus trees of ITS and LEGCYC1A (Figs. 3.6 and 3.8, Chapter 3). The placement of *L. parvifolius* in the GPAT1 consensus tree (Fig. 3.7, Chapter 3) is questionable (see discussion in Chapter 3) and thus leaf characters have not been optimised onto this alternative topology.

Two different codings of chromosome number data have been optimised onto the *Lupinus* phylogeny. The first does not contain any polymorphisms with the most commonly reported chromosome number was chosen or, in the case of *L. linearis* Desr. and *L. bracteolaris* Desr. where specific accessions had been counted and found to be different, the exact count for that accession was utilised. The second approach included reported polymorphisms, excluding halves or doubles of the most common chromosome number.

6.5 Results and discussion

6.5.1 Leaf evolution

Compound leaf development is of considerable interest to those studying the evolution of leaf morphology (Bharathan & Sinha, 2001; Champagne & Sinha, 2004; Gleissberg & Kadereit, 1999; Piazza *et al.*, 2005). Lack of well-resolved and robustly supported phylogenies for groups with compound and simple leaves can obscure evolutionary leaf developmental patterns. The multiple occurrences of transitions between simple and compound leaves across the angiosperms provide opportunities to investigate and compare their genetic bases.

Optimisation of metaphyll states onto the *Lupinus* phylogeny (Fig. 6.2) shows that the plesiomorphic condition is digitate with two independent transitions to unifoliolate leaves. The placement of one of these transitions is unambiguous (clade D Fig. 6.2) whilst the position of the other differs between ACCTRAN and DELTRAN optimisations (clade E Fig. 6.2) due to the occurrence of both digitate and unifoliolate leaves in *L. paraguariensis* which was scored as polymorphic for this character.

Unambiguous optimisation of first eophyll states onto the *Lupinus* phylogeny (Fig. 6.3) suggests that digitate eophylls are plesiomorphic with three transitions to unifoliolate throughout the genus. The placements of these transitions are unequivocal.

Regarding the *Lupinus* leaf there are three main areas of discussion: 1) The evolution of digitate and non-digitate leaves, 2) Heterochronic development and 3) Adaptive value of the different leaf types.

In contrast to Dunn's (1984) study of the unifoliolate species this study does not find the non-digitate metaphyll state to be the plesiomorphic condition within *Lupinus.* The non-digitate metaphyll state is homoplastic (Fig. 6.2) strongly suggesting parallel evolution of unifoliolate leaves in North and South America. Parallel or convergent evolution is a common phenomenon throughout living organisms (Brower, 1994; Dahlgren, 1970; Losos *et al.*, 1998; Weins *et al.*, 2006). This *in situ* character gain removes the need to invoke long-distance dispersal (Planchuelo & Dunn, 1984) to account for the disjunct distribution of the unifoliolate-leaved species. No obvious fixed morphological differences differentiate these two unifoliolate groups but they can be distinguished by different chromosome numbers 2n=52 in Florida and 2n=36 in South America. The position of the transition to non-digitate metaphylls is ambiguous in South America due to the occurrence of both leaf types in *L. paraguariensis.* DELTRAN optimisation places the transition at the base of the clade sister to *L. paraguariensis* and would thus imply a third independent origin of the unifoliolate character state, within *L. paraguariensis.* Given the sister group relationship of *L.*

paraguariensis to the remaining eight South American unifoliolate species the ACCTRAN placement of the transition at the base of clade E (Fig. 6.2) is more parsimonious. However, this optimisation does not alter the fact that metaphyll state fails the conjunction test.

The idea that character states which fail the congruence test and are homoplasious could still be homologous in terms of conserved underlying developmental mechanisms has been suggested (Ashby, 1948; Hawkins, 2002). Dissociation of morphological and developmental homology has been identified in flowers (Albert *et al.*, 1998; Gustaffson & Albert, 1999) and various new terms including, latent homology, syngeny and allogeny have been suggested to describe this phenomenon (Butler & Saidel, 2000; de Beer, 1971). Syngeny and allogeny refer to homologies which have either the same or different developmental pathways respectively. Characters which are syngenic may not define the same monophyletic groups as phenotypic data (Butler & Saidel, 2000). Similarly, the term latent homology, refers to occurrence of characters that are expressed in some but not all the members of a monophyletic group, i.e. a genetically-based homology which is cryptic in some members (de Beer, 1971). Thus, the non-digitate lupin leaf state could be syngenic, i.e. with a common genetic pathway, which may be cryptic in multifoliolate species. Non-digitate leaves in the divergent North and South American clades may be latently homologous.

The presence of unifoliolate eophylls in the digitate-metaphylled species more widely across the clade in which the South American unifoliolate species are nested (Figs. 6.2-6.4) immediately suggests a link between the juvenile and adult non-digitate condition. Furthermore, all digitate-leaved lupins exhibit a progression in leaflet number from eophyll to metaphyll. The data suggest two independent cases of paedomorphosis within *Lupinus.* I hypothesize that earlier offset of digitate-leaf development, progenesis (Li & Johnston, 2000), has occurred to retain the unifoliolate morphology in the metaphyll. Further stages of progenesis would also explain the loss/reduction of the pulvinule and latterly the petiole in the nested sessile group (Fig. 6.4). Similar suggestions have been made for such an

occurrence in palms (Tomlinson, 1960) and *Cyanea* (Lammers, 1990).

Most digitate-leaved species have eophylls with three leaflets. It is possible that either a genetic or environmental stimulus could cause a reduction but this is unclear. The homoplasious occurrence in three independent clades suggests that this transition is labile and evolutionarily and genetically straightforward.

The occurrence of both digitate and non-digitate adult metaphylls in individuals of *L. paraguariensis* is intriguing and does not neatly fit any explanation of paedomorphosis. *Lupinus paraguariensis* shows a reiteration of the supposed juvenile to adult transition on each individual shoot. I agree with Planchuelo and Dunn (1984) that it represents a superficially 'intermediate' state between digitate and non-digitate leaves but suggest it represents partial progenesis and rather than the 'transition' of simple to compound leaves.

The habitats where the North and South American unifoliolate species grow show striking similarities which distinguish them from the habitats in which digitate-leaved lupins are commonly found. The North American unifoliolate species are found in or besides dry pine woods or open woodland in parts of Florida, Georgia and North and South Carolina, while the South American species have a wider distribution in lowland Brasil, Argentina, Paraguay and Uruguay in open cerrado, disturbed woodland, dry sandy regions or grassy roadsides. These habitats in both regions are virtually frost-free, with regular hot humid seasons as well as being fire-prone and containing vegetation adapted to a regular fire-regime (Eiten, 1972; Menges & Kohfeldt, 1995). There is little literature discussing the adaptive value of compound versus simple leaves. Dunn (1984) suggested that digitate leaves cool more effectively and can control water loss more tightly. However, since many digitate-leaved lupins grow in habitats which experience relatively hot and dry periods, this provides no adaptive explanation for different leaf types. A possible hypothesis to explain the retention of unifoliolate leaves is selection pressure for fast growth in fire-prone environments (Gwil Lewis, personal communication). This is consistent with the paedomorphic hypothesis and

the fact that many species growing in fire-prone environments display a range of adaptations to promote rapid growth. Detailed ecological studies will be required to ascertain the adaptive significance of unifoliolate leaves with greater certainty.

6.5.2 Seed coat surface pattern evolution

Optimisation of seed coat surface pattern data onto the *Lupinus* phylogeny (Fig. 6.5) implies that the plesiomorphic state for *Lupinus* is simple-foveolate, consistent with the common occurrence of this seed coat type across the Genisteae (Lersten, 1981; Polhill, 1976). Other seed coat types can be viewed as specialised forms that have evolved from this common type (Fig. 6.5). Although the greatest seed coat pattern diversity is seen among the Old World species, it is interesting to note the difference in seed coat surface pattern diversity between the two large New World clades (clades A and B Fig. 6.5). While species in clade A have simple foveolate seed coat surface patterns, albeit with some limited variation (Fig. 5.13a-d, Chapter 5), clade B shows four different states. This is perhaps contrary to expectation given their evolutionary histories (Chapter 3). The ocellate seed coat surface pattern was observed in two Mexican taxa which were not sampled in the phylogenetic analyses. On the basis of morphology (non simple-foveolate seed coat) they are placed within clade B (Fig. 6.5) and are likely to group with the other Mexican accessions. Two areas of minor ambiguity (clades B and C Fig 6.5) are resolved differently by ACCTRAN and DELTRAN optimisation (Fig. 6.6). The ambiguity at the base of clade B may be due to the sparse taxon sampling in this part of the tree.

Only two character states, substriate simple-foveolate and papillose, fail the congruence test. In the light of the phylogeny, delimitation of the papillose, simple-foveolate and substriate simple-foveolate character states was reconsidered. Despite the presence of very shallow crateriform centres in the substriate simple-foveolate taxa, the substriate simple-foveolate state with virtually no apical depression is distinguished from simple-foveolate pending increased taxon sampling that could shed further light on this boundary.

L. angustifolius was coded as polymorphic for papillose and simple-foveolate states. Although seeds of this species are most commonly reported as papillose (Ainouche *et al.*, 2004; Heyn & Herrnstadt, 1977; Plitmann & Heyn, 1984), the simple-foveolate seed coat pattern has also been recorded (Ainouche *et al.*, 2004). Some papillae show slight craters which would suggest coding as simple-foveolate. Again this border is unclear, but suggests an evolutionary transition between these two states and the phylogeny indicates the polarity from simple-foveolate to papillose. Thus the closing of the centres of the palisade cells could constitute the papillose state.

A link between the rugulate and rugulate and protuberant states is also suggested as both seed coat patterns are fundamentally rugulate. Since optimisation at the base of clade B (Fig. 6.6) is resolved differently by ACCTRAN and DELTRAN optimisation whether these are related and the polarity of a transition between them remains uncertain.

6.5.3 Cotyledon evolution

Unambiguous optimisation of cotyledon states onto the *Lupinus* phylogeny (Fig. 6.7) shows that connate cotyledons provide a unique synapomorphy for the clade containing *L. odoratus, L. densiflorus* and *L. microcarpus.* Connate cotyledons have long been recognised in *Lupinus* but are here shown for the first time to support a monophyletic group which is recognised as Section Occidentalis Series Platycarpos (see Chapter 7). Although data are missing for the outgroup for this character, both the single transition within *Lupinus* and the rarity of this cotyledon type across the Genisteae and the angiosperms as a whole, strongly suggests that the petiolate state is plesiomorphic within *Lupinus*.

6.5.4 Chromosome evolution

There is broad congruence between the optimisations using the alternative approaches to coding chromosome numbers (Figs. 6.8 - 6.10). Most of the differences are found around the root of the tree, where the majority of the variation in chromosome number is found.

This is not surprising given that branching patterns remain highly uncertain due to short branch lengths and lack of support. All chromosome numbers are homoplasious reflecting the labile nature of chromosome evolution within *Lupinus*. The different coding methods suggest either an ancestral number of 2n=48 or 2n=52 and both these numbers are frequently reported from species across the Genisteae (Goldblatt, 1981; Polhill, 1976). Lupin species, particularly those from the Old World, commonly display a range of chromosome numbers (Appendix 16). The polymorphic coding option represents all this diversity and thus is considered more informative than the 'most common' coding method. The plesiomorphic chromosome number reconstructed using polymorphic coding is 2n=48. It is possible that ancestral lineages with further chromosome numbers existed and resolving uncertainty surrounding the plesiomorphic state in *Lupinus* may not be possible even with denser sampling.

A range of monoploid numbers have been suggested for *Lupinus*. The data are consistent with the proposed monoploid number of six (Conterato & Schifino-Wittmann, 2006; Dunn, 1984; Gladstones *et al.*, 1998; Plitmann & Pazy, 1984). A well-sampled, well-resolved and well-supported phylogeny can provides a basis for assessing chromosome evolution in terms of the likely transitions that have occurred and their polarity (e.g. Goldblatt *et al.*, 2002). Figure 6.11 speculates how some of the transitions, from 2n=48 to 36 (clade A Fig. 6.9) and from 36 to 34 to 32 (within clade A Fig. 6.10a) can be explained by euploidy and serial aneuploidy. This is purely speculative as the directional interpretations of transitions on the tree depend on fully robust phylogeny and mechanisms for these changes remain unclear. It is also clear that there are more recent polyploids within the group e.g. reports of 2n=96, 100 and 104. However, other transitions, for example from 48 to 32 (within clade C Fig. 6.10b), from 52 to 32 (within clade C Fig. 6.10a) and of course from 48 to 52 (within clades H and C Fig. 6.10a) require multiple steps. It is possible that taxa exhibiting intermediate chromosome numbers e.g. 2n=48 are extinct. Changes in chromosome number would appear to be a widespread phenomenon and may be due to the small size of lupin chromosomes such that such losses can be easily compensated. A monoploid number

n=12 is less parsimonious. For example the transition from 48 to 36 would require either the loss of half a set of chromosomes or result in a unmatched chromosome set which would be unstable, infertile or/and behave abnormally during cell division. Plitmann and Pazy (1984) suggested failure to pair correctly, due to high or low temperatures, as a possible cause of chromosome loss. However, observations of cell division have not so far revealed abnormal patterns (Conterato & Schifino-Wittmann, 2006).

6.6 Conclusion

Studies of character evolution whilst providing important insights into morphological evolution and evo-devo also reveal patterns of character distribution across phylogenies which can be utilised in the construction and diagnosis of classifications (Chapter 7). This is only informative if the homology of characters has been rigorously assessed.

Optimisation of five characters based on a detailed comparative study onto a 3-locus *Lupinus* phylogeny has provided insights into their evolution. Previously unifoliolate-leaved lupins from North and South America were assumed to be monophyletic. Here evidence strongly suggests this is not the case and that unifoliolate leaves evolved independently in each continent. This negates the need to invoke long-distance dispersal to explain the disjunct distribution of the unifoliolate-leaved species between North and South America as proposed by Dunn (1971). Seed coat surface pattern variation within New World species is relatively limited yet similar to the Old World species seed coat surface types largely characterise monophyletic groups. In contrast to most characters within *Lupinus,* which are highly labile, transition of cotyledon type has occurred only once. Despite the likely potential complexity of chromosome evolution within *Lupinus* chromosome number is highly congruent with the phylogeny. Overall the Old World species are more variable than the New World species. Persistent uncertainty surrounding the sister group of *Lupinus* introduces ambiguity in the optimisation of chromosome number since this character is also highly variable within Genisteae. Petiolar cotyledons, digitate eophylls, digitate metaphylls, simple foveolate seed coat surface pattern and 2n=48 chromosome number are hypothesised as plesiomorphic within *Lupinus*.

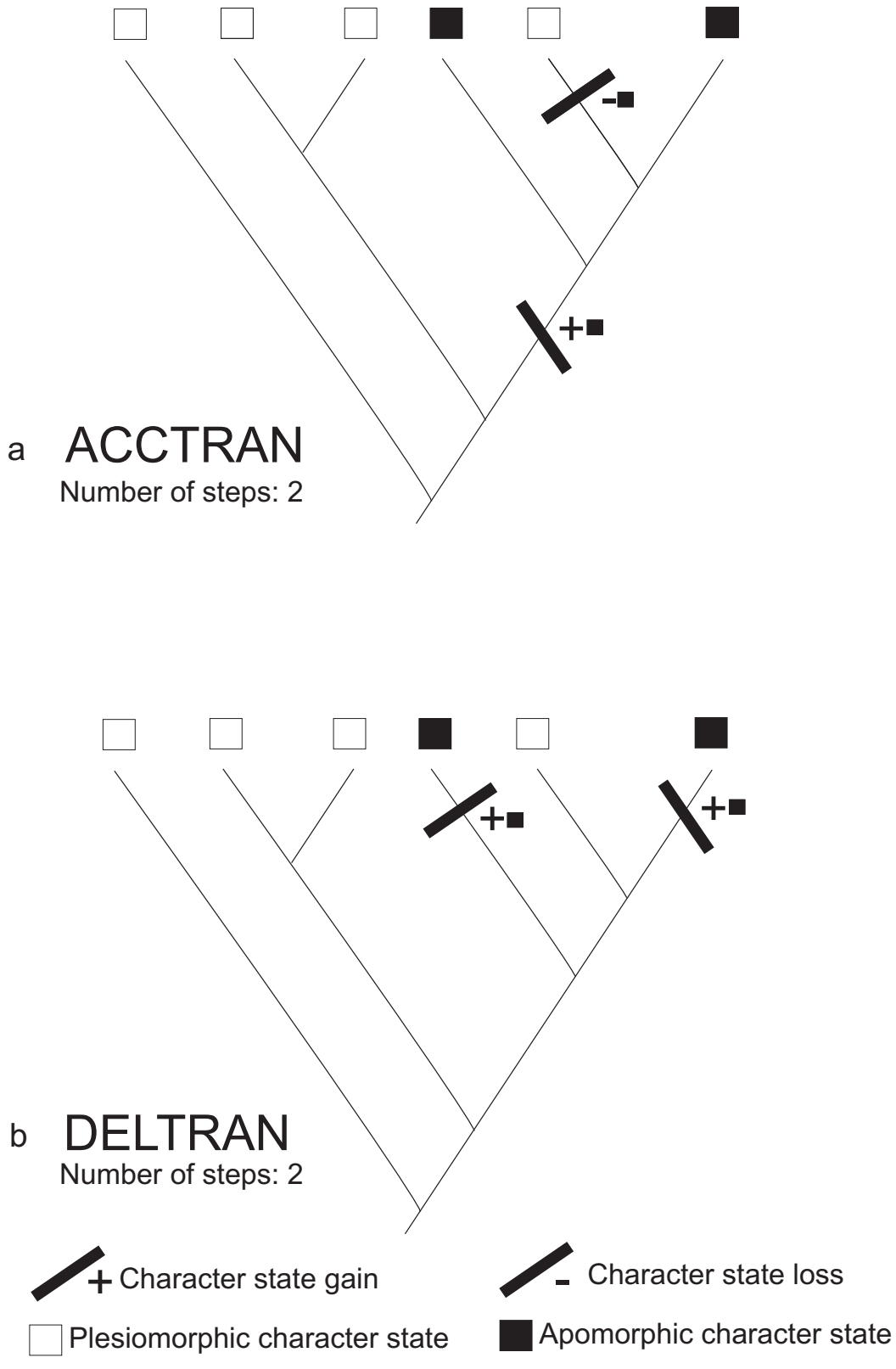


Fig. 6.1 ACCTRAN versus DELTRAN optimisation. a. ACCTRAN optimisation favours reversals over parallel evolution and places character state changes nearer the root of the tree. b. DELTRAN optimisation favours parallel evolution over reversals and thus places character state changes near the tips of the tree.

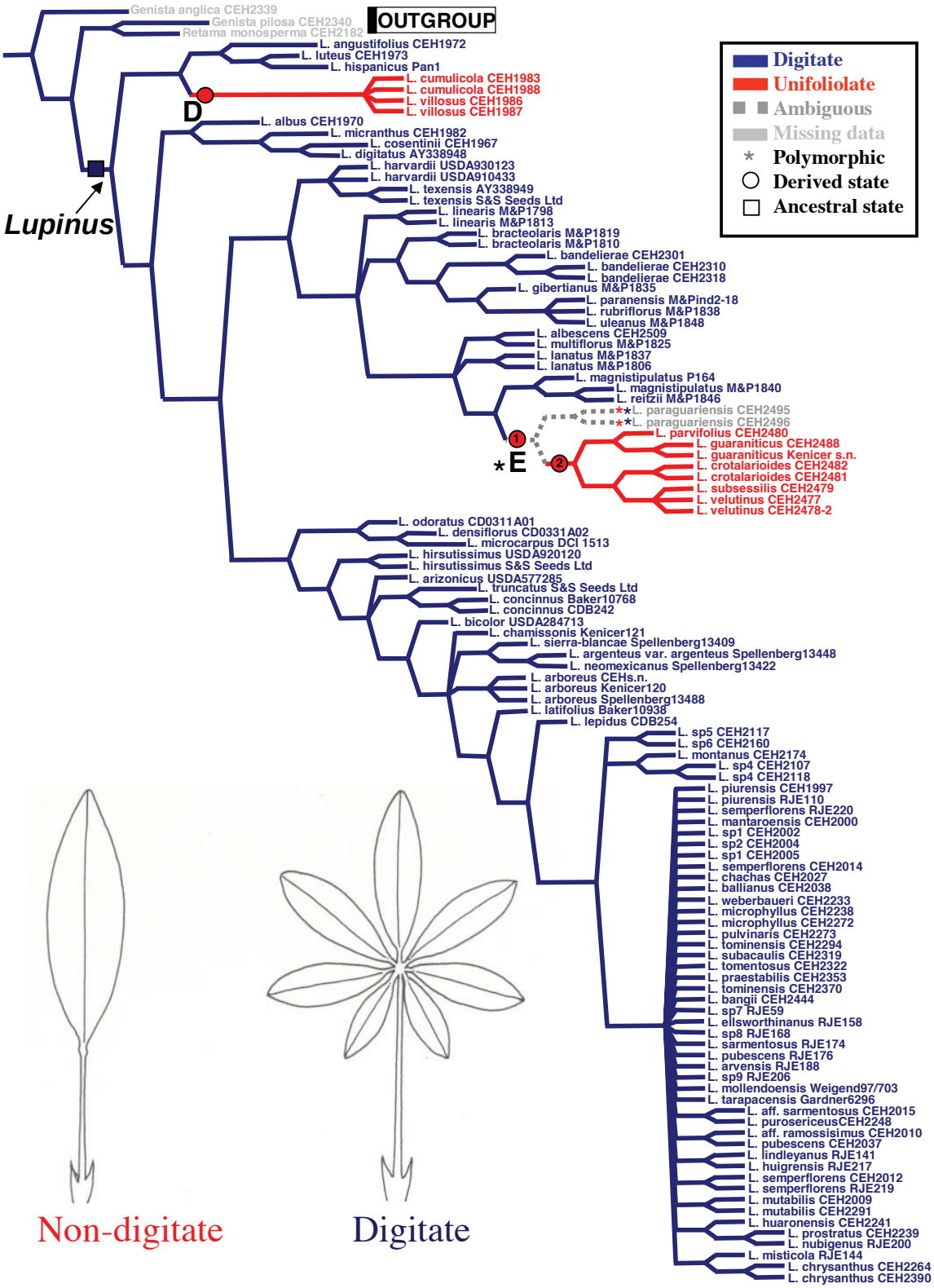


Fig. 6.2 Evolution of metaphylls in Lupinus. Unambiguous optimisation of metaphyll states onto the *Lupinus* phylogeny. There have been at least two independent transformations from digitate to unifoliolate leaves within *Lupinus* (clades D and E). The position of the transition in clade E is ambiguous due to the polymorphic coding of *L. paraguariensis* and ACCTRAN and DELTRAN optimisations resolve character state placement differently (❶ and ❷ respectively). Placement at position ❷ would infer a third independent origin of the unifoliolate state within *L. paraguariensis*.

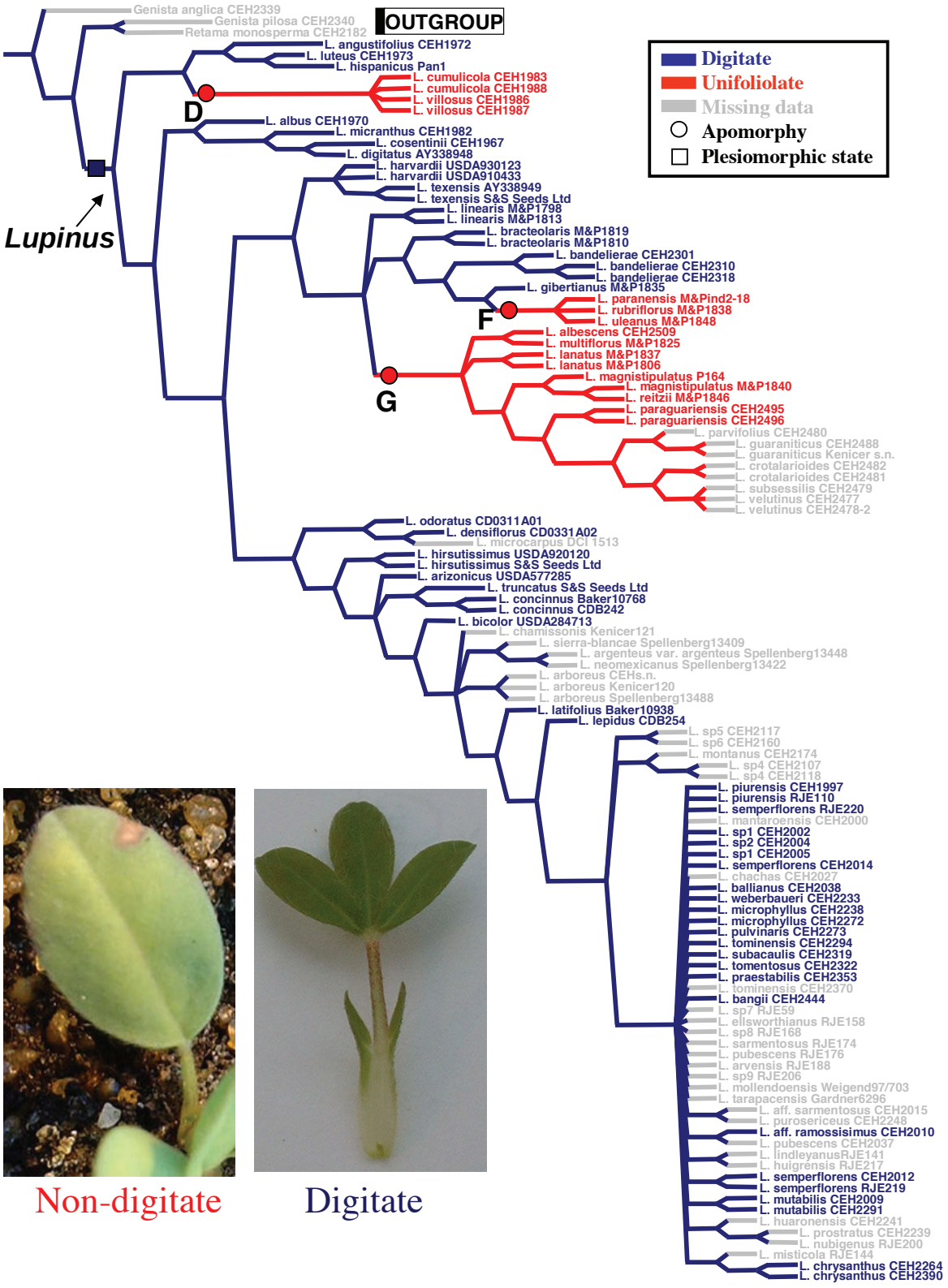
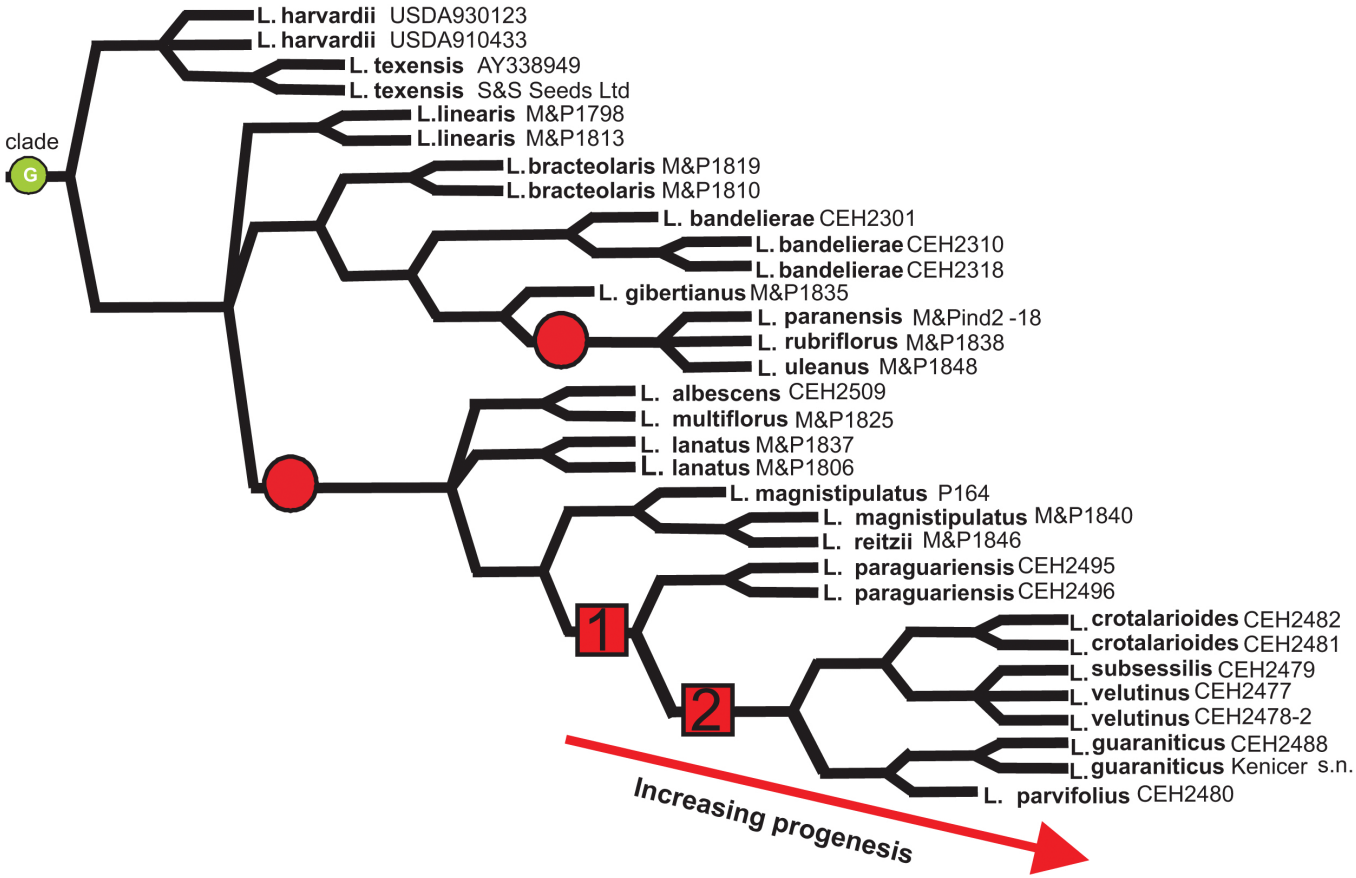


Fig. 6.3 Evolution of eophylls in *Lupinus*. Unambiguous optimisation of first eophyll states onto the *Lupinus* phylogeny**.** The unifoliolate character state is optimised onto the phylogeny in three places (clades D, F and G). The outgroup taxa were not sampled because eophyll data were not available for these species and were scored as missing.



First eophyll character state ○	Metaphyll character state □	Metaphyll petiolate /sessile
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Digitate	Digitate	Petiolate
Unifoliolate	Digitate	Petiolate
Unifoliolate	Digitate	Petiolate
Unifoliolate	Digitate	Petiolate
Unifoliolate	Digitate	Petiolate
Unifoliolate	Digitate	Petiolate
Unifoliolate	Digitate	Petiolate
Unifoliolate	Digitate	Petiolate
Unifoliolate	Digitate	Petiolate
Unifoliolate	Digitate	Petiolate
Unifoliolate	Unifoliolate/Digitate	Petiolate
Unifoliolate	Unifoliolate/Digitate	Petiolate
Unifoliolate	Unifoliolate	Petiolate
Unifoliolate	Unifoliolate	Petiolate
Unifoliolate	Unifoliolate	Petiolate
Unifoliolate	Unifoliolate	Petiolate
Unifoliolate	Unifoliolate	Petiolate
Unifoliolate	Unifoliolate	Petiolate
Unifoliolate	Unifoliolate	Petiolate
Unifoliolate	Unifoliolate	Sessile

Fig. 6.4 Leaf evolution within clade G (Section Orientalis). Optimisation of leaf character states onto clade G (see Fig. 3.9). The unifoliolate eophyll state is optimised twice upon the tree. Unifoliolate metaphylls map once on the tree to a clade nested with the larger unifoliolate clade. The only sessile-leaved species sampled *L. parvifolius* is deeply nested within the unifolioate metaphyll clade. The trend of descreased leaf complexity can be explained by increased progenesis.

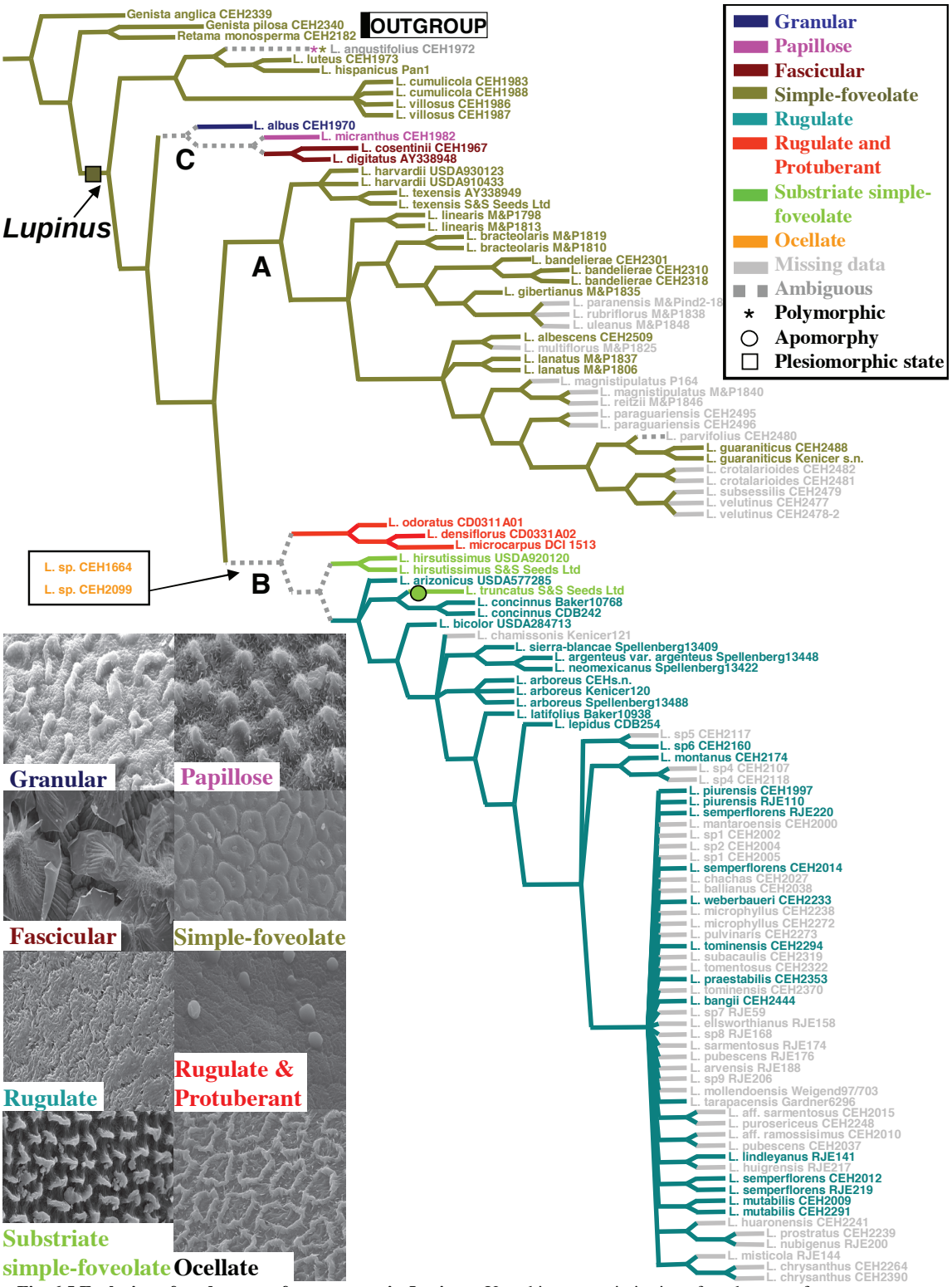


Fig. 6.5 Evolution of seed coat surface patterns in *Lupinus*. Unambiguous optimisation of seed coat surface pattern states onto the *Lupinus* phylogeny suggests that the plesiomorphic seed coat surface state for *Lupinus* is simple-foveolate. Clades A and B differ in their seed coat surface types. Whilst clade A only exhibits simple-foveolate, seed coat surface patterns in clade B are far more diverse. The substriate simple-foveolate seed coat surface pattern is homoplastic, appearing twice on the tree. There are areas of ambiguity at the base of clades B and C. ACCTRAN and DELTRAN optimisations shown in Figs 7.4. The taxa with ocellate seed coat surface patterns were not sampled in the phylogeny study. On the basis of their overall morphology they would be placed within clade B.

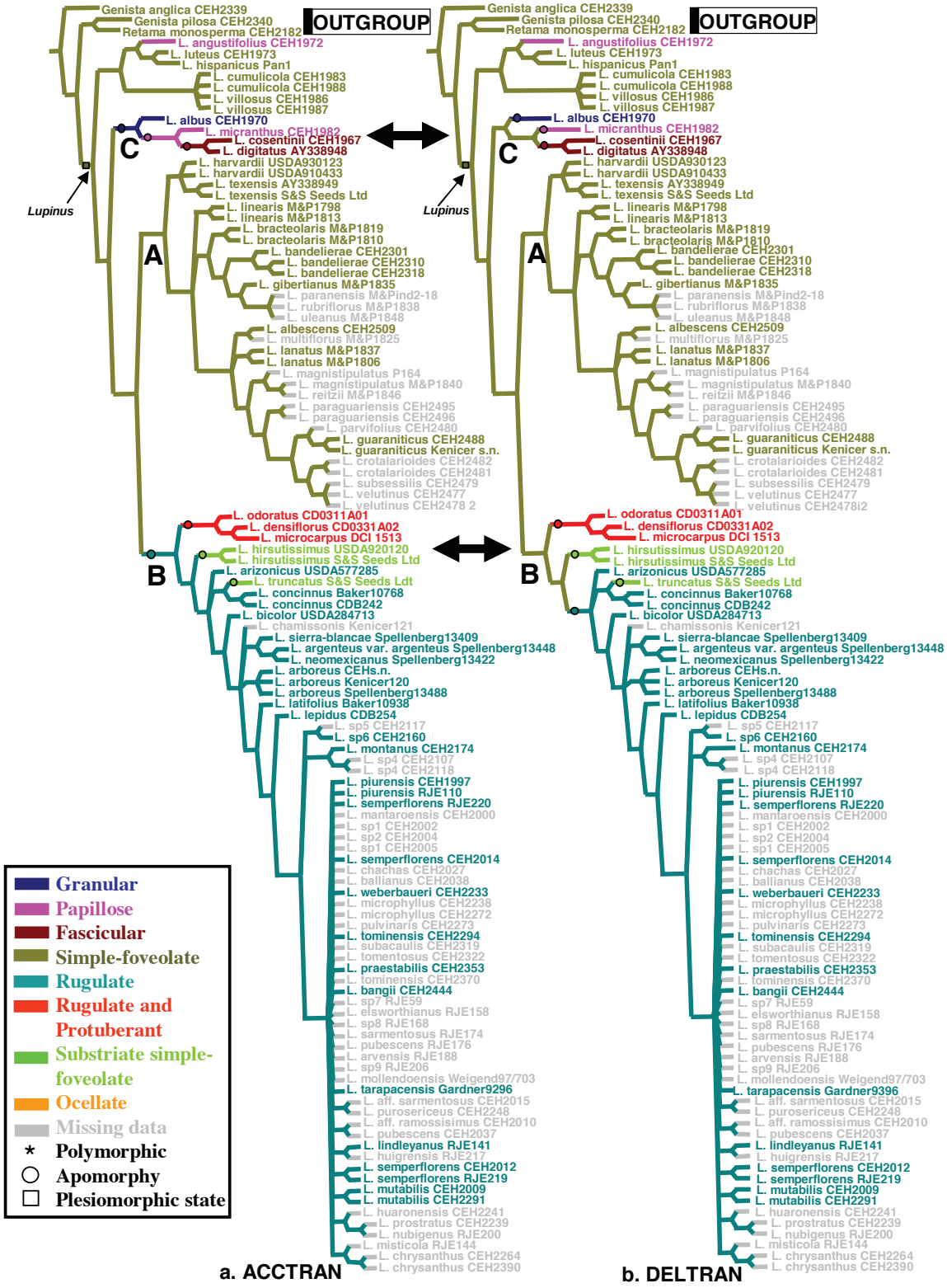


Fig. 6.6 Evolution of seed coat surface pattern in *Lupinus*. ACCTRAN and DELTRAN optimisation of seed coat surface pattern states onto the *Lupinus* phylogeny. There are two areas of the topology (clades B and C, indicated by black arrows) where different optimisations resolve different placements of character states. ACCTRAN (a.) optimisation favours gain of rugulate seed coat surface pattern at the base of clade B where as DELTRAN (b.) favours a later gain of this state. At clade C it is the position of the gain of the granular seed coat pattern which is in question.

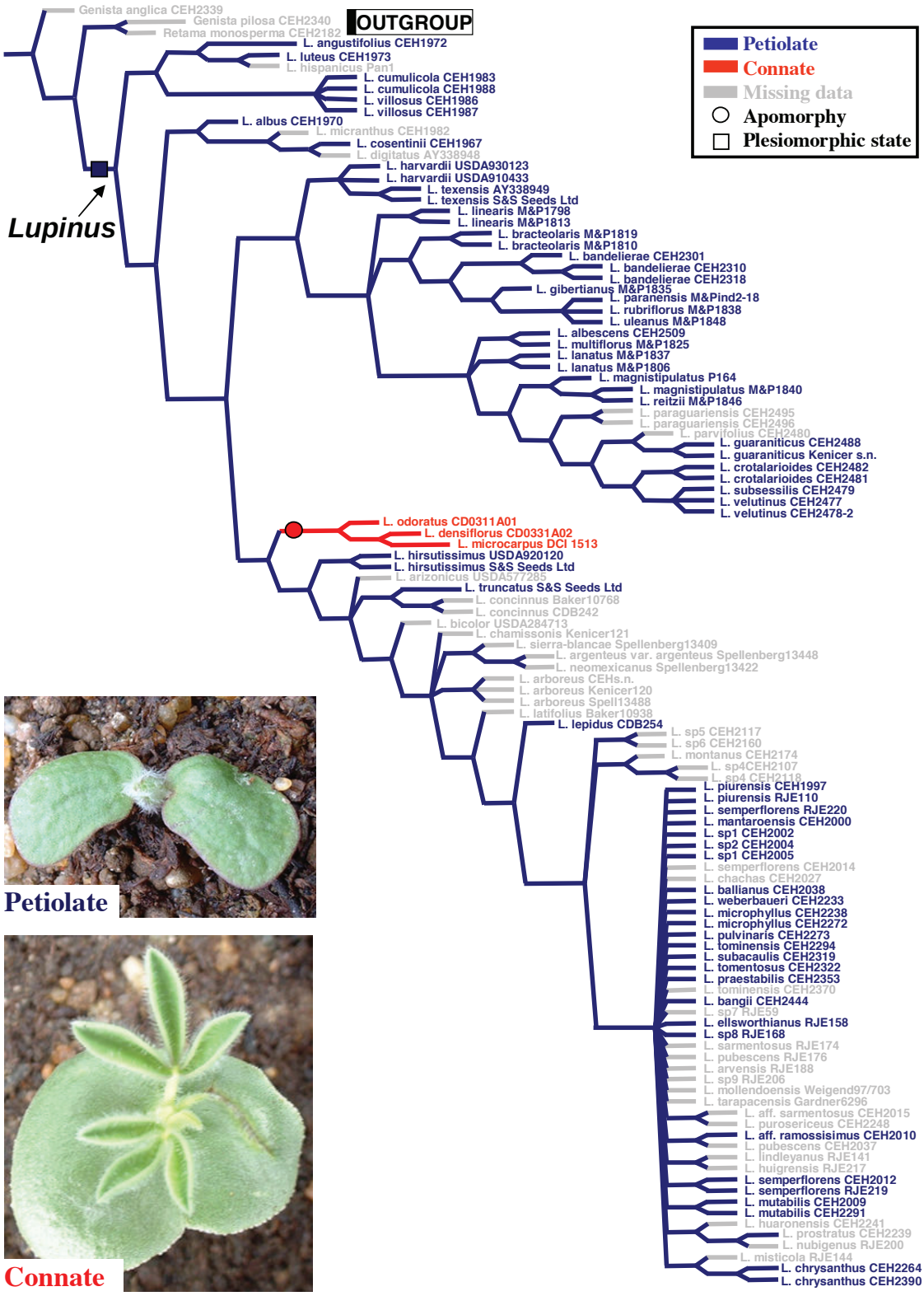


Fig. 6.7 Evolution of cotyledons in *Lupinus*. Unambiguous optimisation of cotyledon states onto the Lupinus phylogeny. The pleisomorphic cotyledon state for *Lupinus* is petiolate. The connate state is derived and placed once (red circle) on the tree, as a synapomorphy for Section Occidentalis, Series Platycarpos (see Chapter 7). Although the outgroup taxa are not coded many Genisteae are reported to have petiolate cotyledons.

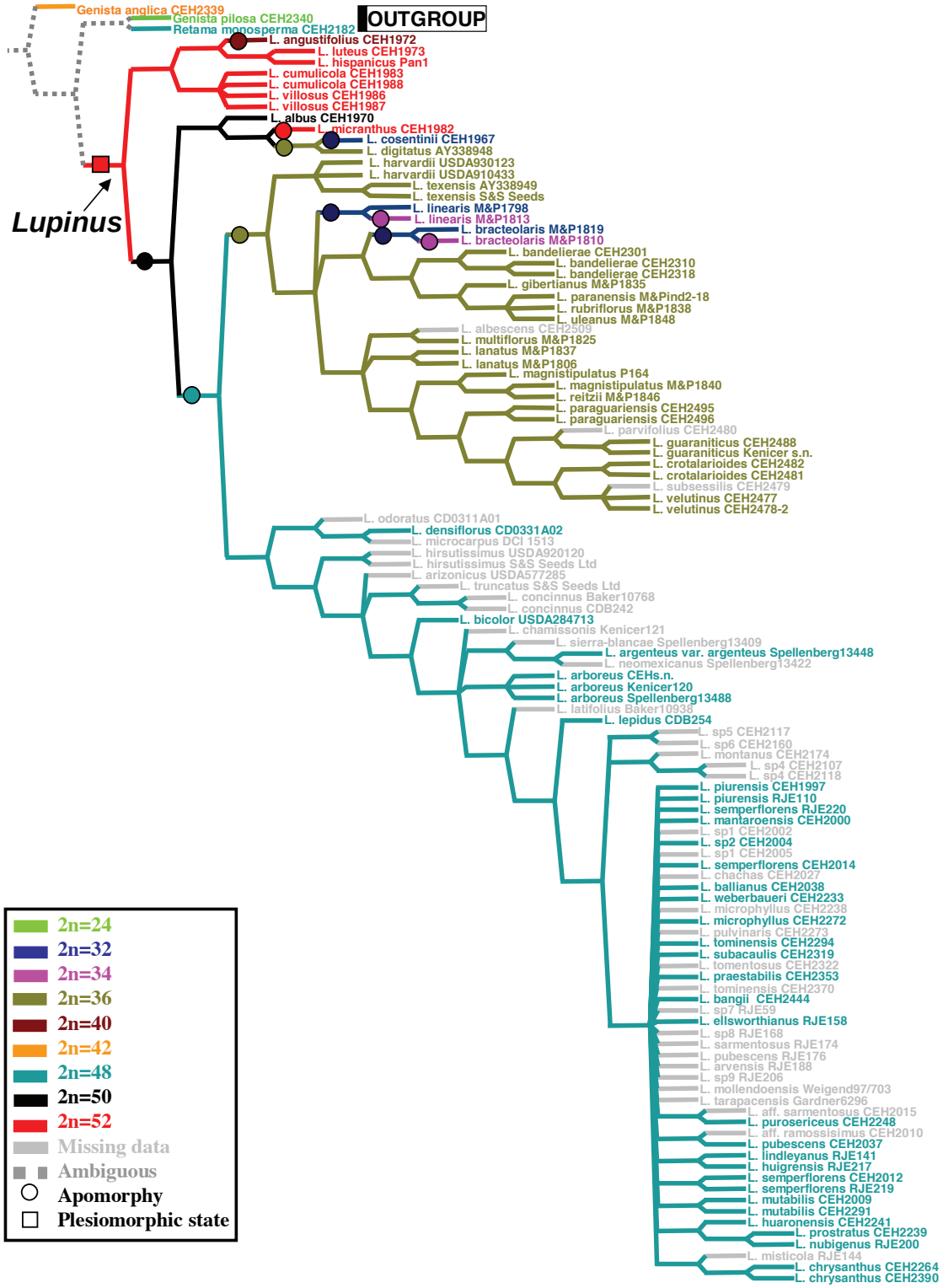


Fig. 6.8 Chromosome evolution in *Lupinus*. Unambiguous optimisation of commonest chromosome number onto the *Lupinus* phylogeny. This reconstructs 2n=52 at the base of the *Lupinus* phylogeny. ACCTRAN/DELTRAN optimisations (not shown) reconstruct the ambiguous region at the root as 2n=52.

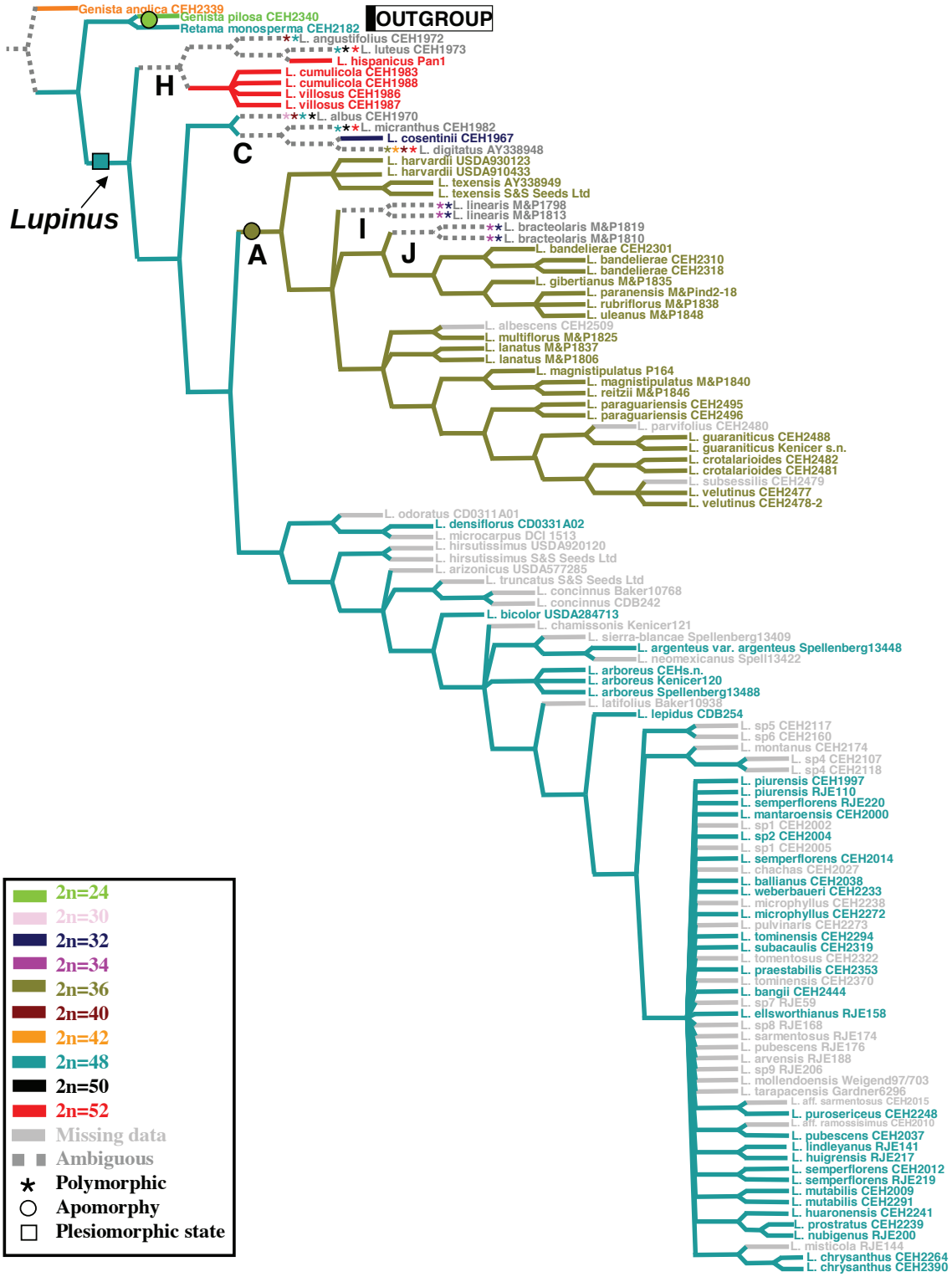


Fig. 6.9 Chromosome evolution in *Lupinus*. Unambiguous optimisation of chromosome number onto the *Lupinus* phylogeny. Terminals were coded polymorphically to represent intraspecific variation in chromosome number. This reconstructs 2n=48 at the base of the *Lupinus* phylogeny. ACCTRAN and DELTRAN optimisations are given in Fig. 7.8. and resolve areas of ambiguous (clades C, H, I and J).

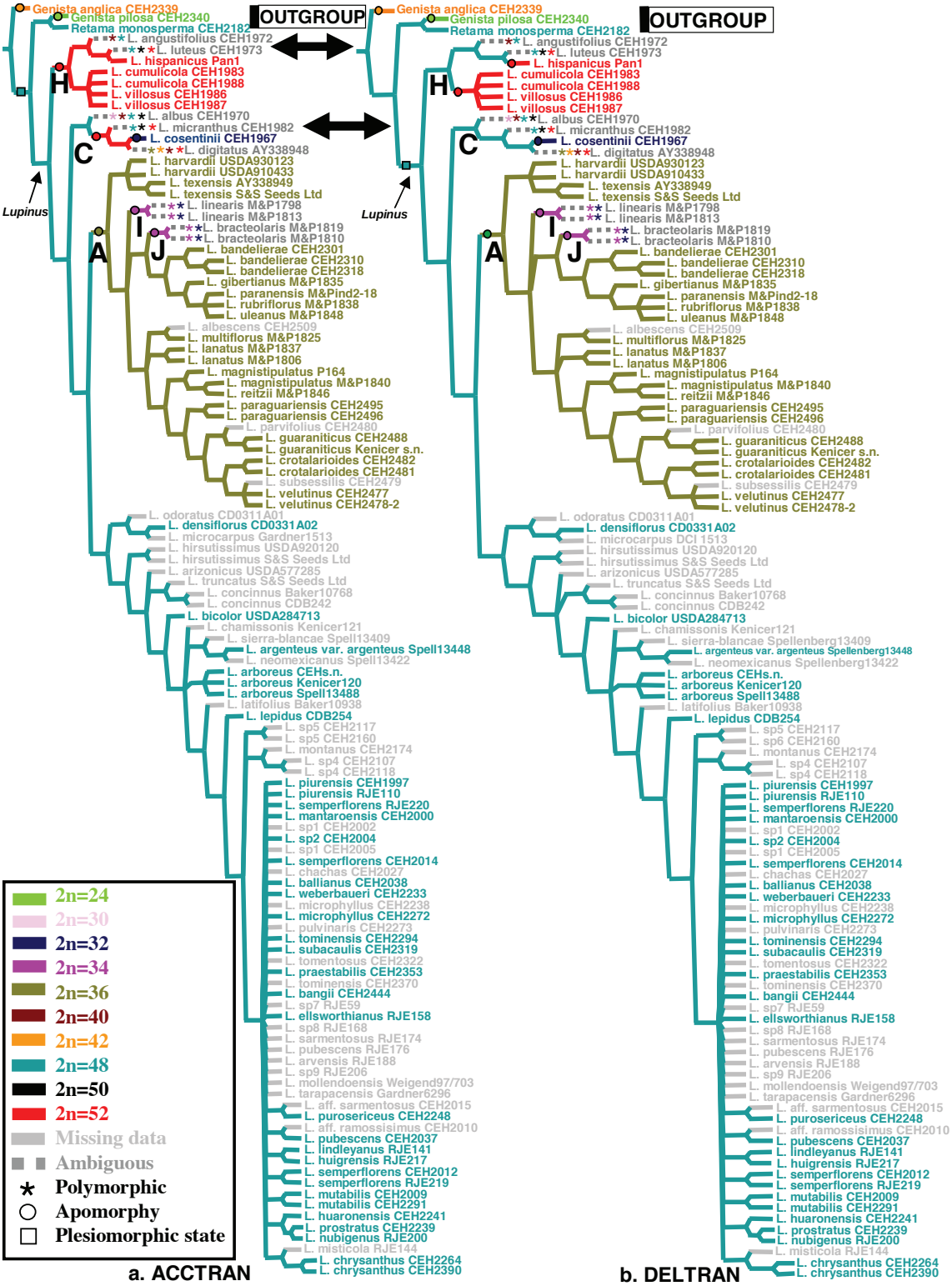


Fig. 6.10 Chromsome evolution in *Lupinus*. ACCTRAN and DELTRAN optimisation of chromosome number onto the *Lupinus* phylogeny. Terminals were coded polymorphically to represent the range of chromosome numbers found in species. These reconstruct 2n=48 at the base of the *Lupinus* phylogeny. Both ACCTRAN (a.) and DELTRAN (b.) optimisations are shown above. These differ in the character placements within clades C and H shown by black arrows).

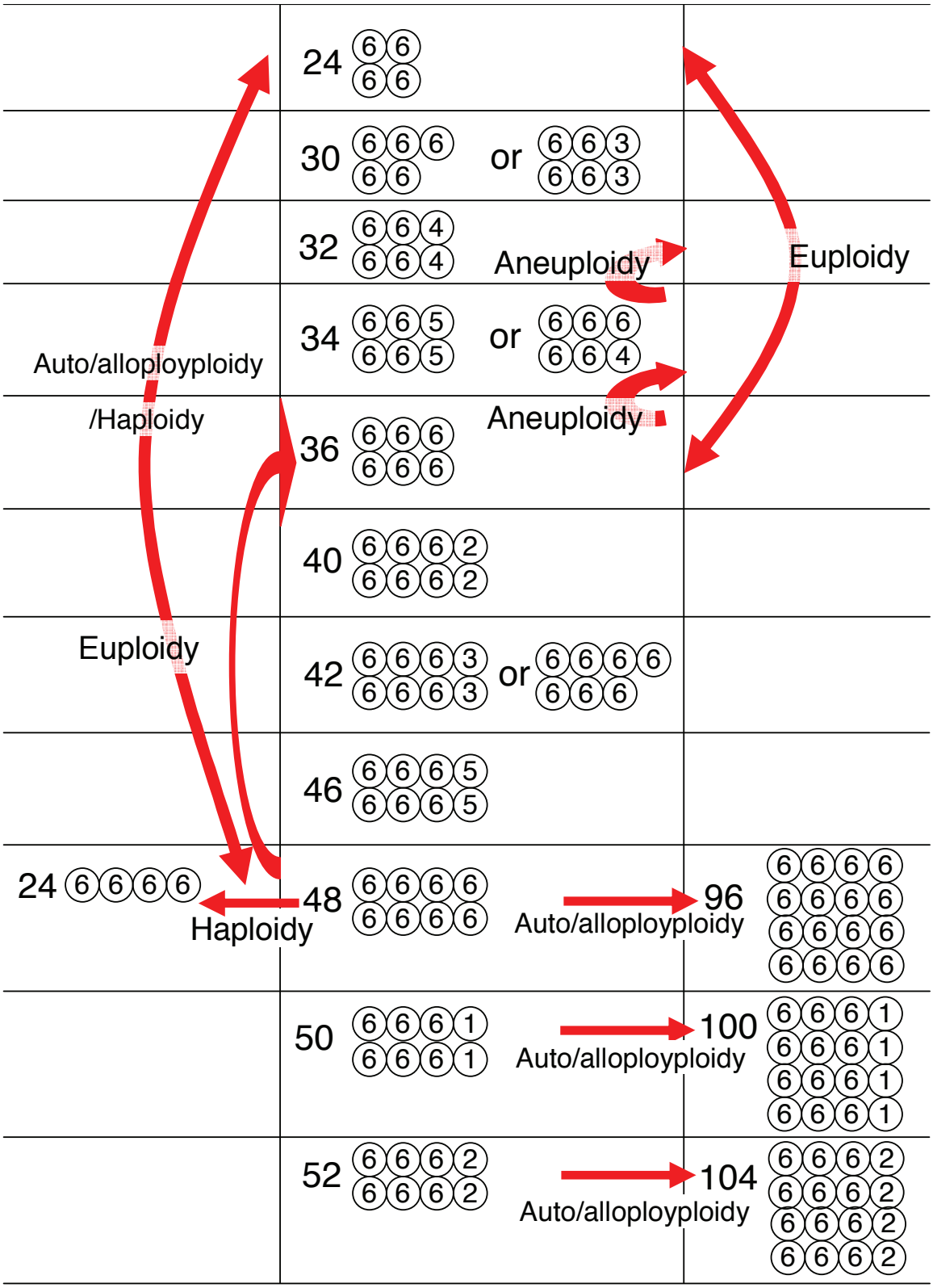


Fig. 6.11 Possible chromosome evolution within *Lupinus*. Chromosome evolution in *Lupinus* is complex. If the monoploid number for *Lupinus* is 6 (represented by (6)) then a combination of mechanisms (euploidy, serial aneuploidy, auto/alloploidy) have to be invoked to explain all the transitions within the genus. However, this is speculation as part of the phylogeny remain uncertain.

7 A new infrageneric classification of *Lupinus*

7.1 Taxonomic history

The monophyly of *Lupinus* has never been questioned since the genus was erected by Linnaeus (1771). *Lupinus* has been considered to be an isolated genus within the tribe Genisteae (Monteiro, 1987b) suggested by a number of unusual characteristics including: palmately digitate (rarely simple) leaves, stipules adnate to the petiole, pollination syndrome, unique root-nodule morphology (Sprent, 2001) and geography (Bisby, 1981). Rouy (1893) first recognised this by erecting the subtribe Lupinées containing only *Lupinus*. This was subsequently duplicated by Bisby (1981) when he erected the subtribe Lupininae. An affinity with *Aygyrolobium* was suggested when Hutchinson (1964) proposed the tribe Lupinieae containing *Aygyrolobium* and *Lupinus*. On the basis of morphology and geography *Lupinus* has most recently been placed with *Anarthrophyllum* and *Sellocharis* in the subtribe Lupininae Rouy (Lewis *et al.*, 2005a). However, this remains speculative because *Anarthrophyllum* has not been resolved as sister to *Lupinus* in any phylogenetic studies to date (Fig. 3.5, Chapter 3) (Lavin *et al.*, 2005) and the affinities of *Sellocharis* are not clearly known.

In the 16th and 17th centuries the first species of *Lupinus* were described from the Old World by Lobel, Clusius and Bauhin (Kurlovich, 2002). However, formal recognition of the genus and these first species, dates to Linneaus (1771) who described five Old World and one North American species. The type species for the genus is *L. albus* (LINN). The only monographic treatment of the genus, the *Synopsis Generis Lupini*, of Agardh (1835) recognised 83 species, including 17 species described as new. Descriptions, discussions (both in latin), lists of specimens examined and literature references are given for each species. Much of the reference material used in Agardh's revision came from Lindley's herbarium, now at CGE. Even at this early date ten of the currently known 13 Old World species were recognised. An indication of the species richness of *Lupinus* in the New World is already apparent with 40 and 20 species documented from North and South America

respectively. Interestingly, several of the New World species, including the South American domesticate *L. mutabilis,* were reported to be in cultivation in Europe. The early occurrence of a large proportion of horticulturally interesting species may be due to Lindley's link with the Royal Horticultural Society (amongst other positions he was Assistant Secretary from 1822) and also reflects how important garden introductions were in our early knowledge of non-European species (Walters, 1961).

Since Agardh's monograph several other authors have presented infra-generic classifications for the genus as a whole (Ascherson & Graebner, 1907; Gross, 1986; Kurlovich & Stankevich, 2002; Maisurian & Atabekova, 1974; Taubert, 1894). Many other classifications have focused on regional subsets of species to provide partial classifications (e.g. Old World Ainouche *et al.*, 2004; or Brazil, Bentham, 1859). This fragmentary approach reflects taxonomic work more generally within *Lupinus,* which has been on a geographic rather than monographic basis. Appendix 18 lists the published infrageneric classifications of *Lupinus*. A comparison of the infrageneric groups in those classifications which have considered the whole genus is presented in Figure 7.1. This demonstrates the striking lack of consensus between different classifications.

Maisurian and Atabekova's (1974) "The Lupin" was published in Russian and has never been translated. Their classification however can be interpreted and shows strong similarities to Agardh's system (Fig. 7.1). It would seem, from their illustrations, that this book has extensive discussions on many species, cytology and mutations and it lists species country by country with some synonymy but much of this remains inaccessible in the absence of a full translation. What is clear is that Maisurian and Atabekova used finely differentiated infraspecific taxa (up to seven) to delimit different flower colour variants within species.

Recently another Russian, Kurlovich (2002) published a book entitled "Lupins: geography, classification, genetic resources and breeding". Translated into English, this book made Russian lupin research accessible to non-Russian speakers for the first time. Kurlovich

pulled together data from a large number of disparate sources. Unfortunately, none of the early phylogenetic data were incorporated into this synthesis. This partially explains the differences between the infrageneric classification proposed by Kurlovich & Stankevich (2002) and that presented here which is directly based on a new phylogeny. Furthermore, the book fails to acknowledge that revision of *Lupinus* is still in the early stages and that far more work is needed before a comprehensive reference text can be produced. It also suffers from numerous typing and translational errors which detract from it's quality and usefulness.

There are ca. 1391 species names (ca. 1931 names in total including infraspecific names), published by more than 200 different authors, associated with the ca. 290 species of lupin. However, almost 40% of these names are attributable to one particularly prolific life-long lupin enthusiast Charles Piper Smith (see below). The other taxonomists who described substantial numbers of new species were Edward Greene (85), Amos Heller (69), Alice Eastwood (55), Sereno Watson (36), George Bentham (35), David Douglas (32), Per Rydberg (30), Jakob Agardh (25), Joseph Rose (25), Asa Gray (24), John Torrey (24), Joseph Rose (18) and Oskar Ulbrich (17). Although ~100 names were published before 1900 the vast majority of names were published in the 20th century. Perhaps surprisingly given the very large number of existing names, new species are still being described e.g. *L. anatolicus* (Swiecicki *et al.*, 1994), *L. jamiehintoniana* (Turner, 1995) and *L. mariae-josephi* (Pascual, 2004).

Charles Piper Smith, a forester by training, devoted much of his life to the study of North and South American lupins and between 1938 and 1953 privately published most of his work in a series of papers forming his *Species Lupinorum*, (Smith, 1938-1953; Stafleu & Cowan, 1976). He also contributed to the Flora of the Pacific States and the first Jepson Manual (Smith, 1944b; 1951). His approach to the taxonomy of *Lupinus* has caused considerable taxonomic chaos (Barneby, 1990; Standley & Steyermark, 1946) because not only was his work generally focused within political boundaries and lacked synthesis but

also because he was a taxonomic splitter to the highest degree. Barneby (p 257, 1990) highlighted this problem stating "*Undue emphasis on differences between these minor races leads inexorably to the quicksands that ultimately engulfed Charles Piper Smith*". These problems are immediately apparent when type specimens, which he recognised as separate species, are examined. In the space of 34 years Smith described some 675 species and infraspecific taxa, many of which have been sunk into synonymy and many more of which are destined to follow (Hughes & Eastwood unpubl. data). In fact Smith's names in general can be equated with type specimens rather than species. Throughout *Species Lupinorum* and the various flora accounts that he wrote, Smith also recognised innumerable informal groups. The circumscription of these groups was inconsistent and I have not attempted to reconcile these with any other recognised classifications. It is widely recognised that, at least in later years, Smith suffered from senility which may account for some of his erratic and unjustified decisions (Thomas, 1974). While classificatory efforts proposed by Smith can readily be discounted, the huge surfeit of names is a legacy that has burdened all subsequent lupin taxonomists and that 50 years later still remains largely unresolved.

Due to the wide distribution and large number of species in the genus many workers have created partial and often informal infrageneric classifications encompassing species from one geographic region. Broad consensus between the classifications with regard to the Old World species has been reached. These species have been intensively studied and revised partly because they include a number of commercially important domesticates (*L. albus*, *L. angustifolius* and *L. luteus*) (Gladstones, 1974, 1984), but also because the Old World lupins were among the first species known and are naturally distributed in and around the Mediterranean, the key centre for early taxonomic work (Walters, 1961), and where they were easy to collect. Agardh (1835) was aware of the use of *L. albus* as a food and also noted some medicinal uses of lupins. Informally, the Old World species have been designated as smooth- or rough-seeded (Gladstones, 1974). The rough-seeded species have also been formally recognised as Section Scabrispermae (Plitmann & Heyn, 1984).

The formal classification of New World *Lupinus* has been dealt with in two contrasting ways, extreme lumping or extreme splitting. Firstly, Agardh (1835) along with Piper and Robinson (1906), Rydberg (1917), Smith (1938-1953; 1944b), (Maisurian & Atabekova, 1974) and Sholars (1996) divided New World lupins into numerous small groups defined by combinations of mainly quantitative characters. This framework was not successful because authors disagreed not only about which groups to recognise but also the circumscription of these groups. Geographically based, piecemeal efforts at classifications inevitably result in inconsistencies in circumscription and naming of groups, especially when the character combinations used to define groups are based on continuously variable traits. Alternative much simpler classification schemes were devised by Bentham (1859), Watson (1873), Taubert (1894), Ascherson & Graebner (1907), Gross (1986) and Kurlovich & Stankevich (2002) who recognised a maximum of five subgroups of New World species distinguished by one or a few clearly circumscribed discrete characters. Despite this, there is still considerable incongruence between these systems (Fig. 7.1 and Appendix 18).

Dunn and his students worked extensively on North American lupins, grouping them into informal 'complexes'. Dunn suggested that delimitation of New World lupins is difficult not only because there are many morphological continuities but also because putative interspecific hybridisation is widespread. Dunn also worked with Plancheulo on the lowland South American species (Dunn & Planchuelo, 1981; Planchuelo & Dunn, 1984, 1989). Their works alongside that of Monteiro and Gibbs (1986a) and Pinheiro and Miotto (2001), have clarified much of the taxonomy of the lowland South American species. Barneby (1990) started work on clarifying both species delimitation and synonymy of the North American lupins. In his treatment of lupins for the Intermountain Flora he took a broader view of species exemplified by the fact that 83 names (including subspecific taxa) were synonymised within *L. argenteus*.

Some New World infrageneric groups appear in more than one classification system (Appendix 18) e.g. Arborei, Caespitosi, Concinni, Leucophylli, Microcarpi, Polyphylli,

Pusilli, Saxosi, Sericei, and Sulphurei, and others are unique, including the segregation of *L. uncialis* in a monospecific section by Watson (1873). One factor contributing towards the inconsistent circumscription and nomenclature of North American groups is differential sampling resulting from the different distributions of species in America. An example of inconsistent circumscription is the placement of *L. ornatus* and *L. leucophyllus* along with 13 North American species in his Series Sericei by Agardh (1835). Rydberg (1917) however placed these in his informal group Ornati with six different species. Even placement of the highly distinctive *L. microcarpus* has been highly unstable in the face of competing classification schemes. It has been placed in group Lutei (Agardh, 1835), Section Platycarpos (Watson, 1873), Subgenus Platycarpos (Piper & Robinson, 1906) and informal group Microcarpi (Smith, 1944b). However, it is extremely difficult and in some cases impossible to compare these competing classifications systems due to the chaotic taxonomy, use of the same specific names for more than one entity and failures to state clearly which species belong to which group in much of the literature.

A further inconsistency amongst previous infrageneric classifications is that many authors did not specify formal ranks or type specimens for their infra-generic groups (Agardh, 1835; Bentham, 1859; Rydberg, 1917; Smith, 1938-1953). For those who did, ranks varied from subgenera (Piper & Robinson, 1906) to sections (Plitmann & Heyn, 1984; Taubert, 1894; Watson, 1873) and most involved more than one hierarchical level.

Thus, the infrageneric classification of *Lupinus* remains very confused and a practical classification reflecting evolutionary relationships is much needed. *Lupinus* is a relatively large genus of ca. 290 species, for which some sort of internal taxonomic structure is required to facilitate identification and information retrieval. While, many attempts have been made to provide an infrageneric classification for *Lupinus*, these are unsatisfactory because they have been generally incomplete, often over-differentiated, lacking diagnosability, often rely on informal groups and none have been based on a well-supported hypothesis of species relationships.

7.2 Data and characters used in previous infrageneric classifications

All previous infrageneric classifications of *Lupinus* have been based solely on morphological data except for one classification of the Old World species in which an ITS gene tree was taken into consideration (Ainouche *et al.*, 2004). The morphological approach has contributed to discordant classifications because of lack of consensus over which characters to use. The majority of morphological variation within *Lupinus* is quantitative, e.g. size of parts, number of leaflets. Even some superficially qualitative characters (e.g. petal shape) are essentially quantitative in nature (Stevens, 1991). These characters show continuous variation that is difficult to partition into discrete unambiguous states. Both quantitative and qualitative characters have been used to define infrageneric groups within *Lupinus* but the most commonly and effectively used characters have been the qualitative ones e.g. cotyledon type, leaf type, and seed coat (see Chapter 5 for a description of these).

The recognition that only a small set of characters from the superficially high number of potential morphological characters are useful, i.e. those showing fixed differences between species or groups, has long been highlighted for *Lupinus*. A similar situation is found in *Hibiscus* (Pfeil & Crisp, 2005). However, more characters are useful to delimit species than groups. As early as 1835 Agardh commented that species delimitation within *Lupinus* was difficult. Watson (p 518, 1873) commented that: "*It will be seen, however, that there are three well-marked sections, the largest of which is conveniently divided into perennials and annuals. In minor subdivisions, use is made mainly of differences in habit and in the number of ovules, which are found to be in most cases very constant. Little use is made of the characters of the seed, the differences in the calyx are mostly slight and not easily expressed, the color of the petals is often variable and changeable and in dried specimens uncertain, and peculiarities of pubescence or foliage, which in some cases seem very constant are not to be relied on*." Smith (1918; p 3, 1919) also discussed the diagnostic characters for species delimitation within some of the infra-generic groups he recognised. He noted: "*reliable diagnostic characters are not as readily selected as I expected them to be*." and discussed how environmental conditions can influence morphology, attributing variation in height, branching, caulescence/acualescence and fistulous/non-fistulous characters in

part to phenotypic plasticity. Characters exhibiting phenotypic plasticity were disregarded by Smith for species delimitation as they are not fixed within a group. On the same basis Smith (1918; 1919) stated that he further disregarded petal colour and shape, pubescence, stem type, leaf, stipule and floral bract length for species delimitation. In practise however he did not adhere to these principles.

Seed coat type (Chapters 6 and 7) is one of the few useful characters within *Lupinus* although observations have focused mainly on Old World species where macromorphologically rough and smooth seeds are found. The rough seed coat type (Fig. 6.5) provides a unique synapomorphy for the rough-seeded Section Scabrispermae (Plitmann & Heyn, 1984). The term Malacospermae was first used by Kass and Wink (1996) to represent the smooth-seeded Old World taxa. It is not a formal name for this group but has subsequently been used by Kass and Wink (1997a) and Naganowska *et al.* (2006). Heyn and Herrnstadt (1977) first recognised the heterogeneous nature of the smooth-seeded group at the micromorphological level. The four smooth-seeded Old World groups each show different seed coat micromorphologies. These five seedcoat groups are congruent with well-supported clades in a series of phylogenetic analyses of ITS sequences (Fig. 3.5, Chapter 3) (Ainouche *et al.*, 2004), trnS-G, trnT-trnL and trnL intron sequences (Drummond, 2004 and unpubl. data), LEGCYC1A (Fig. 3.8, Chapter 3) and GPAT1 (Fig. 3.7, Chapter 3) as well as the combined 3-locus phylogeny (Fig. 3.10, Chapter 3). Until now seed coat surface pattern has not been investigated for more than a handful of New World species (Chapter 5).

A second major character, leaf type, has been used, ever since Agardh's (1835) first classification of *Lupinus* into groups of digitate-leaved and non-digitate-leaved (see discussion Chapter 6) species. This character has been used in most subsequent systems accounting for the whole genus (Gross, 1986; Taubert, 1894) and those focusing on geographic regions such as Brazil, where both digitate and unifoliolate species occur (Bentham, 1859). Exceptions to this are Watson (1873), Ascherson & Graebner (1907), Kurlovich & Stankevich (2002), whose classifications did not separate unifoliolate from digitate-leaved species (Fig. 7.1).

A third discrete character, first used by Watson (1873) to delimit the *Platycarpos* of *L. microcarpus* and allies is cotyledon type (Chapter 6). Species within this group exhibit connate cotyledons whereas all other *Lupinus* species have petiolate cotyledons.

A fourth character, life history, and specifically annual or perennial habit, has also been widely implemented to divide *Lupinus* into infrageneric groups. However, this is not a clear cut discrete character. One problem is that life history traits are affected by environmental conditions. A second problem is the difficulty of dividing life history into discrete character states. Biennials and short-lived perennials are found as well as annuals and true perennials. This suggests that life history can not be coded subjectively into character states.

A final character utilised for infrageneric classification was the number of ovules per ovary. Although a discrete character it is not constant within all species. Many authors used this character to support *L. microcarpus* and allies, as these species have just two ovules per ovary (Ascherson & Graebner, 1907; Piper, 1906; Rydberg, 1917; Smith, 1946; Taubert, 1894; Watson, 1873). Kurlovich & Stankevich (2002) however, divided the genus into two subgenera circumscribed by a combination of characters including two, four or more ovules (Subgenus *Platycarpos* - all New World species) vs. four or more ovules (Subgenus *Eulupinus* - all Old World species). However, the ovule character is not diagnostic for these groups under this classification.

7.3 Criteria for classification

7.3.1 Natural versus artificial classification

Classifications can be natural or artificial depending on whether they aim to represent or at least reflect evolutionary history or not (Orthia *et al.*, 2005). Natural systems can claim to be predictive for a wide range of characters because members of groups share common evolutionary histories. Artificial systems are only predictive about the specific character or characters upon which they are built. Almost all modern methods aim to build natural

classifications. Indeed, this has been the aim of many historical classifications, but our ability to build more accurate phylogenies based on DNA sequence data in recent years means that classifications can now be based on much more explicit phylogenetic hypotheses.

7.3.2 Diagnosability

Diagnosability, the ability to recognise a group using diagnostic morphological, or at least non-molecular, characters is a desirable property of any formally recognised group. Diagnosable characters allow placement of an individual or taxon in a group without the need to derive molecular information. Diagnosability can rely on a single unique character, several unique characters or a unique combination of characters (Helbig *et al.*, 2002). In general, to be practical in the field and herbarium, useful diagnostic characters that can be readily assessed (e.g. macromorphological) are preferable. However, molecular data is necessary to recognise the monophyletic or paraphyletic nature of the taxa (Pfeil & Crisp, 2005).

There has been surprisingly little discussion of "cryptic clades", i.e. those which are resolved with high statistical support in molecular datasets but which are not diagnosable using non-molecular characters. Most taxonomists are reluctant to recognise these clades in formal classifications but have sometimes acknowledged their presence with informal names (Richardson *et al.*, 2000; Wojciechowski *et al.*, 1993). In some cases re-examination of morphology, in the light of knowledge of a well-supported tree based on molecular data, has led to the discovery of previously unrecognised non-molecular synapomorphies for such clades (Lavin *et al.*, 2001a).

7.3.3 Monophyly

Since the rise of cladistics, monophyly has been widely advocated as the most important criteria for classification because it is non-subjective (Backlund & Bremer, 1998). In cladistic theory natural groups are monophyletic groups, i.e. those that contain all and

only the extant taxa which share a common ancestor (Forey, 1992). Only classifications based on monophyly can truely represent evolutionary history. Monophyly is tested by the construction of cladograms based on morphological and/or molecular data and monophyletic groups are those which share derived characters (synapomorphies) and are thus visualised as clades in cladograms. Statistical confidence in a clade can be tested objectively by bootstrap or jacknife methods in parsimony or likelihood analysis or via posterior probabilities in Bayesian analysis. However, the potential incompatibility of the Linnaean classification system and strictly monophyletic classifications has frequently been pointed out (Brummitt, 1997, 2002, 2003; de Queiroz & Gauthier, 1992, 1994).

Classification based on monophyletic groups in the Tree of Life is a major goal of taxonomic research (Cracraft & Donoghue, 2004; Patterson, 1982a; Sosef, 1997). DNA sequence data are most commonly used to construct cladograms for this purpose (Scotland *et al.*, 2003). As a method for identifying non-monophyletic groups this has been highly successful for many groups (e.g. Archibald *et al.*, 2005; Carine & Scotland, 2002; Kim & Sun, 2004; Lavin *et al.*, 2001a; Mols *et al.*, 2004; Wang *et al.*, 2004). For the most part, established infrageneric classifications are monophyletic (e.g. Chen *et al.*, 2005; Kenicer *et al.*, 2005), and establishment of new classifications or altering existing systems to ensure monophyly is less common at the infrageneric level than higher ranks, but examples of new infrageneric classifications include *Fritillaria* (Ronsted *et al.*, 2005), *Orchis* (Bateman *et al.*, 1997) and *Incarvillea* (Chen *et al.*, 2005). A cautious approach to incorporation of knowledge of non-monophyly into formal infrageneric classifications, often stems from concerns over poor taxon sampling (e.g. Alice & Campbell, 1999) and/or lack of resolution and weak support and thus formal infrageneric taxonomic decisions remain pending until further data are generated.

The perceived prominence of monophyly has meant that other criteria traditionally used in classification, such as information content and diagnosability are little considered. This may stem from the fact that most phylogenies built to test monophyly are based on DNA

sequence data with consideration of morphological data a secondary objective. Carine and Scotland (2002) provided an exception to this as they discuss the importance of monophyly, diagnosability and taxonomic structure (implementation of several hierarchical ranks) when proposing a classification for Strobilanthinae. Less consideration of other criteria may also be due to the presence of existing classifications which present a hypothesis to be tested for monophyly and a framework to be modified. This can be achieved using molecular data alone without the need to consider morphological characters, informativity and diagnosability.

7.3.4 Information content

The information content of any groups circumscribed is another attribute of classification to be considered. Information content is a combination of the number of synapomorphies and the number of taxa in a group. A group with three synapomorphies will be favoured over that with one. Backlund & Bremer (1998) equated information content with phylogenetic relationship only. However, a broader definition that incorporates other information e.g. character states and biogeography into classifications is preferable.

Some authors have suggested that all information from a cladogram should be available in a classification and that a cladogram can be reconstituted from a classification (Humphries & Funk, 1984). In this way a classification can convey information about relationships of groups without reference to a cladogram. The ease of achieving this goal depends upon the complexity and size of the group in question. In some cases this could be achieved succinctly whereas in others it may require the addition of many informal groupings which may complicate the classification. This approach is not commonly implemented in practice.

7.3.5 Stability

Stability, of names and their meanings is desirable because it decreases potential confusion for users. However, it is obvious that complete stability in classification would be nonsensical

as it would preclude improvement in the light of increasing knowledge (Schuh, 2003). To ensure a certain level of stability, precedence is given to the first name published for a particular group. It is not always practical to retain an older name but this criterion can at least be taken into consideration when classifications are re-evaluated. The issue of stability is most pressing when well-known agricultural, horticultural or medicinal species are involved.

Stability can also depend on the data e.g. the likelihood of relationships changing with further data. Measures of statistical support are often interpreted as indicators of the accuracy of a tree or true likelihood of a clade, but these values relate only to the data in hand. Stability can be maximised by acknowledging the possibility that phylogenetic hypotheses can be progressively improved with more data and adopting a conservative approach which recognises only clades which are well supported by many characters.

7.3.6 Rank

Classification systems can employ several hierarchical levels or ranks. Family and genus are mandatory in the Linnaean system but additional ranks e.g. tribes, sections and series are optional. Ranks are arbitrarily specified but in Linnaean classification nested groups are required to hold a lower rank than the group within which they are nested (Brummitt, 2002). Under Linnaean classification rank is denoted by a formalised set of word endings i.e. –aceae for the rank of family (Stace, 1980). The number of ranks recognised in a classification depends on personal preferences as well as monophyly, diagnosability, informativity and size. For example, a genus containing only a few species is unlikely to require an infrageneric classification whilst one containing more than 500 species could merit an infrageneric classification comprised of several levels.

7.3.7 Linnaean classification versus phylogenetic nomenclature

The system which has provided the bedrock of biological classification for the last 250 years, allowing biological information to be stored and retrieved, is referred to as the Linnaean system even though it pertains to few of the classification rules outlined by Linnaeus (Moore, 2003). The most important feature of the Linnaean system is that hierarchical representation of relationships is reflected in nomenclature. Other key features are the mandatory specification of type specimens, binomial species names, name endings, that all taxa must be accounted for at every rank and additionally taxon diagnosis and nomenclatural priority within rank. This system, in plants, is governed by the International Code of Botanical Nomenclature (Greuter, 2000). Similar sets of rules exist in zoology (Internation Commission for Zoological Nomenclature, 1999) and bacteriology (Lapage *et al.*, 1976). More recently with the explosion of new phylogenies particularly from sequence data an alternative system of classification, phylogenetic nomenclature, has been proposed (http://www.ohiou.edu/phylocode/) (de Queiroz & Gauthier, 1990), generating considerable debate within the taxonomic community.

Phylogenetic nomenclature arose from the view that the existing system is not sufficient to name all clades and species. The emphasis of phylogenetic nomenclature is naming clades in phylogenetic trees by explicit definitions involving specifiers (species, specimens or apomorphies). It is a rank-free system governed by a set of rules known as the Phylocode. This means that spelling is constant and does not imply rank. No rules for naming species exist to date due to considerable debate amongst the proponents of phylogenetic nomenclature but proposals include uninominal names (Cantino *et al.*, 1999). In contrast to the Linnaean code where only mutually exclusive groupings of the same rank are permitted, phylogenetic nomenclature allows names to be applied to partially overlapping clades. Both systems recognise the need for one recognised name for each taxon and invoke precedence of names with the caveat that these can be overturned in special circumstances. In phylogenetic nomenclature precedence is not restricted to rank as in the Linnaean system. The advocated benefits of phylogenetic nomenclature are increased stability, the opportunity to name

clades as they are found and the elimination of subjective ranking. However, phylogenetic nomenclature is still in it's infancy and not yet governed by a complete set of clear rules.

Naturally these proposals have sparked debate surrounding the perceived flaws in the Linnaean system (Baum *et al.*, 1998; Moore, 1998; Nixon *et al.*, 2003; van Welzen, 1998; Wheeler, in press and references therein.). One of the main issues is the loss of information content in phylogenetic nomenclature due to loss of ranks and specific rank endings. This would be most apparent if species names were to lose their generic names, a proposal mentioned in a recent debate of the phylocode (Wheeler, in press). There is also a question of which phylogeny estimate should be used to base clade names on and a constant reliance to consult "the phylogeny" to understand the content of any group (Nixon *et al.*, 2003).

The problems with the Linnaean system relate to the potential incompatibility of dividing a phylogeny into mutually exclusive monophyletic taxonomic groups (Brummitt, 2002). Under the current rules undesirable non-monophyletic groups that are created as a bi-product of naming nested clades need to be named (Nixon *et al.*, 2003; Sosef, 1997). For example, the recognition of subgenus X automatically assigns the remaining taxa not included in subgenus X to subgenus Y. To avoid this difficulty many informal names have been applied to nested clades and some of these are now widely used and accepted. Prominent higher-level examples include the Eudicots and Fabales while within the legumes several large clades the genistoid and dalbergioid groups have been informally recognised in this way (Lavin *et al.*, 2005; Wojciechowski *et al.*, 2004 and further examples therein)**.** Relaxation of the rules governing mutual exclusivity within the Linnaean system, to allow clades to be recognised while not forcing unresolved or non-monophyletic residues to be grouped together as a consequence, would solve this problem (Nixon *et al.*, 2003). Such a solution, incorporated into a much-need modernisation of the Linnaean system, would offer significant advantages over the turmoil that would be created by introducing a new or parallel system such as the phylocode**.**

7.4 Methods

7.4.1 Criteria

In order to construct a new infrageneric classification for *Lupinus* a set of explicit criteria for recognising groups were applied. In order of priority these were diagnosability, monophyly, information content, stability and rank. The order of priority is somewhat different from other workers although all these factors have generally been recognised as important (Backlund & Bremer, 1998; Carine & Scotland, 2002). I have chosen to classify *Lupinus* using the established Linnaean classification system which in this case can be satisfactorily used to convey the most important phylogenetic information whilst being diagnosable, practical, governed by strict rules and providing reference material in the form of type specimens.

The main criterion for a new classification was diagnosability even though this compromised and limited the number of groups that could be recognised. I took this approach because there is only a small set of constant qualitative characters within *Lupinus* which are amenable to objective partitioning into discrete character states. Diagnosability based qualitative characters is lacking across many nodes in the lupin phylogeny. When considering a classification for *Lupinus,* rank was the least important factor although an attempt was made to ensure that groups of the same rank were roughly comparable in terms of phylogenetic position i.e. where possible sister taxa have the same rank. Unlike many other groups, the *Lupinus* phylogeny is relatively straightforward to divide into well-supported diagnosable monophyletic groups and none of the problems of non-monophyletic residues associated with the Linnaean classification system, discussed above, were encountered.

7.4.2 Phylogeny data

Phylogenetic relationships constructed using nrDNA ITS, nDNA LEGCYC1A and GPAT1 individually and in combination formed the basis for the classification. These data are presented and discussed in Chapter 3.

7.4.3 Characters

A survey of non-molecular (morphological and cytological) characters within *Lupinus* (Chapter 5) resulted in only six discrete characters which could be unambiguously delimited and partitioned across the genus. These were: first eophylls, eophyll type, mature leaf (metaphyll) type, cotyledons, seed coat surface pattern and chromosome number. These characters were optimised onto the 3-locus *Lupinus* phylogeny (Figs. 6.2 – 6.10, Chapter 6). With a single exception, nodes which were diagnosable i.e. supported by non-molecular synapomorphies or unique combinations of these, were considered for recognition.

A detailed description of the characters used is given in Chapters 5 and 6.

7.5 Results

7.5.1 Infrageneric classification

The phylogenetic basis and morphological character states that provide diagnosability are illustrated in Figure 7.2. Eight Sections and five Series are recognised:

Lupinus

- Section Albi
- Section Angustifolii
- Section Lutei
- Section Micranthi
- Section Scabrispermae
- Section Boreo-unifolius
- Section Occidentalis
    - § Series Platycarpos
    - § Series Occidentalis
- Section Orientalis
    - § Series Boreo-orientalis
    - § Series Austro-orientalis
    - § Series Austro-unifolius

The geographic distributions of the sections/series is shown in Fig. 7.3.

7.6 Discussion

7.6.1 Diagnosability of *Lupinus* clades

Due to the small number of discrete morphological characters within the genus few clades of the molecular phylogeny can be reliably diagnosed. However, most of the well-supported clades are diagnosable (Fig. 7.2). Some Sections and Series are diagnosed by unique combinations of characters because no one character is sufficient. Geography can also be helpful in distinguishing between closely related clades.

7.6.2 Monophyly of infrageneric groups within *Lupinus*

The topology of the *Lupinus* phylogeny is such that the whole genus can be conveniently partitioned into monophyletic groups (Fig. 7.2). However, clearly this is only possible because not all clades in the phylogeny have been recognised in the classification (Fig. 7.2). For example, the strongly supported Andean clade which is nested within Section Occidentalis Series Occidentalis, and which is not diagnosable by any morphological characters, was not given formal recognition. All of the groups recognised except Series Austro-orientalis are monophyletic and well supported (Figs. 3.5 - 3.10 & 3.12 – 3.16, Chapter 3). All members of Series Austro-orientalis except *L. linearis* are strongly supported as a clade. The placement of *L. linearis* differs between the different gene trees and in the 3-locus gene tree it is unresolved relative to Series Austro-orientalis and Series Austro-unifolius. However, morphological data a chromosome number of 36, a simple foveolate seed coat surface pattern and digitate eophylls clearly support inclusion of *L. linearis* in Series Austro-orientalis (Planchuelo, 1996b).

Five Old World groups are recognised as individual monophyletic sections. Gene trees constructed from LEGCYC1A and LEGCYC1A & ITS in combination both support the monophyly of a group comprising the five Old World Sections and Section Boreo-unifolius

(the Florida unifoliolate species). However, these clades form a paraphyletic grade in the ITS, GPAT1 and chloroplast DNA (combined trnS-G, trnT-trnL and trnL intron) gene trees (Figs. 3.7 & 3.7 Chapter 3 and Chris Drummond pers. comm.). Sections Angustifolii and Lutei are strongly supported as sister taxa in the ITS, GPAT1 and LEGCYC1A single and combined gene trees (Figs. 3.5 - 3.10 & 3.12 – 3.16, Chapter 3) but these are retained as separate sections to maintain stability (see below).

7.6.3 Stability of the infrageneric classification

Within *Lupinus* there are several species of considerable economic importance in agriculture and horticulture importance (Gladstones, 1998). Most of the economically important lupins are from the Old World where the existing classification is congruent with the molecular phylogeny and was deemed robust and has been maintained. The only possible alternative arrangement that was considered was the merging of Sections Angustifolii and Lutei since they are consistently resolved together in a well-supported clade. However, this clade is not readily diagnosable and its recognition would only confuse users familiar with the current well-established classification of Old World species.

In order to avoid future disruption, only well-supported clades or those which can be reliably diagnosed with current data are formally named. It is possible that denser sampling of the North American, Mexican and Central American species in Section Occidentalis will highlight additional monophyletic clades which could merit recognition as additional series. It is anticipated that this can be encompassed within the proposed framework without causing significant alteration of names or ranks.

7.6.4 Rank

Two ranks, Sections and Series are used. The rank of section is used as the primary division within *Lupinus* due to the large number and nature of the groups recognised. Use of subgenus for lineages containing one or two species is inappropriate. Furthermore, the

botanical code advocates the use of section for the first division of the genus (Greuter, 2000). Treatment of the small Old World Sections at the same rank as the two large New World lineages is justified as the characters used to diagnose them are similar and the Old World sections are separated from each other by branches similar in length to those separating Old and New World clades and are thus comparably genetically different. Series are used to recognise additional strongly supported monophyletic, informative and diagnosable groups nested within the two large New World Sections. The three Series within Section Orientalis are not phylogenetically equivalent. This approach has been taken to avoid unnecessary confusion which would result from the recognition of a further level of hierarchy to account for the clade containing all the South American 2n=36 species. I would advocate the use of informal names to further subdivide Series Occidentalis e.g. the 'Andean' lupins to represent the lupins resulting from the large Andean radiation nested in Section Occidentalis Series Occidentalis.

7.6.5 Diagnostic characters utilised in the infrageneric classification of *Lupinus*

The characters which provide diagnosability of the classification are chromosome number, micromorphological seed coat surface pattern, cotyledon and leaf morphology, although geography is also useful in some cases. I agree with Watson (1873) that characters such as pubescence and stipule length which are constant in some species are not constant in others. Although chromosome number is not a morphological character it is largely congruent with the phylogeny and most notably so for the two large New World lineages. In contrast some of the Old World species show significant within-species variability in chromosome number (Chapters 5 and 6).

Macromorphology of seed coat surfaces although generally a good indicator for micromorphology in the Old World species, is not applicable over the genus as a whole. Some *L. hispanicus* seeds feel rough to the touch yet this is not a rough-seeded species as defined here, or by others (Ainouche & Bayer, 1996; Ainouche *et al.*, 2004; Heyn &

Herrnstadt, 1977). The surface sculpture of the smooth-seeded species is much finer in comparison with true rough-seeded species (Fig. 6.11, Chapter 6). Two examples show that seed coat surface macromorphology is not a reliable indicator for micromorphology. *Lupinus anatolicus* was originally described as a smooth-seeded species. However, it is firmly placed in the rough-seeded clade in the ITS gene tree (Ainouche *et al.*, 2004). The members of Section Platycarpos also have macromorphologically rough seeds but these have a different micromorphology from Section Scabrispermae (Fig. 6.12c & f, Chapter 6).

Of the characters providing diagnosability connate cotyledons, rough-seed coats and mature leaf characters were discovered early in the taxonomic history of the genus. However, the homoplasious nature of the unifoliolate leaf character has only been uncovered in the light of new estimate of the phylogeny (Chapters 3 and 6). The other useful characters - chromosome number, micromorphological seed coat surface pattern and eophyll type and number - have been more recently observed.

7.6.6 Discussion of new system relative to previous systems

The classification presented here (Appendix 18) adopts a conservative approach in line with some previous authors e.g. Taubert (1894) and Gross (1986). This system recognises relatively few groups diagnosed by unique combinations of characters and is based on a well-resolved and robustly-supported phylogeny derived from multiple genes. The resultant classification comprises groups of very variable sizes with some groups, particularly Series Occidentalis with ca. 238 species, that are very large compared to some previous classifications. Some may feel that a group of this size is not informative enough. However, this system clearly reflects the evolutionary history of the genus and it is recognised that variation in group size with many small groups and few large ones, as illustrated by the hollow curve, is an intrinsic feature of hierarchical systems of classification (Scotland & Sanderson, 2004; Willis & Yule, 1922).

The Old World Sections Scabrispermae, Angustifolii, Lutei, Micranthi and Albi have traditionally been circumscribed as groups (Ainouche *et al.*, 2004; Gladstones, 1984; Plitmann & Heyn, 1984). These five sections are retained in this classification because multiple gene trees from separate loci recognise these as monophyletic lineages, which are diagnosable.

In all previous classifications the unifoliolate species from both North and South America have been classified as a single group. Based on close morphological similarities, Dunn (1971) and Monteiro & Gibbs (1986a) assumed that the unifoliolate species were closely related and explained the notable geographic disjunction between South America and SE USA by long distance dispersal. However, this hypothesis has not been explicitly tested until now. Morphologically these groups are difficult to separate. They both have unifoliolate leaves and eophylls and simple-foveolate seed coats. In fact, such are the similarities that Monteiro & Gibbs (1986a) suggested that *L. villosus* from Florida might even be conspecific with *L. sellowianus* from southern Brasil. However, the Florida species can be readily distinguished from the South American unifoliolate species by chromosome number and beaked/non-beaked pods. The monophyly of the unifoliolate species is not supported by any of the molecular datasets analysed in this study, suggesting that the unifoliolate condition represents a convergence. Thus, Section Boreo-unifolius has been erected to account for the Florida unifoliolate species i.e. the independent reduction of digitate to unifoliolate leaves in parallel in North and South America. It was not deemed practical to recognise the South American unifoliolate species as a group despite their monophyly because *L. paraguariensis* has both unifoliolate and digitate adult leaves (Fig. 5.7c, Chapter 6). A more pragmatic approach, which also reflects evolutionary history, was to recognise Series Austro-unifolius, a well-supported clade diagnosable by the character four or more unifoliolate eophylls.

The New World species are placed in the three Sections Boreo-unifolius, Occidentalis and Orientalis. The species circumscribed in Section Occidentalis were recognised as a group, *Digitatae neogeae,* by Taubert (1894) whereas the other two sections are new (Fig. 7.1). In

the past the 2n=36 digitate-leaved species have been not been recognised as a group separate from the 2n=48 digitate-leaved species (Gross, 1986). Section Occidentalis is divided into two Series Platycarpos and Occidentalis. The Platycarpos group has long been recognised in many classifications. Some authors recognised a group which corresponds to Series Occidentalis i.e. Section *Lupinus* (Watson, 1873) and Subgenus *Lupinus* (Smith, 1938-1953). This arrangement differs from the more recent classifications which divide the New World, particularly the North American species, into many smaller sections (Sholars, 1996; Smith, 1951). All the Series within Section Orientalis are newly recognised.

The phylogeny of *Lupinus* shows a high degree of geographic structure with most of the sections/series occupying largely allopatric distributions (Fig. 7.3). This geographic structure is also specifically reflected in the names of the sections/series and in that sense the classification superficially resembles a classification based on geography although it is not. However, because the geographic structure is so predictable it allows geography to be used, with caution, as an additional feature facilitating diagnosis.

7.6.7 Limitations of the new system

A drawback to this new classification is the nature of some of the character states used to diagnose groups. Most are not characters that can be readily used for immediate placement of a species in a group, but rather require technical investigation or availability of seedlings. However, there are additional characters which, although not completely reliable because they are not fixed, can provide more accessible clues to identification. These include pod apex morphology, banner shape, calyx morphology, branching morphology, seed shape and presence of a seed aril rim as highlighted in Table 7.1.

A more serious problem of the proposed classification is that the species in all Sections and Series are not listed in full. Until comprehensive taxonomic accounts are available for all taxa it is impossible to list all the species especially in Section Occidentalis Series

Occidentalis, where species delimitation problems and extensive synonmy remain to be fully resolved.

The classification presented here is based on a new combined 3-locus phylogeny (Chapter 3). The sampling of the phylogeny although representing a far wider geographic and taxon sampling than all previous studies, is still sparsely sampled for North American species. Extrapolation from the phylogeny regarding the position of North American lupins more widely, in Section Occidentalis Series Occidentalis, should be taken with some caution.

7.6.8 Description and discussion of Sections and Series

The proposed classification is presented in Table 7.1. Latin diagnoses of new sections/series will be presented in due course at time of publication of the new classification.

Table 7.1 Conspectus of Sections and Series	
Section Albi Agardh	1 species
Type:	*L. albus* L.
Diagnosable characters:	Macromorphologically smooth seeds, granular micromorphological seed coat surface pattern, chromosome number of 2n=50 (most common), entire upper calyx lobe.
Distribution:	Mediterranean.
Phylogenetic affinities:	Closely related to other Old World sections but the precise relationships remain unclear.
Section Angustifolii Agardh	1 species
Type:	*L. angustifolius* L.
Diagnosable characters:	Macromorphologically smooth seeds, papillose micromorphological seed coat surface pattern, chromosome number of 2n=40 (most common), entire or almost entire lower calyx lobe, linear leaflets.
Distribution:	Mediterranean.
Phylogenetic affinities:	Strongly supported (bootstrap values of 73% in 3-locus analysis, 75% ITS, 33% LEGCYC1A, 75% GPAT1) as sister group to Section Lutei.
Section Lutei Agardh	2 species
Type:	*L. luteus* L.
Diagnosable characters:	Macromorphologically smooth seeds, simple-foveolate micromorphological seed coat pattern, chromosome number of 2n=52 (most common).
Distribution:	Mediterranean.
Support:	Bootstrap values of 99% (3-locus analysis), 87% (ITS), 97% (LEGCYC1A) and 92% GPAT1.
Phylogenetic affinities:	Sister group to Section Angustifolii.
Section Micranthi Agardh	1 species
Type:	*L. micranthus* Guss.

Diagnosable characters:	Macromorphologically smooth seeds, papillose micromorphological seed coat surface pattern, chromosome number of 2n=52 (most common).
Distribution:	Mediterranean
Phylogenetic affinities:	Closely related to other Old World sections but the precise relationships remain unclear.
Section Scabrispermae Plitm. & Heyn	8 species
Type:	*L. pilosus* Murr.
Diagnosable characters:	Macromorphologically rough seeds, fascicular micromorphological seed coat surface pattern.
Distribution:	Circum-Mediterranean with endemics in the Middle East, the Atlas Mountains and tropical mountains of eastern Africa.
Support:	Bootstrap values of 100% (3-locus analysis), 99% (ITS) and 99% (LEGCYC1A). Phylogenetic affinities: Closely related to other Old World sections but the precise relationships remain unclear.
Section Boreo-unifolius Eastwood & Hughes sect. nov.	ca. 4 species
Synonymy:	Section *Lupinus* Watson pro parte, (Watson, 1873)
Type:	*L. villosus* Willd.
Diagnosable characters:	Chromosome number 2n=52, non-beaked legume, unifoliolate leaves.
Distribution:	Southeastern USA.
Support:	Bootstrap values of 100% (3-locus analysis), 99% (ITS, LEGCYC1A, GPAT1).
Phylogenetic affinities:	Closely related to Old World sections but the precise relationships remain unclear.
Section Occidentalis Eastwood & Hughes sect. nov.	ca. 250 species
Synonymy:	Section *Lupinus* Watson pro parte, (Watson, 1873)
Type:	*L. formosus* Greene
Diagnosable characters:	Chromosome number 2n=48.
Distribution:	North, Central and South America - predominately west coast from Alaska to Argentina but some North American species distributed across to the east coast. Centres of diversity in the Rocky Mountains and the Andes.
Support:	Bootstrap values of 97% (3-locus analysis), 51% (ITS) and 67% (LEGCYC1A). Phylogenetic affinities: Sister to Section Orientalis.
Series Platycarpos (Watson) Eastwood & Hughes stat. nov .	ca. 12 species
Type:	*L. microcarpus* Sims
Diagnosable characters:	Connate cotyledons, regulate and protuberant seed coat surface pattern, annual life history, two ovules per ovary.
Distribution:	western USA, southeastern Canada and coastal Chile
Support:	Bootstrap values of 100% (3-locus analysis), 62% (ITS) and 99% (LEGCYC1A, GPAT1).
Phylogenetic affinities:	Sister to Series Occidentalis.
Series Occidentalis Eastwood & Hughes ser. nov.	ca. 238 species
Type:	*L. formosus* Greene
Diagnosable characters:	Petiolate cotyledons.
Distribution:	North, Central and South America - predominately west coast from Alaska to Argentina but some North American species distributed across to the east coast. Centres of diversity in the Rocky Mountains and the Andes.
Support:	Bootstrap value of 70% (3-locus analysis). Unresolved relative to Series Platycarpos in other analyses.
Phylogenetic affinities:	Sister to Series Platycarpos.

Section Orientalis Eastwood & Hughes sect. nov.	ca. 33 species
Type:	*L. havardii* Watson
Diagnosable characters:	Chromsome number 2n=36, simple-foveolate seed coat surface pattern.
Distribution:	lowland South America and reaching into the Bolivian Andes. Four representatives in south-central USA and NE Mexico.
Support:	Bootstrap values of 100% (3-locus analysis), 84% (ITS), 94% (LEGCYC1A) and 87% GPAT1.
Phylogenetic affinities:	Sister to Section Occidentalis.
Series Boreo-orientalis Eastwood & Hughes ser. nov.	ca. 4 species
Type:	*L. havardii* Watson
Diagnosable characters:	Seeds square in transverse section.
Distribution:	south-central USA and possibly into northern Mexico.
Support:	Bootstrap values of 100% (3-locus analysis and ITS), 87% (LEGCYC1A) and 98% GPAT1.
Phylogenetic affinities:	Sister to a clade containing Series Austro-orientalis and Austro-unifolius.
Series Austro-orientalis Eastwood & Hughes sect. nov.	ca. 14 species
Type:	*L. gibertianus* C. P. Smith
Diagnosable characters:	Digitate adult leaves, digitate or two unifoliolate eophylls.
Support:	Bootstrap values of 91% (3-locus analysis), 96% (ITS), and 85% GPAT1 excluding *L. linearis*. Bootstrap value of 59% (LEGCYC1A), including *L. linearis*.
Phylogenetic affinities:	Sister to Series Austro-unifolius.
Series Austro-unifolius Eastwood & Hughes sect. nov.	ca. 14 species
Type:	*L. arenarius* Gardn. in Hooker
Diagnosable characters:	More than two unifoliolate eophylls.
Distribution:	Southern central Brazil, Uruguay, Paraguay and Argentina.
Support:	Bootstrap values of 99% (3-locus analysis), 78% (ITS), 60% (LEGCYC1A) and 18% GPAT1.
Phylogenetic affinities:	Sister to Series Austro-orientalis.

7.7 Conclusion

The combination of strongly-supported monophyly in a molecular phylogeny and non-molecular diagnosability used here provides a more compelling basis for classification than either alone. Translating the *Lupinus* phylogeny into a Linnaean classification has been straightforward compared to many other groups. The resulting classification resembles one based on geography due to the high levels of geographic structure exhibited by the phylogeny. The size of the sections recognised is very uneven representing the evolutionary history of *Lupinus* as currently understood. However, because the classification can be used in conjunction with the phylogeny more detailed knowledge of the relationships between the species in the larger sections is available.

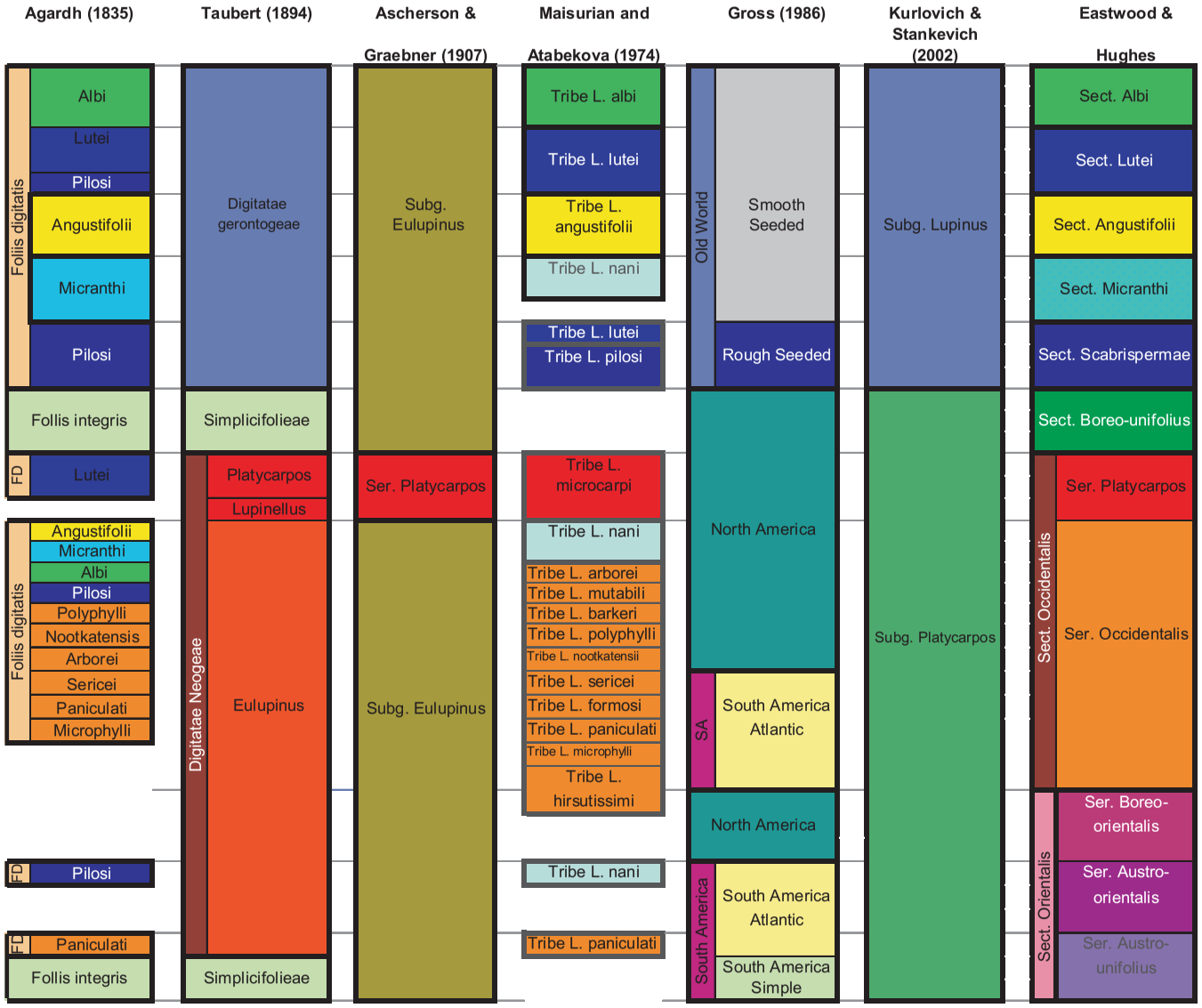
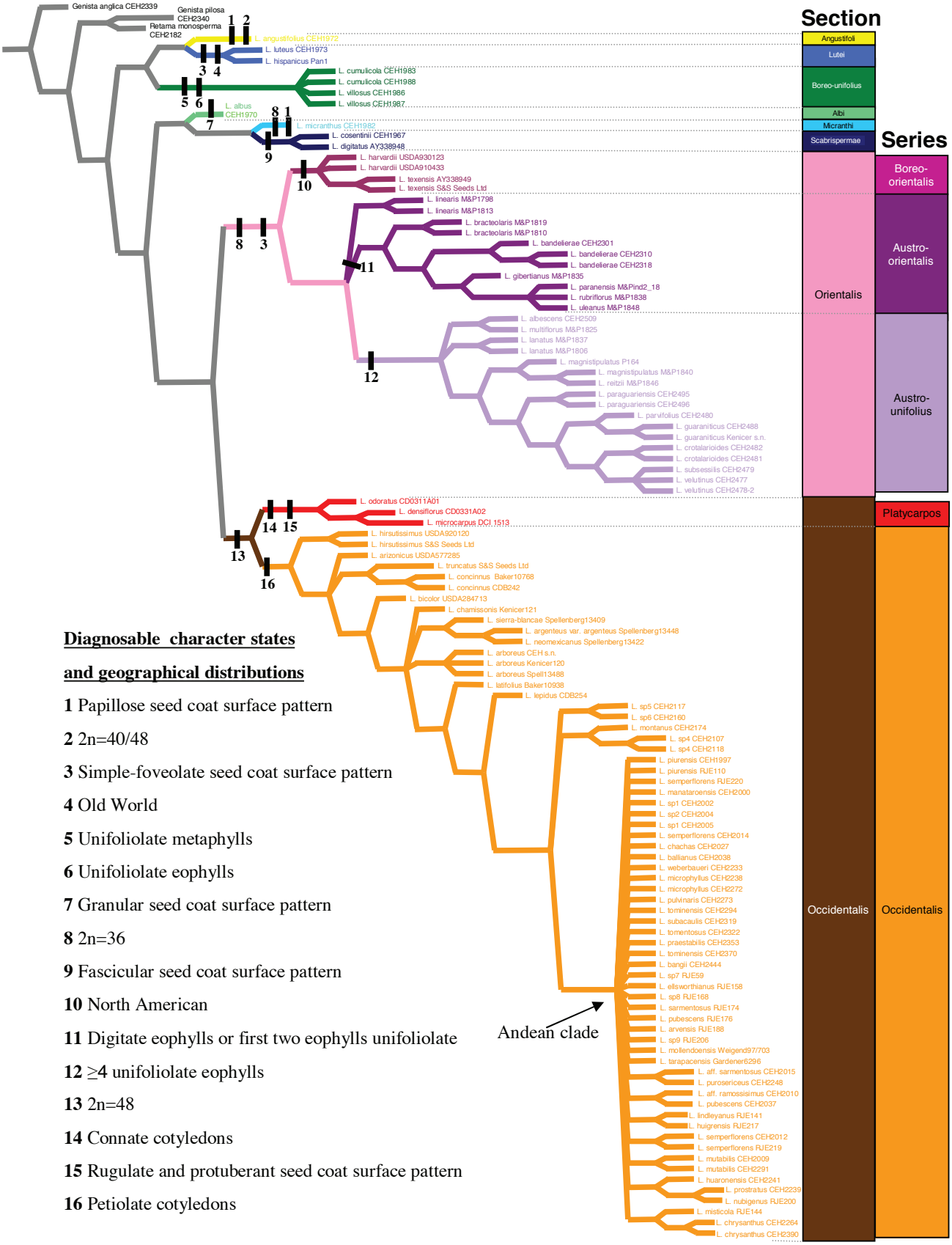


Figure 7.1 Comparision of genus-wide classifications of *Lupinus*. Identical groups are shown in the same colour.



Diagnosable character states and geographical distributions

1 Papillose seed coat surface pattern

2 2n=40/48

3 Simple-foveolate seed coat surface pattern

4 Old World

5 Unifoliolate metaphylls

6 Unifoliolate eophylls

7 Granular seed coat surface pattern

8 2n=36

9 Fascicular seed coat surface pattern

10 North American

11 Digitate eophylls or first two eophylls unifoliolate

12 ≥4 unifoliolate eophylls

13 2n=48

14 Connate cotyledons

15 Rugulate and protuberant seed coat surface pattern

16 Petiolate cotyledons

Fig. 7.2 Infrageneric classification and the *Lupinus* phylogeny. The strict consensus tree from the parsimony analysis of the 3-locus data set is shown. Section/series are colour coded. The character states and geographical distributions that provide diagnosibility of the sections/series are numbered. The undiagnosable but strongly supported Andean node is indicated.

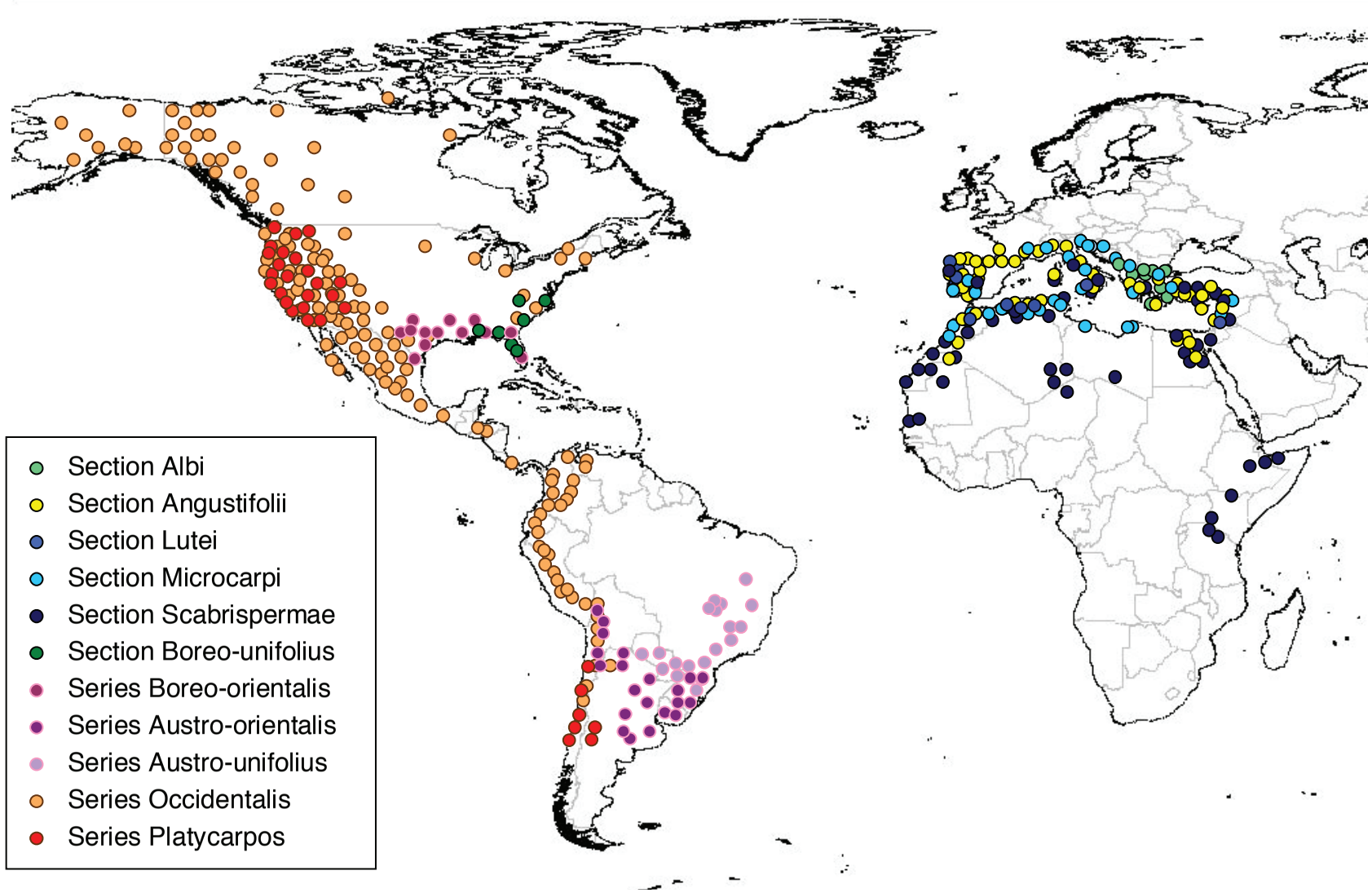


Fig. 7.3 Distribution Map of *Lupinus* Sections/Series. Data from Gladstones (1984), Monteiro & Gibbs (1986), Isley (1998), USDA http://plants.usda.gov/, and Hughes & Eastwood (unpubl. data).

8 Conclusions

8.1 Achievements and broader implications

Fully-resolved, well-sampled and robustly-supported species-level phylogenies for plant groups remain few and far between (Appendix 3, Hughes *et al.*, 2006). While analyses of a set of exemplar taxa that resolve a subset of well-supported nodes will often be sufficient for the purpose of providing a robust phylogenetic classification, poorly sampled, weakly supported and partially resolved gene and species trees are of limited use in studies of hybridisation, polyploidy, character evolution, species diversification, speciation and domestication. Species-level phylogenies for detailed evolutionary studies often demand not just complete or near-complete taxon sampling, but ideally multiple accessions within species, as well as complete or near-complete resolution with high statistical support (Hawkins, 2006; Appendix 2, Hughes *et al.*, 2006). In this thesis I discuss how to build such trees and then use appropriate methods to reconstruct a well-sampled, well-resolved and well-supported phylogeny for my study group *Lupinus.*

The 3-locus phylogeny presented here represents the **best estimate of *Lupinus* relationships** to date. With the exception of the Andean clade the phylogeny is highly resolved and well-supported. Taxon sampling is representative of the morphological diversity and geographic distribution of *Lupinus* species. Sequence data from three independent loci have been combined. The phylogeny identified two previously unrecognised and well-supported New World clades with distinct geographic signatures. One is predominately western New World whilst the other is primarily distributed in the eastern New World. Chromosome numbers provide synapomorphies for these clades. Although Andean lupins are not monophyletic the majority form a large well-supported polytomy nested within North American and Mexican species. The few Andean species which fall outside this clade are members of the eastern New World clade. Uncertainty remains surrounding relationships at the base of the phylogeny, particularly between Old World species.

This study adds to a growing body of evidence and datasets which suggest that reconstructing phylogenies for very rapidly evolving groups can be extremely challenging (Kocher, 2003; Richardson *et al.*, 2001a; 2001b; Rokas *et al.*, 2005; Scherson *et al.*, 2005; Whittall *et al.*, 2006). Commonly utilised loci e.g. ITS and cpDNA will continue to be useful for building framework phylogenies but are insufficiently variable for many species-level investigations (Bayer, 1996; Emshwiller & Doyle, 1998; Hodges & Arnold, 1994; Miller *et al.*, 1999; Samuel *et al.*, 1998; Schilling *et al.*, 1998; Sonnante *et al.*, 2003; Vargas *et al.*, 1998; Wink *et al.*, 1999). The difficulties associated with locating and developing "hypervariable" loci suggest that universal primers will not be found and that lineage-specific solutions will have to be developed (Sang, 2002). The low levels of sequence variation found between closely related species also suggests that DNA barcoding for such groups may be much more challenging than so far acknowledged (Janzen, 2004; Tautz *et al.*, 2002; 2003).

The new phylogeny for *Lupinus* has provided a solid basis for a set of studies on species diversification, domestication, character evolution and infrageneric classification. These have transformed our knowledge of the evolutionary history and systematics of this well-known and economically important genus. Using the new phylogeny and evolutionary rates analysis I have been able to present a detailed time-calibrated hypothesis of diversification for *Lupinus*. The most noteable finding from this study is the discovery of the **Andean radiation** of lupins. The Andean crown clade containing about 81 species, as estimated through re-assessment of species delimitation in the Andes, is dated at 1.47 ± 0.29 Myr implying a net species diversification rate of 2.49–3.72 species per Myr. This rate is **faster than any other plant diversification** documented to date. Furthermore, it suggests that the high cichlid diversification rates, which have long been an order faster than any other group, are not unique. These analyses have contributed to a major paper on Andean plant diversification published recently in PNAS (Appendix 3, Hughes *et al.*, 2006).

The discovery of the fastest plant diversification documented in the Andean lupin radiation has highlighted the Andes as an ideal location to study rapid species diversification. The

recency of orogeny and appearance of high upland environments are such that many other groups are likely to show similar patterns and may indeed have faster diversification rates.

Comparative morphological studies of wild Andean species and the domesticate *L. mutabilis* have **identified a putative progenitor** and a set of potentially closely related species. These suggest that **seed size has increased at least six-fold** through domestication. The distribution of the putative progenitor implies an **origin for *L. mutabilis* in northern Peru**. Archaeological data suggest domestication occurred **between 5500 – 2500 BP**.

The benefits of integrating data from disparate fields e.g. comparative morphology, traditional taxonomy, molecular systematics and archaeology in studies of domestication as advocated by Smith (2001) and Zeder *et al*. (2006) are demonstrated within this project. Molecular data uncovered the explosive Andean radiation explaining why molecular data are uninformative about the sister group relationships of the Andean domesticate *L. mutabilis.* Comparative morphology was able to pinpoint putative close relatives to *L. mutabilis.* Archaeological data provided a time frame for domestication and a hypothesis for the geographic origin of *L. mutabilis* was suggested on the basis of fieldwork and specimen-based mapping of species distributions. This suggests that it is important that traditional areas of research are not swamped by more fashionable modern molecular techniques.

These results provide a new and explicitly based hypothesis for the origins of this important Andean crop. However, the extreme lack of sequence divergence, attributable to the exceptionally rapid diversification and recency of the Andean lupin radiation, means that finding independent support for this hypothesis from molecular data remains extremely challenging. This is in line with findings for other Andean domesticates such as oca and potato (Emshwiller & Doyle, 1998, 2002; Van den Berg *et al.*, 1998), where similar challenges posed by domestications embedded in similar large and recent radiations have been encountered. Evidence to confirm the Andes as a hotspot for crop domestication is growing and concurrently indicating that crop origins in this region may be particularly

difficult to reconstruct due to the recent evolution of the Andean flora as a whole.

Leaf evolution was investigated via character optimisation onto the 3-locus phylogeny. Results contradict a previous hypothesis suggesting that unifoliolate species formed a natural group and that this condition was plesiomorphic within *Lupinus*. These new sets of data imply **two independent origins of unifoliolate leaves in North and South America**. A hypothesis proposing a change in developmental timing as a response to selection for rapid growth is presented to explain this convergent evolution of unifoliolate leaves.

Currently there is much interest in the development and evolution of leaves (Bharathan & Sinha, 2001; Champagne & Sinha, 2004; Gleissberg & Kadereit, 1999; Harrison *et al.*, 2005; Piazza *et al.*, 2005). Unifoliolate leaves are widespread across the legumes. This is in line with the discovery of convergent evolution within *Lupinus* presented here in suggesting that leaf morphology is highly labile. Questions about how to define leaf types are raised by the evidence presented here. Detailed comparative re-assessment of other genera is needed to test whether these "fuzzy" boundaries between leaf types are more widespread than acknowledged in the literature. The labile nature of leaf morphology provides multiple independent examples of genera with both uni- and multifoliolate leaves which make powerful models for testing the genetic basis of leaf diversity.

There has been considerable recent debate surrounding whether and how best to integrate molecular phylogenetic information into classification (Brummitt, 1997, 2002, 2003; Brummitt & Sosef, 1998; Carine & Scotland, 2002; de Queiroz & Gauthier, 1990, 1992, 1994; Grant, 2003; Harlin, 2003; Jardine, 1969a; Moore, 1998; Nelson *et al.*, 2003; Pfeil & Crisp, 2005; Potter & Freudenstein, 2005; van Welzen, 1998; Wheeler, in press; Williams *et al.*, 2005). Molecular phylogenetic data were used as the basis for the new **infrageneric classification** of *Lupinus* proposed here. The genus is divided into eight sections and five series. Three sections and four series are new.

Lupinus provides an example of a phylogeny that can be easily and directly translated into an infrageneric classification based on monophyly and diagnosability suggesting that more densely sampled phylogenies and detailed comparative morphological data can provide the basis for integration of phylogeny data into Linnaean classification in many other groups. This result supports the views of Nixon *et al.* (2003) that small improvements, e.g. naming of monophyletic groups only, to the existing Linnaean system can solve the difficulties associated with constructing phylogenetic classifications without resorting to new classification systems as suggested by proponents of the phylocode (de Queiroz & Gauthier, 1992, 1994). However, the application of phylogenetic data to classification is reliant on well-sampled, well-resolved and well-supported phylogeny estimates.

8.2 Challenges

Perhaps the greatest challenge encountered during this project was lack of sequence variation between Andean taxa. This hampered reconstruction of relationships within this group. However, discovery of the exceptionally rapid rate of diversification within the Andean clade explains the reason for this problem. Phylogenetic studies of other groups which have high net rates of diversification e.g. cichlid fish (Albertson *et al.*, 1999; Kocher, 2003, 2004), *Astragalus* (Sanderson & Wojciechowski, 1996; Scherson *et al.*, 2005; Wojciechowski, 2005; Wojciechowski *et al.*, 1999) and *Aquilegia* (Hodges, 1997; Hodges & Arnold, 1994; Whittall *et al.*, 2006) have also encountered similar problems. A range of methods (SCARs (Bailey *et al.*, 2004; Lockhart *et al.*, 2001; McLenachan *et al.*, 2000; Rameau *et al.*, 1998; Shaw *et al.*, 2003; Yang *et al.*, 2001), CATs (Chandappa *et al.*, 2005; Choi *et al.*, 2006; Fulton *et al.*, 2002; Lyons *et al.*, 1997; Scherson *et al.*, 2005) and 3'-UTR anchored primers (Whittall *et al.*, 2006) to locate hyper-variable loci for species-level phylogeny reconstruction have had mixed and so far generally very limited success implying that perhaps the limits of sequence data utility may have been reached for these portions of the Tree of Life that show exceptionally high rates of species diversification. These studies all show that highly variable DNA sequence loci are rare. My preliminary SCAR-based

study sampled only nine loci, thus in this context it is not surprising that none of the loci were particularly promising.

8.3 Further work

Ideas for further work fall into three main categories: taxon sampling, additional data sources and new analysis methods.

Denser taxon sampling is a common feature of almost all successive phylogenetic studies of particular groups. This is the case for *Lupinus* with earlier studies sampling just 31 and 41 species (Ainouche & Bayer, 1999; Kass & Wink, 1997a respectively) and this study sampling 134 (Chapter 3). However, current sampling of *Lupinus* species, still only accounts for 46% of the extant species. Complete or near-complete taxon sampling would open the way for application of more sophisticated techniques for analysing species diversity (e.g. lineage-through-time plots) and more detailed biogeographical analyses. Denser sampled trees would allow more rigorous analyses of shifts in diversification rate and the possible cause of such shifts e.g. via SymmeTREE (Chan & Moore). Another inevitable feature of multiple phylogenetic analyses is the mismatch of loci and accessions, posing challenges for straightforward total evidence approaches that combine all available data. These factors are very much to the fore in studies in *Lupinus.* One approach to overcome sampling mismatches is the production of supertrees. A supertree approach could be utilised to incorporate the partially overlapping cpDNA sequence datasets with the nDNA data presented here in order to widen sampling of North American species. However, a major obstacle to this remains complete and consistent delimitation of species across the genus as a whole, and this is perhaps the most urgent priority for taxonomic and evolutionary studies in *Lupinus.*

Discovery of the exceptionally rapid diversification of Andean lupins largely explains the low levels of sequence variation between taxa. The recency of their origin suggests that a

more population-based approach using PCR-based markers may be a more successful way to resolve relationships between species than via searching for additional DNA sequence loci. This would need to include multiple population and multiple accessions per population per species and utilise molecular markers e.g. AFLPs in the expectation of locating more differences between species by sampling from across the whole genome. This approach has been highly successful in *Senecio* L. (Pelser *et al*, 2003) and *Lactuca* L. (Koopman, 2005). More exhaustive Bayesian analyses should be carried out to test whether the additional resolution, recovered utilising this method, is real and supported.

Finally, there is good scope to investigate the genetic basis of the contrasting leaf types in *Lupinus*. Hybridisation studies between unifoliolate and digitate Brazilian species would provide material for observations of the eophyll to metaphyll transition and QTL analysis to investigate the genetic basis of this character change.

References

Acuña, N. & Planchuelo, A.M. (1996) Comparative morphology of Lupin seedlings. In *Towards the 21st Century. Proceedings of the 8th International Lupin Conference* (ed. G. Hill, D.), pp. 386-393, Asilomar, California.

Adams, R.J. & Smith, M.V. (1977) Scanning electron and light microscope studies of pollens of some Leguminosae. *Journal of Apicultural Research*, **16**, 99-106.

Agardh, J.G. (1835) *Synopsis Generis Lupini*, Berlin.

Ainouche, A. & Bayer, R. (1996) Phylogenetic relationships among and within the old and new world *Lupinus* species (Fabaceae) based on internal transcribed spacer sequences of nuclear ribosomal DNA. In *Proceedings of the 8th International Lupin Conference,* California (ed. G. Hill, D.), pp. 236-244, California.

Ainouche, A. & Bayer, R. (1999) Phylogenetic relationships in *Lupinus* (Fabaceae: Papilionoideae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. *American Journal of Botany*, **86**, 590-607.

Ainouche, A. & Bayer, R. (2000) Genetic evidence supports the new anatolian lupine accession, *Lupinus anatolicus*, as an Old World "rough-seeded" lupine (Section Scabrispermae) related to *L. pilosus*. *Folia Geobotanica*, **35**, 83-95.

Ainouche, A., Bayer, R., Cubas, P. & Misset, M.-T. (2003) Phylogenetic relationships within tribe Genisteae (Papilionoideae) with special reference to genus *Ulex*. In *Advances in Legume Systematics part 10, Higher Level Systematics.* (eds. B. Klitgaard & A. Bruneau), pp. 239-252. Royal Botanic Gardens, Kew.

Ainouche, A., Bayer, R. & Misset, M.-T. (2004) Molecular phylogeny, diversification and character evolution in *Lupinus* (Fabaceae) with special attention to Mediterranean and African lupines. *Plant Systematics and Evolution*, **246**, 211-222.

Ajibade, S.R., Weeden, N.F. & Chite, S.M. (2000) Inter simple sequence repeat analysis of genetic relationships in the genus *Vigna*. *Euphytica*, **111**, 47-55.

Akiyama, S. (1988) A revision of the genus *Lespedeza* Section *Macrolespedeza* (Leguminosae) *Bulletin of the University Museum and the University of Tokyo*, **33**.

Albert, V.A., Gustafsson, M.H.G. & DiLaurenzio, L. (1998) Ontogenetic systematics, molecular developmental genetics and the angiosperm petal. In *Molecular Systematics of Plants II: DNA Sequencing.* (eds. D.E. Soltis, P.S. Soltis & J.J. Doyle), pp. 349-374. Chapman and Hall, London.

Albertson, R.C., Markert, J.A., Danley, P.D. & Kocher, T.D. (1999) Phylogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences, USA*, **96**, 5107-5110.

Alfaro, M.E. & Zoller, S.L. (2003) Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution*, **20**, 255-266.

Alice, L.A. & Campbell, C.S. (1999) Phylogeny of *Rubus* (Rosaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *American Journal of Botany*, **86**, 81-87.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389-3402.

Alvarez, I. & Wendel, J.F. (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**, 417-434.

Andreasen, K. & Baldwin, B.G. (2003) Nuclear ribosomal DNA sequence polymorphism and hybridization in checker mallows (*Sidalcea*, Malvaceae) *Molecular Phylogenetics and Evolution*, **29**, 563-581.

Angiosperm Phylogeny Group (2003) An update for the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society*, **141**, 399-436.

Aniszewski, T., Kupari, M.H. & Leinonen, A.J. (2001) Seed number, seed size and seed diversity in Washington Lupin (*Lupinus polyphyllus* Lindl.) *Annals of Botany*, **87**, 77-82.

Antúnez de Mayolo, S. (1982) Tarwi in Ancient Peru. In *Proceedings of the1st International Lupin Conference*, pp. 1-12, Lima/Cuzco, Peru.

Archibald, J.K., Mort, M.E. & Crawford, D.J. (2003) Bayesian inference of phylogeny: a non-technical primer. *Taxon*, **52**, 187-191.

Archibald, J.K., Mort, M.E. & Wolfe, A.D. (2005) Phylogenetic relationships within *Zaluzianskya* (Scrophulariaceae s.s., tribe *Manuleeae*): Classification based on DNA sequences from multiple genomes and implications for character evolution and biogeography. *Systematic Botany*, **30**, 196-215.

Ascherson, P. & Graebner, P. (1907) *Sinopsis der Mitteleuropaischen Flora*, Leibzig.

Ashby, E. (1948) Studies in the morphogenesis of leaves. I. An essay on leaf shape. *New Phytologist*, **47**, 153-176.

Atkins, C.A., Smith, P.M.C., Gupta, S., Jones, M.G.K. & Caligari, P.D.S. (1998) Genetics, cytology and biotechnology. In *Lupins as Crop Plants: Biological, Production and Utilization.* (eds. J.S. Gladstones, C.A. Atkins & J. Hamblin), pp. 67-92. CAB International, Wallingford.

Backlund, A. & Bremer, B. (1998) To be or not to be - principles of classification and monotypic plant families. *Taxon*, **47**, 391-400.

Badr, A. (1994) Chloroplast DNA restriction site polymorphism in Genisteae (Leguminosae) suggests a common origin for European and American lupines. *Plant Systematics and Evolution*, **193**, 95-106.

Bailey, C.D. & Doyle, J.J. (1999) Potential phylogenetic utility of the low-copy number nuclear gene *pistillata* in dicotyledonous plants: comparison to nrDNA ITS and trnL intron in *Sphaerocardamum* and other Brassicaceae. *Molecular Phylogenetics and Evolution*, **13**, 20-30.

Bailey, C.D., Carr, T.G., Harris, S.A. & Hughes, C.E. (2003) Characterisation of angiosperm nrDNA polymorphism, paralogy and pseudogenes. *Molecular Phylogenetics and Evolution*, **29**, 435-455.

Bailey, C.D., Hughes, C.E. & Harris, S.A. (2004) Using RAPDs to develop DNA sequence loci for species level phylogeny reconstruction: an example from *Leucaena* (Fabaceae) *Systematic Botany*, **29**, 4-12.

Baldwin, B. (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution*, **1**, 3-16.

Baldwin, B. & Sanderson, M.J. (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae) *Proceedings of the National Academy of Sciences, USA*, **95**, 9402-9406.

Baldwin, B., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S. &

Donoghue, M.J. (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of Missouri Botanical Garden*, **82**, 247-277.

Barkman, T.J. & Simpson, B.B. (2002) Hybrid origin and parentage of *Dendrochilum acuiferum* (Orchidaceae) inferred in a phylogenetic context using nuclear and plastid DNA sequence data. *Systematic Botany*, **27**, 209-220.

Barneby, R. (1990) *Lupinus*. In *Intermountain Flora: Vascular Plants of the Intermountain West, USA.* (ed. A. Cronquist), Vol. 3B, pp. 237-267. New York Botanical Garden Press, New York.

Barthlott, W. (1984) Microstructural features of seed surfaces. In *Current Concepts in Taxonomy.* (eds. V.H. Heywood & D.M. Moore), Vol. 25, pp. 95-105. Academic Press.

Bateman, R.M. (1999) Integrating molecular and morphological evidence of evolutionary radiations. In *Molecular Systematics and Plant Evolution.* (eds. P.M. Hollingsworth, R.M. Bateman & R.J. Gornall), Vol. 1, pp. 432-471. Taylor and Francis, London.

Bateman, R.M., Pridgeon, A.M. & Chase, M.W. (1997) Phylogenetics of subtribe Orchidinae (Orchidoideae, Orchidaceae) based on nuclear ITS sequences. 2. Infrageneric relationships and reclassification to achieve monophyly of *Orchis sensu stricto*. *Lindleyana*, **12**, 113-141.

Baum, D.A., Alverson, W.S. & Nyffeler, R. (1998) A durian by any other name: taxonomy and nomenclature of the core Malvales. *Harvard Papers in Botany*, **3**, 315-330.

Bayer, R.J. (1996) Phylogenetic relationships in Sarraceniaceae based on rbcL and ITS sequences. *Systematic Botany*, **21**, 121-134.

Bell, C.D. & Donoghue, M.J. (2005) Phylogeny and biogeography of Valerianaceae (Dipsacales) with special reference to South American valerians. *Organisms, Diversity and Evolution*, **5**, 147-159.

Bellido, L.L. (1984) World report on lupin. In *Proceedings of the 3rd International Lupin Conference,* pp. 465-510, La Rochelle, France.

Bennett, T., Sieberer, T., Willett, B., Booker, J., Luschnig, C. & Leyser, O. (2006) The *Arabidopsis* MAX pathway controls shoot branching by regulating auxin transport. *Current Biology*, **16**, 553-563.

Bentham, G. (1859) Papilionaceae. In *Flora Brasiliensis.* (ed. C.F.P. Martius), Vol. 15(1), pp. 9-16, Munich.

Berry, P.E., Hahn, W.J., Sytsma, K.J., Hall, J.C. & Mast, A. (2004) Phylogenetic relationships and biogeography of *Fuchsia* (Onagraceae) based on noncoding nuclear and chloroplast DNA data. *American Journal of Botany*, **91**, 601-614.

Besnard, G., Khadari, B., Baradat, P. & Bervillé, A. (2002) *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. *Applied Genetics*, **104**, 1353-1361.

Bharathan, G. & Sinha, N.R. (2001) The regulation of compound leaf development. *Plant Physiology*, **127**, 1533-1538.

Bharathan, G., Goliber, T.E., Moore, C., Kessler, S., Pham, T. & Sinha, N. (2002) Homologies in leaf form inferred from KNOX1 gene expression during development. *Science*, **296**, 1858-1860.

Bhella, R.S. & Mackenzie, S.L. (1994) Nucleotide sequence of a cDNA from *Carthamus tinctorius* encoding a glycerol-3-phosphate acyltransferase. *Plant Physiology*, **106**, 1713.

Bieniek, M.E. & Millington, W.F. (1967) Differentiation of lateral shoots as thorns in *Ulex europaeus*. *American Journal of Botany*, **54**, 61-70.
Bisby, F.A. (1981) Tribe 32. Genisteae (Adans.) Benth. In *Advances in Legume Systematics.* (eds. R.M. Polhill & P.H. Raven), Vol. 1, pp. 409-425. Royal Botanic Gardens, Kew.
Blanco Galdes, O. (1982) Genetic variability of Tarwi (*Lupinus mutabilis* Sweet) In *Proceedings of the 1st International Lupine Conference*, pp. 33-50.
Blanco Galdes, O. (1984) *Lupinus mutabilis* germplasm resources in South America. In *Proceedings of the 3rd International Lupin Conference*, pp. 159-164.
Blanco Galdes, O. (1986) Wild stock of *Lupinus mutabilis* Sweet. In *Proceedings of the 4th International Lupin Conference*, Vol. 1, pp. 284, Geraldton, Western Australia.
Böhle, U.-R., Hilger, H.H. & Martin, W.F. (1996) Island colonization and evolution of the insular woody habit in *Echium* L. (Boraginaceae) *Proceedings of the National Academy of Sciences, USA*, **93**, 11740-11745.
Bolmgren, K. & Oxelman, B. (2004) Generic limits in *Rhamnus* L. s.l. (Rhamnaceae) inferred from nuclear and chloroplast DNA sequence phylogenies. *Taxon*, **53**, 383-390.
Borchiellini, C., Chombard, C., Manuel, M., Alivon, E., Vacelet, J. & Boury-Esnault, N. (2004) Molecular phylogeny of Demospongiae: implications for classification and scenarios of character evolution. *Molecular Phylogenetics and Evolution*, **32**, 823-837.
Bragg, L.H. (1983) Seed coats of some *Lupinus* species. *Scanning Electron Microscopy*, **IV**, 1739-1745.
Brako, L. & Zarucchi, J.L., eds.. (1993) *Catalogue of the Flowering Plants and Gymnosperms of Peru: Catálogo de las Angiospermas y Gimnospermas del Perú.* pp xl, 1286. Missouri Botanical Garden, St. Louis, Mo.
Braun, E.L. & Kimball, R.T. (2001) Polytomies, the power of phylogenetic inference, and the stochastic nature of molecular evolution: a comment on Walsh et al. (1999) *Evolution*, **55**, 1261-1263.
Briggs, D. & Walters, S.M. (1984) *Plant Variation and Evolution*, 2nd edn. Cambridge University Press, Cambridge.
Britten, R.J., Rowan, L., Williams, J. & Cameron, R.A. (2003) Majority of divergence between closely related DNA samples is due to indels. *Proceedings of the National Academy of Sciences, USA*, **100**, 4661-4665.
Brower, A. (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences, USA*, **91**, 6491-6495.
Brower, A.V.Z. & Schawaroch, V. (1996) Three steps of homology assessment. *Cladistics*, **12**, 265-272.
Brown, G.R., Kadd, E.E., Bassoni, D.L., Kiehne, K.L., Temesgen, B., van Buijtenen, J.P., Sewell, M.H., Marshall, K.A. & Neale, D.B. (2001) Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. *Genetics*, **159**, 799-809.
Brown, T.A. (1999) How ancient DNA may help in understanding the origin and spread of agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **354**, 89-98.
Brummitt, N. & Lughadham, E.N. (2003) Biodiversity: where's hot and where's not? *Conservation Biology*, **17**, 1442-1448.

Brummitt, R., K. (1997) Taxonomy versus cladonomy, a fundamental controversy in biological systematics. *Taxon*, **46**, 723-735.
Brummitt, R., K. & Sosef, M.S.M. (1998) Paraphyletic taxa are inherent in Linnaean classification - a reply to Freudenstein. *Taxon*, **47**, 411-412.
Brummitt, R., K. (2002) How to chop a tree. *Taxon*, **51**, 31-41.
Brummitt, R., K. (2003) Further dogged defence of paraphyletic taxa. *Taxon*, **52**, 803-804.
Brush, S.B. (1976) Man's use of an Andean ecosystem. *Human Ecology*, **4**, 147-166.
Buirchell, B. & Cowling, W.A. (1992) Domestication of rough-seeded lupins. *Journal of Agriculture, Western Australia*, **33**, 131-137.
Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L. & Waddell, P.J. (1993) Partitioning and combining data in phylogenetic analysis. *Systematic Biology*, **42**, 384–397.
Bullitta, A. & Hayward, M.D. (1996) Application of RAPD markers to a study of species relationships in the genus *Trifolium*. In *Legumes of Economic Importance*. (eds. B. Pickersgill & J.M. Lock), Vol. 8, pp. 127-136. Royal Botanic Gardens, Kew.
Burger, D. (1972) *Seedlings of Some Tropical Trees and Shrubs, Mainly of South East Asia.* Centre for Agricultural Publishing and Documentation, Wageningen.
Burger, R.L. (1989) An overview of Peruvian archaeology (1976-1986) *Annual Review of Anthropology*, **18**, 37-69.
Burnham, R.J. & Graham, A. (1999) The history of Neotropical vegetation: new developments and status. *Annals of the Missouri Botanical Garden*, **86**, 546-589.
Buss, C.C., Lammers, T.G. & Wise, R.R. (2001) Seed coat morphology and its systematic implications in *Cyanea* and other genera of Lobelioideae (Campanulaceae). *American Journal of Botany*, **88**, 1301-1308.
Bussell, J.D., Waycott, M. & Chappill, J. (2005) Arbitrarily amplified DNA markers as characters for phylogenetic inference. *Perspectives in Plant Ecology, Evolution and Systematics*, **7**, 3-26.
Butler, A.B. & Saidel, W.M. (2000) Defining sameness: historical, biological, and generative homology. *Bioessays*, **22**, 846-853.
Calvillo-Canadell, L. & Cevallos-Ferriz, S.R.S. (2005) Diverse assemblage of Eocene and Oligocene Leguminosae from Mexico. *International Journal of Plant Science*, **166**, 671-692.
Camacho, F.J., Gernandt, D.S., Liston, A., Stone, J.K. & Klein, A.G. (1997) Endophytic fungal DNA, the source of contamination in spruce needle DNA. *Molecular Biology and Evolution*, **6**, 983-987.
Cantino, P.D., Bryant, H.N., de Queiroz, K., Donoghue, M.J., Eriksson, T., Hillis, D.M. & Lee, M.S.Y. (1999) Species names in phylogenetic nomenclature. *Systematic Biology*, **48**, 790-807.
Cantino, P.D. & de Queiroz, K. (2006) Phylocode, pp. 1-83, http://www.ohiou.edu/phylocode/.
Carine, M.A. & Scotland, R.W. (2002) Classification of Strobilanthinae (Acanthaceae): trying to classify the unclassifiable? *Taxon*, **51**, 259-279.
Carlsward, B.S., Whitten, W.M., Williams, N.H. & Bytebier, B. (2006) Molecular phylogenetics of Vandeae (Orchidaceae) and the evolution of leaflessness. *American Journal of Botany*, **93**, 770-786.
Chacón, S.M.I., Pickersgill, B. & Debouck, D.G. (2005) Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and

Andean cultivated races. *Theoretical and Applied Genetics*, **110**, 432-444.
Chakrabarti, S., Pattanayak, D., Sarkar, D., Chimote, V. & Naik, P. (2006) Stability of RAPD fingerprints in potato: effect of source tissue and primers. *Biologia Plantarum*, **50**, 531-536.
Champagne, C. & Sinha, N. (2004) Compound leaves: equal to the sum of their parts? *Development*, **131**, 4401-4412.
Chan, K.M.A. & Moore, B.R. SymmeTREE: whole-tree analysis of differential diversification rates.
Chandappa, L.H., Feltus, F.A., Singh, H.P. & Paterson, A.H. (2005) Conserved PCR primers from rice-banana and rice-onion alignments. In XIII Plant-Animal Genome Conference, San Diego, USA.
Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, P.R., Haidar, N. & Savolainen, V. (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1889-1895.
Chat, J., Jauregui, B., Petit, R.J. & Nadot, S. (2004) Reticulate evolution in kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. *American Journal of Botany*, **91**, 736-747.
Chen, S., Guan, K., Zhou, Z., Olmstead, R. & Cronk, Q. (2005) Molecular phylogeny of *Incarvillea* (Bignoniaceae) based on ITS and trnL-F sequences. *American Journal of Botany*, **92**, 625-633.
Chernoff, M., Plitmann, U. & Kislev, M.E. (1992) Seed characters and testa texture in species of the Vicieae - their taxonomic significance. *Israel Journal of Botany*, **41**, 167-186.
Choi, H.-K., Luckow, M., Doyle, J. & Cook, D. (2006) Development of nuclear gene-derived molecular markers linked to legume genetic maps. *Molecular Genetics and Genomics*, **276**, 56-70.
Cieslak, T., Polepalli, J.S., White, A., Müller, K., Borsch, T., Barthlott, W., Steiger, J., Marchant, A. & Legendre, L. (2005) Phylogenetic analysis of *Pinguicula* (Lentibulariaceae): Chloroplast DNA sequences and morphology support several geographically distinct radiations. *American Journal of Botany*, **92**, 1723-1736.
Citerne, H.L. (2006) A primer set for specific amplification of two CYCLOIDEA-like genes in the genistoid clade of Leguminosae subfamily Papilionoideae. *Edinburgh Journal of Botany*, **62**, 119-126.
Citerne, H.L., Luo, D., Pennington, R.T., Coen, E. & Cronk, Q.C.B. (2003) A phylogenomic investigation of CYCLOIDEA-like TCP genes in the Leguminosae. *Plant Physiology*, **131**, 1042-1053.
Clark, R.M., Wagler, T.N., Quijada, P. & Doebley, J. (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature Genetics*, **38**, 594-597.
Clarkson, J.J., Knapp, S., Garcia, V.F., Olmstead, R.G., Leitch, A.R. & Chase, M.W. (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. *Molecular Phylogenetics and Evolution*, **33**, 75-90.
Clements, J.C., Dracup, M., Buirchell, B.J. & Smith, C.G. (2005a) Variation for seed coat and pod wall percentage and other traits in a germplasm collection and historical cultivars of lupins. *Australian Journal of Agricultural Research*, **56**, 75-83.
Clements, J.C., Sweetingham, M.W., Buirchell, B.J., Sipsas, S., Thomas, G., Quealy, J., Jones, R., Francis, C., Smith, C. & Francis, G. (2005b) The potential of the pearl

lupin (*Lupinus mutabilis*) for southern Australia. In Crop Updates, pp. 43.
Colless, D.H. (1980) Congruence between morphometric and allozyme data for *Menidia* species: a reappraisal. *Systematic Zoology*, **29**, 288-299.
Comes, H.P. & Abbott, R.J. (2001) Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). *Evolution*, **55**, 1943-1962.
Cong, B., Liu, J. & Tanksley, S.D. (2002) Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proceedings of the American Academy of Arts and Science*, **99**, 13606-13611.
Conterato, I.F. & Schifino-Wittmann, M.T. (2006) New chromosome numbers, meiotic behaviour and pollen fertility in American taxa of *Lupinus* (Leguminosae): contributions to taxonomic and evolutionary studies. *Botanical Journal of the Linnean Society*, **150**, 229-240.
Cook, L.G. & Crisp, M.D. (2005) Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography. *Journal of Biogeography*, **32**, 741-754.
Cook, O.F. (1925) Peru as a center of domestication - Tracing the origin of civilization through the domesticated plants. *Journal of Heredity*, 32-110.
Cowling, W.A., Buirchell, B. & Tapia, M. (1998) *Lupin Lupinus L.* Promoting the Conservation and Use of Underutilized and Neglected Crops 23. International Plant Genetic Resources Institute, Rome.
Cracraft, J. & Donoghue, M.J. (2004) Assembling the tree of life - where we stand at the beginning of the 21st century. In *Assembling the Tree of Life.* (eds. J. Cracraft & M.J. Donoghue), pp. 553-562. Oxford University Press, Oxford.
Crawford, D.J. & Mort, M.E. (2005) Phylogeny of eastern North American *Coreopsis* (Asteraceae-Coreopsideae): insights from nuclear and plastid sequences, and comments on character evolution. *American Journal of Botany*, **92**, 330-336.
Crayn, D.M., Winter, K. & Smith, J.A.C. (2004) Multiple origins of crassulacean acid metabolism and the epiphytic habit in the Neotropical family Bromeliaceae. *Proceedings of the National Academy of Sciences, USA*, **101**, 3703-3708.
Crisp, M.D., Gilmore, S. & Van Wyk, B.-E. (2000) Molecular phylogeny of the genistoid tribes of papilionoid legumes. In *Advances in Legume Systematics.* (eds. P.S. Herendeen & A. Bruneau), Vol. 9, pp. 249-276. Royal Botanic Garden Kew, London.
Cristofolini, G. (1989) A serological contribution to the systematics of the genus *Lupinus* (Fabaceae). *Plant Systematics and Evolution*, **166**, 265-278.
Cronn, R.C., Small, R.L., Haselkorn, T. & Wendel, J.F. (2002) Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany*, **89**, 707-725.
Cronquist, A. (1988) *The Evolution and Classification of Flowering Plants.*, 2nd edn. New York Botanical Garden, New York.
Cubas, P., Pardo, C. & Tahiri, H. (2005) Genetic variation and relationships among *Ulex* (Fabaceae) species in southern Spain and northern Morocco assessed by chloroplast microsatellite (cpSSR) markers. *American Journal of Botany*, **92**, 2031-2043.
Culley, T.M. & Wolfe, A.D. (2001) Population genetic structure of the cleistogamous plant species *Viola pubescens* Aiton (Violaceae) as indicated by allozyme and ISSR molecular markers. *Heredity*, **86**, 545-556.

Daehler, C.C. (2005) Upper-montane plant invasions in the Hawaiian Islands: patterns and opportunities. *Perspectives in Plant Ecology, Evolution and Systematics*, **7**, 203-216.
Dahlgren, R. (1970) Parallelism, convergence, and analogy in some South African genera of Leguminosae. *Botaniska Notiser*, **123**, 552-568.
D'Altroy, T.N. & Hastorf, C.A. (1984) The distribution and contents of Inca state storehouses in the Xauxa region of Peru. *American Antiquity*, **49**, 334-349.
Darlington, C.D. & Wylie, A.P. (1955) *Chromosome Atlas of Flowering Plants.*, 2nd edn. Allen & Unwin, London.
Darwin, C.R. (1868) *The Variation in Animals and Plants Under Domestication.* John Murray, London.
de Beer, G.R. (1971) *Homology, an Unsolved Problem.* Oxford University Press, London.
de Candolle, A. (1884) *Origin of Cultivated Plants.* Kegan Paul Trench & Co., London.
de Pinna, M.C.C. (1991) Concepts and tests of homology in the cladistic paradigm. *Cladistics*, **7**, 367-394.
de Queiroz, K. & Gauthier, J. (1990) Phylogeny as a central principle in taxonomy: phylogenetic definitions of taxon names. *Systematic Zoology*, **39**, 307-322.
de Queiroz, K. & Gauthier, J. (1992) Phylogenetic taxonomy. *Annual Review of Ecology, Evolution, and Systematics*, **23**, 449-480.
de Queiroz, K. & Gauthier, J. (1994) Towards a phylogenetic system of biological nomenclature. *Trends in Ecology and Evolution*, **9**, 27-31.
de Vogel, E.F. (1980) *Seedlings of Dicotyledons: Structure, Development, Types. Descriptions of 150 Woody Malesian Taxa* Centre for Agricultural Publishing and Documentation, Wageningen.
Debouck, D.G., Linan Jara, J.H., Campana Sierra, A. & de la Cruz Rojas, J.H. (1987) Observations on the domestication of *Phaseolus lunatus* L. *FAO/IBPGR Plant Genetic Resources Newsletter*, **70**, 26-32.
DeFilippis, V., R. & Moore, W.S. (2000) Resolution of phylogenetic relationships among recently evolved species as a function of amount of DNA sequence: An empirical study based on Woodpeckers (Aves: Picidae). *Molecular Phylogenetics and Evolution*, **16**, 143-160.
Delgado-Salinas, A., Turley, T., Richman, A. & Lavin, M.T. (1999) Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). *Systematic Botany*, **24**, 438-460.
Dengler, N.G. (1984) Comparison of leaf development in normal (+/+), entire (e/e), and lanceolate (La/+) plants of tomato, *Lycopersicon esculentum* 'Ailsa Craig'. *Botanical Gazette*, **145**, 66-77.
DeSalle, R., Egan, M.G. & Siddal, M. (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1905-1916.
Després, L., Gielly, L., Redoutet, B. & Taberlet, P. (2003) Using AFLP to resolve phylogenetic relationships in a morphologically diverse plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Molecular Phylogenetics and Evolution*, **27**, 185-196.
Diamond, J. (1997) *Guns, Germs, and Steel. The Fate of Human Societies.* W.W. Norton & Co., New York.
Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, **418**, 700-707.

Diers, L. (1961) Der anteil an polyploiden in den vegetationsgüteln der westkordillere Perus. *Zeitschrift für Botanik*, **49**, 437-488.
Doebley, J. (2006) Plant science: unfallen grains: how ancient farmers turned weeds into crops. *Science*, **312**, 1318-1319.
Doebley, J., Stec, A. & Hubbard, L. (1997) The evolution of apical dominance in maize. *Nature*, **386**, 485-488.
Doebley, J. & Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. *The Plant Cell*, **10**, 1075-1082.
Doganlar, S., Frary, A. & Tanksley, S.D. (2000) The genetic basis of seed-weight variation: tomato as a model system. *Theoretical and Applied Genetics*, **100**, 1267-1273.
Donoghue, M.J., Bell, C.D. & Winkworth, R.C. (2003) The evolution of reproductive characters in Dipsacales. *International Journal of Plant Sciences*, **164**, S453-S464.
Donoghue, M.J. & Cracraft, J. (2004) Introduction - charting the tree of life. In *Assembling the Tree of Life.* (eds. J. Cracraft & M.J. Donoghue), pp. 1-4. Oxford University Press, Oxford.
Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F. & Douzery, E.J.P. (2003) Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molecular Biology and Evolution*, **20**, 248-254.
Doyle, J.A. & Hickey, L.J. (1976) Pollen and leaves from the mid-Cretaceous Potomac Group and their bearing on early angiosperm evolution. In *Origin and Early Evolution of the Angiosperms.* (ed. C.A. Beck). Columbia University Press, New York.
Doyle, J.J. (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany*, **17**, 144-163.
Doyle, J.J. & Doyle, J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11-15.
Doyle, J.J., Kanazin, V. & Shoemaker, R.C. (1996) Phylogenetic utility of histone H3 intron sequences in the perennial relatives of Soybean (*Glycine*: Leguminosae). *Molecular Phylogenetics and Evolution*, **6**, 438-447.
Doyle, J.J. & Doyle, J.L. (1999) Nuclear protein-coding genes in phylogeny reconstruction and homology assessment: some examples from Leguminosae. In *Molecular Systematics and Plant Evolution.* (eds. P.M. Hollingsworth, R.M. Bateman & R.J. Gornall), Vol. 1, pp. 229-254. Taylor and Francis, London.
Doyle, J.J., Chappill, J., Bailey, C.D. & Kajita, T. (2000) Towards a comprehensive phylogeny of legumes: evidence from rbcL sequences and non-molecular data. In *Advances in Legume Systematics.* (eds. P.S. Herendeen & A. Bruneau), Vol. 9, pp. 1-20. Royal Botanic Garden, Kew.
Doyle, J.J. & Luckow, M.A. (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiology*, **131**, 900-910.
Doyle, J.J., Doyle, J.L. & Harbison, C. (2003a) Chloroplast expressed glutamine synthetase in *Glycine* and related Leguminosae: phylogeny, gene duplication and ancient polyploidy. *Systematic Botany*, **28**, 567-577.
Doyle, J.J., Doyle, J.L., Rauscher, J.T. & Brown, A.H.D. (2003b) Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus Glycine). *New Phytologist*, **161**, 121-132.
Doyle, J.J., Doyle, J.L., Rauscher, J.T. & Brown, A.H.D. (2004) Evolution of the perennial soybean polyploid complex (*Glycine* subgenus Glycine): a study of contrasts. *Biological Journal of the Linnean Society*, **82**, 583-597.

Driskell, A.C., Burleigh, J.G., Ane, C., McMahon, M.M., O'Meara, B.C. & Sanderson, M.J. (2004) Prospects for building the Tree of Life from large sequence databases. *Science*, **306**, 1172-1174.
Drummond, C. (2004) Phylogeny of *Lupinus* (Fabaceae) based on non-coding chloroplast DNA. Botany 2004 Conference, Society of America, Snowbird, Utah.
Drummond, C. (unpublished manuscript) Phylogeny of *Lupinus* (Fabaceae) based on non-coding DNA.
Dübbern de Souza, F.H. & Marcos-Filho, J. (2001) The seed coat as a modulator of seed-environment relationships in Fabaceae. *Revista Brasileira de Botânica*, **24**, 365-375.
Duke, J.A. (1969) On tropical tree seedlings I. Seeds, seedlings, systems, and systematics. *Annals of the Missouri Botanical Garden*, **56**, 125-161.
Duke, J.A. (1981) *Handbook of Legumes of World Economic Importance.* Plenum Press, New York and London.
Dunn, D.B. (1955) Taxonomy of *Lupinus*, group Micranthi (Leguminosae) of the Pacific coast. *Aliso*, **3**, 135-171.
Dunn, D.B. (1957) A revision of the *Lupinus arbustus* complex of the Laxiflori. *Madroño*, **14**, 54-73.
Dunn, D.B. (1959) *Lupinus pusillus* and its relationship. *The American Midland Naturalist*, **62**, 500-510.
Dunn, D.B. (1971) A case of long range dispersal and "rapid speciation" in *Lupinus*. *Transactions, Missouri Academy of Science*, **5**, 26-38.
Dunn, D.B. (1984) Cyto-taxonomy and distribution of the New World lupin species. In *Proceedings of the 3rd International Lupin Conference*, pp. 67-86, La Rochelle, France.
Dunn, D.B. & Gillett, J. (1966) The lupines of Canada and Alaska. *Canada Department of Agriculture Research Branch Monograph 2*, 87.
Dunn, D.B. & Planchuelo, A.M. (1981) *Lupinus heptaphyllus* (Velloso) Hassler vs. *Lupinus hilarianus* Bentham. *Taxon*, **30**, 464-470.
Eames, A.J. (1961) *Morphology of Angiosperms.* McGraw Hill, New York.
Ehrenreich, I.M. & Purugganan, M.D. (2006) The molecular genetic basis of plant adaptation. *American Journal of Botany*, **93**, 953-962.
Eiten, G. (1972) The cerrado vegetation of Brazil. *Botanical Review*, **38**, 201-341.
Eldredge, N. (1992) *Systematics, Ecology, and the Biodiversity Crisis.* Columbia University Press, New York ; Oxford.
Eldredge, N. & Cracraft, J. (1980) *Phylogenetic Pattern and the Evolutionary Process.* Columbia University Press, New York.
Ellison, N.W., Liston, A., Steiner, J.J., Williams, W.M. & Taylor, N.L. (2006) Molecular phylogenetics of the clover genus (*Trifolium*-Leguminosae). *Molecular Phylogenetics and Evolution*, **39**, 688-705.
El-Rabey, H.A., Badr, A., Schäfer-Pregl, R., Martin, W. & Salamini, F. (2002) Speciation and species separation in *Hordeum* L. (Poaceae) resolved by discontinuous molecular markers. *Plant Biology*, **4**, 567-575.
Emshwiller, E. (2006) Genetic data and plant domestication. In *Documenting Domestication, New Genetic and Archaeological Paradigms.* (eds. M.A. Zeder, D.G. Bradley, E. Emshwiller & B.D. Smith), pp. 99-122. University of California Press, Berkeley.
Emshwiller, E. & Doyle, J.J. (1998) Origins of domestication and polyploidy in oca

(*Oxalis tuberosa*: Oxalidaceae): nrDNA ITS data. *American Journal of Botany*, **85**, 975-985.
Emshwiller, E. & Doyle, J.J. (1999) Chloroplast-expressed glutamine synthetase (ncpGS): potential utility for phylogenetic analysis with an example from *Oxalis* (Oxalidaceae). *Molecular Phylogenetics and Evolution*, **12**, 310-319.
Emshwiller, E. & Doyle, J.J. (2002) Origins of domestication and polyploidy in oca (*Oxalis tuberosa*: Oxalidacae) 2. Chloroplast-expressed glutamine synthetase data. *American Journal of Botany*, **89**, 1042-1056.
Endress, P.K. (2002) Morphology and angiosperm systematics in the molecular era. *The Botanical Review*, **68**, 545-570.
Engel, F. (1970) Exploration of the Chilca Canyon, Peru. *Current Anthropology*, **11**, 55-58.
Erickson, D.L., Smith, B.D., Clarke, A.C., Sandweiss, D.H. & Tuross, N. (2005) An Asian origin for a 10,000-year-old domesticated plant in the Americas. *Proceedings of the National Academy of Sciences, USA*, **102**, 18315-18320.
Eubanks, M.W. (1997) Molecular analysis of crosses between *Tripsacum dactyloides* and *Zea diploperennis* (Poaceae). *Theoretical and Applied Genetics*, **94**, 707-712.
Evans, R.C., Alice, L.A., Campbell, C.S., Kellogg, E.A. & Dickinson, T.A. (2000) The granule-bound starch synthase (GBSSI) gene in the Rosaceae: multiple loci and phylogenetic utility. *Molecular Phylogenetics and Evolution*, **17**, 388-400.
Fahn, A. (1982) *Plant Anatomy*, 3rd edn. Pergamon Press, Oxford.
Farris, J.S. (1970) Methods for computing Wagner trees. *Systematic Zoology*, **19**, 83-92.
Farris, J.S., Källersjö, M., Kluge, A.G. & Bult, C. (1995) Testing significance of incongruence. *Cladistics*, **10**, 315-319.
Felsenstein, J. (2004a) *Inferring Phylogenies.*, First edn. Sinauer Associates Ltd, Sunderland, Massachusetts.
Felsenstein, J. (2004b) Restriction sites, RAPDs, AFLPs, and microsatellites. In *Inferring phylogenies*, pp. 230-247. Sinauer Associates Inc., Sunderland, Massachusetts.
Feng, Y., Oh, S.-H. & Manos, P.S. (2006) Phylogeny and historical biogeography of the genus *Platanus* as inferred from nuclear and chloroplast DNA. *Systematic Botany*, **30**, 786-799.
Ferguson, D. & Sang, T. (2001) Speciation through homoploid hybridisation between allotetraploids in peonies (*Paeonia*). *Proceedings of the National Academy of Sciences, USA*, **98**, 3915-3919.
Fishbein, M., Hibsch-Jetter, C., Soltis, D.E. & Hufford, L. (2001) Phylogeny of Saxifragales (Angiosperms, Eudicots): analysis of a rapid, ancient radiation. *Systematic Biology*, **50**, 817-847.
Fishbein, M. & Soltis, D.E. (2004) Further resolution of the rapid radiation of Saxifragales (Angiosperms, Eudicots) supported by mixed-model Bayesian analysis. *Systematic Botany*, **29**, 883-891.
Flores, E.M. (1980) Shoot vascular system and phyllotaxis of *Casuarina* (Casuarinaceae). *American Journal of Botany*, **67**, 131-140.
Fofana, B., Baudoin, J.P., Vekemans, X., Debouck, D.G. & du Jardin, P. (1999) Molecular evidence for an Andean origin and a secondary gene pool for the Lima bean (*Phaseolus lunatus* L.) using chloroplast DNA. *Theoretical and Applied Genetics*, **98**, 202-212.
Fondon, J.W., III & Garner, H.R. (2004) Molecular origins of rapid and continuous morphological evolution. *Proceedings of the National Academy of Sciences, USA*,

101, 18058-18063.
Food and Agriculture Organisation of the United Nations (1996) Food requirements and population growth. In World Food Summit, Vol. 1. FAO.
Ford, V.S. & Gottlieb, L.D. (1999) Molecular characterization of *PgiCI* in a tetraploid plant and its diploid relatives. *Evolution*, **53**, 1060-1067.
Ford, V.S. & Gottlieb, L.D. (2003) Reasseement of phylogenetic relationships in *Clarkia* Sect. *Sympherica*. *American Journal of Botany*, **90**, 284-292.
Forey, P.L. (1992) *Cladistics: A Practical Course in Systematics.* Clarendon Press, Oxford.
Forey, P.L., Humphries, C.J. & Vane-Wright, R.I. (1994) *Systematics and Conservation Evaluation.* Clarendon Press, Oxford.
Foster, A.S. (1951) Heterophylly and foliar venation in *Lacunaria*. *Bulletin of the Torrey Botanic Club*, **78**, 382-400.
Franck, D.H. (1976) Comparative morphology and early leaf histogenesis of adult and juvenile leaves of *Darlingtonia californica* and their bearing on the concept of heterophylly. *Botanical Gazette*, **137**, 20-34.
Friesen, N., Pollner, S., Bachmann, K. & Blattner, F.R. (1999) RAPDs and noncoding chloroplast DNA reveal a single origin of the cultivated *Allium fistulosum* from *A. altaicum* (Alliaceae). *American Journal of Botany*, **86**, 554-562.
Fritsch, P.W. (2001) Phylogeny and biogeography of the flowering plant genus *Styrax* (Styracaceae) based on chloroplast DNA restriction sites and DNA sequences of the internal transcribed spacer region. *Molecular Phylogenetics and Evolution*, **19**, 387-409.
Fuentes, E., Perisse, P. & Planchuelo, A.M. (1993) Seed morphology and lipids analysis of cultivated and wild species of *Lupinus*. In *Proceedings of 7th International Lupin Conference*, pp. 185-190, Edora, Portugal.
Fulton, T.M., Van der Hoeven, R., Eannetta, N.T. & Tanksley, S.D. (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell*, **14**, 1457-1467.
Gade, D.W. (1966) Achira, the edible *Canna*, its cultivation and use in the Peruvian Andes. *Economic Botany*, 407-415.
Galloway, G.L., Malmberg, R.L. & Price, R.A. (1998) Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. *Molecular Biology of Evolution*, **15**, 1312-1320.
Garcia, E.H., Pena-Valdiva, C.B., Rogelio Aguirre, J.R. & Muruaga, J.S.M. (1997) Morphological and agronomic traits of a wild population and improved cultivar of common bean (*Phaseolus vulgaris* L.). *Annals of Botany*, **79**, 207-213.
Gavrilets, S. & Vose, A. (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences, USA*, **102**, 18040-28045.
Gene Codes Corporation. (1991-1998) Sequencher 3.1.1.
Gentry, A.H. (1982) Neotropical floristic diversity: phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Annals of Missouri Botanical Garden*, **69**, 557-593.
Gepts, P. (1990) Biochemical evidence bearing on the domestication of *Phaseolus* (Fabaceae) beans. *Economic Botany*, **44**, 28-38.
Gepts, P. (1993) The use of molecular and biochemical markers in crop evolution studies. *Evolutionary Biology*, **27**, 51-94.
Gepts, P. (1996) Origin and evolution of cultivated *Phaseolus* species. In *Advances in*

Legume Systematics 8: Legumes of Economic Importance. (eds. B. Pickersgill & J.M. Lock), pp. 65-74. Royal Botanic Gardens, Kew.

Geyer, P.S., Larson, T.S. & Stroik, L. (2003) University of San Diego palynological investigation of the Dos Cabezas giants. *International Journal of Osteoarchaeology*, **13**, 275-282.

Ghosh, P., Garzione, C.N. & Eiler, J.M. (2006) Rapid uplift of the Altiplano revealed through ^{13}C-^{18}O bonds in paleosol carbonates. *Science*, **311**, 511-515.

Gil-ad, N.L. (1998) The micromorphologies of seed coats and petal trichomes of the taxa of *Viola* subsect. Boreali-Americanae (Violaceae) and their utility in discerning orthospecies from hybrids. *Brittonia*, **50**, 91-121.

Gill, N.T., Vear, K.C. & Barnard, D.J. (1980) *Agricultural Botany 1. Dicotyledonous Crops*, 3rd edn. Duckworth, London.

Gimenes, M.A., Lopes, C.R. & Valls, J., F.M. (2002) Genetic relationships among *Arachis* species based on AFLP. *Genetic and Molecular Biology*, **25**, 349-353.

Giuffra, E., Kijas, J.M.H., Amarger, V., Carlborg, O., Jeon, J.T. & Andersson, L. (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, **154**, 1785-1791.

Givnish, T.J., Systma, K.J., Smith, J.F. & Hahn, W.J. (1994) Thorn-prickles and heterophylly in *Cyanea*: adaptations to extinct avian browers in Hawaii? *Proceedings of the National Academy of Sciences, USA*, **91**, 2810-2814.

Givnish, T.J. & Sytsma, K.J. (1997) Homoplasy in molecular vs. morphological data: the likelihood of correct phylogenetic inference. In *Molecular Evolution and Adaptive Radiation.* (eds. T.J. Givnish & K.J. Sytsma), pp. 55-101. Cambridge University Press, Cambridge.

Gladstones, J.S. (1974) Lupins of the Mediterranean region and Africa. *Western Australia Department of Agriculture Technical Bulletin no. 26*, 47.

Gladstones, J.S. (1980) Recent developments in the understanding, improvement and use of *Lupinus*. In *Advances in Legume Science.* (eds. R.J. Summerfield & A.H. Bunting), pp. 603-612. Royal Botanic Gardens; Ministry of Agriculture Fisheries and Food, Kew.

Gladstones, J.S. (1984) Present situation and potential of Mediterranean/Africa lupins for crop production. In *Proceedings of the 3rd International Lupin Conference*, pp. 17-38, La Rochelle, France.

Gladstones, J.S. (1998) Distribution, origin, taxonomy, history and importance. In *Lupins As Crop Plants: Biology, Production and Utilization.* (eds. J.S. Gladstones, C. Atkins & J. Hamblin), pp. 1-40. CABI International, Wallingford/New York.

Gladstones, J.S., Atkins, C.A. & Hamblin, J., eds.. (1998) *Lupins As Crop Plants: Biology, Production and Utilization.* CABI, Wallingford.

Gleissberg, S. & Kadereit, J.W. (1999) Evolution of leaf morphogenesis: evidence from developmental and phylogenetic data in Papaveraceae. *International Journal of Plant Sciences*, **160**, 787-794.

Goldblatt, P. (1981) Cytology and phylogeny of *Leguminosae*. In *Advances in Legumes Systematics, Volume 2* (eds. R.M. Polhill & P.H. Raven), pp. 427-464. Royal Botanic Gardens, Kew.

Goldblatt, P., Savolainen, V., Porteous, O., Sostaric, I., Powell, M., Reeves, G., Manning, J.C., Barraclough, T.G. & Chase, M.W. (2002) Radiation in the Cape flora and the phylogeny of peacock irises *Moraea* (Iridaceae) based on four plastid DNA regions. *Molecular Phylogenetics and Evolution*, **25**, 341-360.

Goloboff, P. (1999) NONA. Goloboff, P., Tucumán, Argentina.
Graham, A. (1992) Utilization of the isthmian land bridge during the Cenozoic - paleobotanical evidence for timing, and the selective influence of altitudes and climate. *Review of Palaeobotany and Palynology*, **72**, 119-128.
Graham, A. (1995) Development of affinities between Mexican/Central American and northern South American lowland and lower montane vegetation during the Tertiary. In *Biodiversity and Conservation of Neotropical Montane Forests.* (eds S.P. Churchill, H. Balslev, E. Forero & J.L. Luteyn), pp. 11-22. The New York Botanical Garden, New York.
Grant, V. (2003) Incongruence between cladistic and taxonomic systems. *American Journal of Botany*, **90**, 1263-1270.
Gravendeel, B., Smithson, A., Slik, F.J.W. & Schuiteman, A. (2004) Epiphytism and pollinator specialization: drivers for orchid diversity? *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 1523-1535.
Gregory-Wodzicki, K.M. (2000) Uplift history of the Central and Northern Andes: a review. *Geographical Society of America Bulletin*, **112**, 1091-1105.
Greuter, W. (2000) *International Code of Botanical Nomenclature (Saint Louis Code) Adopted by the Sixteenth International Botanical Congress, St Louis, Missouri, July-August 1999* Koeltz, Kèonigstein.
Grigg, S., P., Canales, C., Hay, A. & Tsiantis, M. (2005) SERRATE coordinates shoot meristem function and leaf axial patterning in *Arabidopsis*. *Nature*, **437**, 1023-1026.
Gross, B.L., Schwarzbach, A.E. & Rieseberg, L.H. (2003) Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). *American Journal of Botany*, **90**, 1708-1719.
Gross, R. (1982) *El cultivo y la utilización del Lupinus mutabilis Sweet.* FAO, Rome.
Gross, R. (1986) First Reinhold von Sengbusch Memorial lecture: lupins in the old and new world - a biological-cultural coevolution. In *Proceedings of the 4th International Lupin Conference*, Vol. 1, pp. 244-277, Geraldton, Western Australia.
Gupta, M. (1991) Seed coat structure in some species of *Trigonella*. *Scanning Microscopy*, **5**, 787-796.
Gustaffson, M.H.G. & Albert, V.A. (1999) Inferior ovaries and angiosperm diversification. In *Molecular Systematics and Plant Evolution.* (eds. P.M. Hollingsworth, R.M. Bateman & R.J. Gornall), pp. 403-431. Taylor and Francis, London.
Guzman, B. & Vargas, P. (2005) Systematics, character evolution, and biogeography of *Cistus* L. (Cistaceae) based on ITS, trnL-trnF, and matK sequences. *Molecular Phylogenetics and Evolution*, **37**, 644-660.
Hall, J.C., Iltis, H.H. & Sytsma, K.J. (2004) Molecular phylogenetics of core brassicales, placement of orphan genera *Emblingia*, *Forchhammeria*, *Tirania*, and character evolution. *Systematic Botany*, **29**, 654-669.
Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.
Halpern, S.L. (2005) Sources and consequences of seed size and variation in *Lupinus perennis* (Fabaceae): adaptive and non-adaptive hypothesis. *American Journal of Botany*, **92**, 205-213.
Hancock, J.F. (2004a) The Dynamics of Plant Domestication. In *Plant Evolution and the*

Origin of Crop Species, pp. 151- 173. CABI Publishing.
Hancock, J.F. (2004b) The Origin of Agriculture. In *Plant Evolution and the Origin of Crop Species*, pp. 129-150. CABI Publishing.
Hancock, J.F. (2004c) *Plant Evolution and the Origin of Crop Species*, 2nd edn. CABI Publishing, Wallingford.
Harlan, J.R. (1975) *Crops and Man.* American Society of Agronomy, Madison, Wisconsin.
Harlin, M. (2003) Taxon names as paradigms: the structure of nomenclatural revolutions. *Cladistics*, **19**, 138-143.
Harries, M. (2005) Lupin and pulse production agronomy and genetic improvement. In Agribusiness Crop Updates.
Harris, D.J., Poulsen, A.D., Frimodt-Møller, C., Preston, J. & Cronk, Q., C.B . (2000) Rapid radiation of *Aframomum* (Zingiberaceae): evidence from nuclear ribosomal DNA internal transcribed spaces (ITS) sequences. *Edinburgh Journal of Botany*, **57**, 377-395.
Harris, S.A. (1999) RAPDs in systematics - a useful methodology? In *Molecular Systematics and Plant Evolution.* (eds. P.M. Hollingsworth, R.M. Bateman & R.J. Gornall). Taylor and Francis, London.
Harris, W.M. (1987) Comparative ultrastructure of developing seed coats of 'hard-seeded' and 'soft-seeded' varieties of soybea, *Glycine max* (L.) Merr. *Botanical Gazette*, **148**, 324-331.
Harrison, C.J., Corley, S.B., Moylan, E.C., Alexander, D.L., Scotland, R.W. & Langdale, J.A. (2005) Independent recruitment of a conserved developmental mechanism during leaf development. *Nature*, **434**, 509-514.
Hastorf, C.A. (1993) *Agriculture and the Onset of Politcal Inequality Before the Inka.*, 1st edn. Cambridge University Press, Cambridge.
Hastorf, C.A. (1999) Cultural Implications of Crop Introductions in Andean Prehistory. In *The Prehistory of Food: Appetites for Change.* (eds. C. Gosden & J.G. Hather), pp. 35-58. Routledge, London and New York.
Hastorf, C.A. & Popper, V.S. (1988) *Current Paleoethnobotany: Analytical Methods and Cultural Interpretations of Archaeological Plant Remains.* University of Chicago Press, Chicago.
Hawkes, J.G. (1983) *The Diversity of Crop Plants.* Harvard University Press, Cambridge, Mass.
Hawkins, J.A. (2000) A survey of primary homology assessment: different botanists perceive and define characters in different ways. In *Homology and Systematics.* (eds. R.W. Scotland & R.T. Pennington), pp. 22-53. Taylor and Francis, London.
Hawkins, J.A. (2002) Evolutionary developmental biology: impact on systematic theory and practice, and the contribution of systematics. In *Developmental Genetics and Plant Evolution.* (eds. Q.C.B. Cronk, R.M. Bateman & J.A. Hawkins), pp. 32-51. Taylor & Francis, London.
Hawkins, J.A. (2006) Using phylogeny to investigate the origins of the Cape flora: the importance of taxonomic, gene and genome sampling strategies. *Diversity and Distributions*, **12**, 27-33.
Hawkins, J.A., Hughes, C.E. & Scotland, R.W. (1997) Primary homology assessment, characters and character states. *Cladistics*, **13**, 275-283.
Hayashi, K., Yoshida, S., Utechs, F. & Kawano, S. (2001) Molecular systematics in the genus *Clintonia* and related taxa based on rbcL and matK gene sequence data.

Plant Species Biology, **16**, 119-137.
He, C., Poysa, V. & Yu, K. (2003) Development and characterisation of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. *Theoretical and Applied Genetics*, **106**, 363-373.
Hedges, S.B. & Maxson, L.R. (1996) Re: molecules and morphology in amniote phylogeny. *Molecular Phylogenetics and Evolution*, **6**, 312-314.
Hegarty, M. & Hiscock, S.J. (2005) Hybrid speciation in plants: new insights from molecular studies. *New Phytologist*, **165**, 411-423.
Heiser, C.B. (1969) Systematics and the origin of cultivated plants. *Taxon*, **18**, 36-45.
Helbig, A.J., Knox, A.G., Parkin, D.T., Sangster, G. & Collinson, M. (2002) Guidelines for assigning species rank. *Ibis*, **144**, 518-525.
Helbig, A.J., Kocum, A., Seibold, I. & Braun, M.J. (2005) A multi-gene phylogeny of aquiline eagles (Aves: Acciptriformes) reveals extensive paraphyly at genus level. *Molecular Phylogenetics and Evolution*, **35**, 147-164.
Heyn, C.C. & Herrnstadt, L. (1977) Seed coat structure of Old World *Lupinus* species. *Botanica Nordic*, **130**, 427-435.
Hillis, D.M. (1996) Inferring complex phylogenies. *Nature*, **383**, 130-131.
Hillis, D.M. (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology*, **47**, 3-8.
Hillis, D.M., Huelsenbeck, J.P. & Swofford, D.L. (1994) Hobgoblin of phylogenetics? *Nature*, **369**, 363-364.
Hillis, D.M., Moritz, C. & Mable, B.K. (1996) *Molecular Systematics.*, 2nd edn. Sinauer Associates, Sunderland, Mass.
Hillis, D.M. & Wiens, J.J. (2000) Molecules versus morphology in systematics: conflicts, artifacts, and misconceptions. In *Phylogenetic Analysis of Morphological Data.* (ed. J.J. Wiens), pp. 1-19. Smithsonian Institution Press, Washington.
Hillis, D.M., Heath, T.A. & St. John, K. (2005) Analysis and visualization of tree space. *Systematic Biology*, **54**, 471-482.
Hilu, K.W., Borsch, T., Muller, K., Soltis, D.E., Soltis, P.S., Savolainen, V., Chase, M.W., Powell, M.P., Alice, L.A., Evans, R.C., Suauquet, H., Neinhuis, C., Slotta, T.A.B., Rohwer, J.G., Campbell, C.S. & Chatrou, L.W. (2003) Angiosperm phylogeny based on matK sequence information. *American Journal of Botany*, **90**, 1758-1776.
Hodges, S.A. (1997) Rapid radiation due to a key innovation in Columbines (Ranunculaceae: Aquilegia). In *Molecular Evolution and Adaptive Radiation.* (eds. T.J. Givnish & K.J. Sytsma), pp. 391-128. Cambridge University Press, Cambridge.
Hodges, S.A. & Arnold, M.L. (1994) Columbine: a geographically widespread flock. *Proceedings of the National Academy of Sciences, USA*, **91**, 5129-5132.
Hofer, J., Turner, L., Hellens, R., Ambrose, M., Matthews, P., Michael, A. & Ellis, N. (1997) UNIFOLIATA regulates leaf and flower morphogenesis in pea. *Current Biology*, **7**, 581-587.
Hondelmann, W. (1984) The lupin - ancient and modern crop. *Theoretical and Applied Genetics*, **68**, 1-9.
Hooghiemstra, H. & Cleef, A.M. (1995) Pleistocene climatic change and environmental and generic dynamics in the north Andean montane forest and páramo. In *Biodiversity and Conservation of Neotropical Montane Forests.* (eds. S.P. Churchill, H. Balslev, E. Forero & J.L. Luteyn), pp. 35-49. The New York Botanical Garden,

New York.
Hooghiemstra, H. & Van der Hammen, T. (2004) Quaternary Ice-Age dynamics in the Colombian Andes: developing an understanding of our legacy. *Proceedings of the Royal Society of London. B.*, **359**, 173-181.
Hopkins, H.C.F. & Silva, M.F.d. (1986) *Parkia (Leguminosae--Mimosoideae)* Published for Organization for Flora Neotropica by the New York Botanical Garden, New York.
Horandl, E., Paun, O., Johansson, J.T., Lehnebach, C., Armstrong, T., Chen, L. & Lockhart, P. (2005) Phylogenetic relationships and evolutionary traits in Ranunculus s.l. (Ranunculaceae) inferred from ITS sequence analysis. *Molecular Phylogenetics and Evolution*, **36**, 305-327.
Howarth, D.G. & Donoghue, M.J. (2006) Phylogenetic analysis of the "ECE" (CYC/TB1) clade reveals duplications predating the core eudicots. *Proceedings of the National Academy of Sciences, USA*, **103**, 9101-9106.
Huelsenbeck, J.P., Bull, J.J. & Cunningham, C.W. (1996) Combining data in phylogenetic analysis. *Trends in Ecology and Evolution*, **11**, 152-158.
Huelsenbeck, J.P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754-755.
Hughes, C.E., Bailey, C.D. & Harris, S.A. (2002) Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: Insights into polyploid origins and nrDNA polymorphism. *American Journal of Botany*, **89**, 1057-1073.
Hughes, C.E. & Eastwood, R.J. (2006) Island radiation on a continental scale: exceptional rates of species diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences, USA*, **103**, 10334-10339.
Hughes, C.E., Eastwood, R.J. & Bailey, C.D. (2006) From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**, 211-225.
Humphries, C.J. & Funk, V.A. (1984) 17 Cladistic Methodology. In *Current Concepts in Plant Taxonomy* (eds. V.H. Heywood & D.M. Moore), Vol. 25, pp. 323-363. Academic Press, London and Orlando.
Hutchinson, J. (1964) Tribe 12. *Lupineae*. In *The Genera of Flowering Plants (Angiospermae): Based Principally on the Genera Plantarum of G. Bentham and J.D. Hooker.*, Vol. 1, pp. 363-364. Clarendon Press, Oxford.
Ingram, A.L. & Doyle, J.J. (2003) The origin and evolution of *Eragrostis tef* (Poaceae) and related polyploids: evidence from nuclear *waxy* and plastid *rps16*. *American Journal of Botany*, **90**, 166-122.
Initiative, T.A.G. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796-815.
International Board for Genetic Resources, ed. (1981) *Lupin Descriptors*. IBPGR, Rome.
International Commission on Zoological Nomenclature, (1999) *International Code of Zoological Nomenclature.*, 4th edn. International Trust for Zoological Nomenclature, London.
Isaacs, M.J.H., Weitz, F.M. & Johnson, C.T. (1993) Seed-coat characteristics of selected Southern African species of *Polygala* L (Polygalaceae). *South African Journal of Botany*, **59**, 592-596.
Isely, D. (1982) Leguminosae and *Homo sapiens*. *Economic Botany*, **36**, 46-70.
Ishizaki, O., Nishida, I., Agata, K., Eguchi, G. & Murata, N. (1988) Cloning and

nucleotide sequence of cDNA for the plastid glycerol-3-phosphate acyltransferase from squash. *FEBS Letters*, **238**, 424-430.
Jackman, T.R., Larson, A., de Queiroz, K. & Losos, J.B. (1999) Phylogenetic relationships and tempo of early diversification in *Anolis* lizards. *Systematic Biology*, **48**, 254-285.
Jackson, B.D. (1928) *A Glossary of Botanical Terms*, 4th edn. Duckworth, London.
Janzen, D. (1967) Why mountain passes are higher in the tropics? *The American Naturalist*, **101**, 233-249.
Janzen, D. (2004) Now is the time. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 731-732.
Jardine, N. (1969a) A logical basis for biological classification. *Systematic Zoology*, 37-52.
Jardine, N. (1969b) The observational and theoretical components of homology: a study based on the morphology of the dermal skull-roofs of rhipidistian fishes. *Biological Journal of the Linnean Society*, **1**, 327-361.
Jenkins, G. & Rees, H. (1991) Strategies of bivalent formation in allopolyploid plants. *Proceedings of the Royal Society of London. B.*, **243**, 209-214.
Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A. & Karp, A. (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, **3**, 381-390.
Jørgensen, P. & Lâeon-Yâanez, S. (1999) *Catalogue of the Vascular Plants of Ecuador.* Missouri Botanical Garden Press, St. Louis, Mo.
Kadereit, G. (2003) Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C4 photosynthesis. *International Journal of Plant Science*, **164**, 959-986.
Källersjö, M., Albert, V.A. & Farris, J.S. (1999) Homoplasy increases phylogenetic structure. *Cladistics*, **15**, 91-93.
Kami, J., Becerra Velàsquez, V., Debouck, D.G. & Gepts, P. (1995) Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. *Proceedings of the National Academy of Sciences, USA*, **92**, 1101-1104.
Kaplan, D., R. (1975) Unifacial leaves in monocotyledons. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie*, **95**, 1-105.
Kaplan, L. (1980) Variation in the cultivated beans. In *Guitarrero Cave Early Man in the Andes.* (ed. T.F. Lynch), pp. 145-148. Academic Press, New York.
Kaplan, L. & Lynch, T.F. (1999) *Phaseolus* (Fabaceae) in archaeology: AMS radiocarbon dates and their significance for pre-colombian agriculture. *Economic Botany*, **53**, 261-272.
Kashi, Y., King, D. & Sollier, M. (1996) Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics*, **13**, 74-78.
Kass, E. & Wink, M. (1996) Molecular phylogeny and phylogeography of the genus *Lupinus*. *In Towards the 21st Century. Proceedings of the 8th International Lupin Conference,* pp. 248-254, Asilomar, California.
Kass, E. & Wink, M. (1997a) Molecular phylogeny and phylogeography of *Lupinus* (Leguminosae) inferred from nucleotide sequences of the rbcL gene and ITS 1 + 2 regions of rDNA. *Plant Systematics and Evolution*, **208**, 139-167.
Kass, E. & Wink, M. (1997b) Phylogenetic relationships in the Papilionoideae (family Leguminosae) based on nucleotide sequences of cpDNA (rbcL) and nrDNA (ITS

1 and 2). *Molecular Phylogenetics and Evolution*, **8**, 65-88.
Kay, K.M., Reeves, P.A., Olmstead, R.G. & Schemske, D.W. (2005) Rapid speciation and the evolution of hummingbird pollination in neotropical *Costus* subgenus *Costus* (Costaceae): evidence from nrDNA ITS and ETS sequences. *American Journal of Botany*, **92**, 1899-1910.
Kazimierski, T. (1984) Spontaneous polyploids in lupine. In *Proceedings of the 3rd International Lupin Conference*, Vol. 1, pp. 535-536, La Rochelle, France.
Kazimierski, T. & Nowacki, E. (1961) Indigenous species of lupins regarded as the initial forms of the cultivated species *Lupinus albus* and *Lupinus mutabilis*. *Flora*, **151**, 202-209.
Keatinge, R.W. (1988) *Peruvian Prehistory: An Overview of Pre-Inca and Inca Society.* Cambridge University Press, Cambridge.
Keepax, C. (1977) Contamination of archaeological deposits by seeds of modern origin with particular reference to the use of flotation. *Journal of Archaeological Sciences*, **4**, 221-229.
Kelchner, S.A. & Clark, L.G. (1997) Molecular evolution and phylogenetic utility of the chloroplast *rp116* intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics and Evolution*, **8**, 385-397.
Kenicer, G.J., Kajita, T., Pennington, R.T. & Murata, J. (2005) Systematics and biogeography of *Lathyrus* (Leguminosae) based on internal transcribed spacer and cpDNA sequence data. *American Journal of Botany*, **92**, 1199-1209.
Kim, C.H. & Sun, B.Y. (2004) Infrageneric classification of the genus Eleutherococcus Maxim. (Araliaceae) with a new section Cissifolius. *Journal of Plant Biology*, **47**, 282-288.
Kim, H.-G., Keeley, S.C., Vroom, P.S. & Jansen, R.K. (1998) Molecular evidence for an African origin of the Hawaiian endemic *Hesperomannia* (Asteraceae). *Proceedings of the National Academy of Sciences, USA*, **95**, 15440-15445.
Kim, M., McCormick, S., Timmermans, M. & Sinha, N. (2003) The expression domain of PHANTASTICA determines leaflet placement in compound leaves. *Nature*, **424**, 438-443.
Kislev, M.E., Hartmann, A. & Bar-Yosef, O. (2006) Early domesticated Fig in the Jordan Valley. *Science*, **312**, 1372-1374.
Klak, C., Reeves, G. & Hedderson, T. (2004) Unmatched tempo of evolution in Southern African semi-desert ice plants. *Nature*, **427**, 63-65.
Knowles, L.L. (2000) Tests of Pleistocene speciation in montane grasshoppers (Genus *Melanoplus*) from the Sky Islands of western North America. *Evolution*, **54**, 1337-1348.
Knox, E.B. & Palmer, J.D. (1995) Chloroplast DNA variation and the recent radiation of the giant senecios (Asteraceae) on the tall mountains of eastern Africa. *Proceedings of the National Academy of Sciences, USA*, **92**, 10349-10353.
Kocher, T.D. (2003) Fractious phylogenies. *Nature*, **423**, 489-491.
Kocher, T.D. (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews: Genetics*, **5**, 288-298.
Kochert, G., Stalker, H.T., Gimenes, M., Galgaro, L., Romero Lopes, C. & Moore, K. (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). *American Journal of Botany*, **83**, 1282-1291.
Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T. & Yano, M. (2006) An

SNP caused loss of seed shattering during rice domestication. *Science*, **312**, 1392-1396.
Koopman, W.J.M., Zevenbergen, M.J. & Van den Berg, R.G. (2001) Species relationships in *Lactuca* s.l. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. *American Journal of Botany*, **88**, 1881-1887.
Koopman, W.J.M. (2005) Phylogenetic signal in AFLP data sets. *Systematic Biology*, **54**, 197-217.
Kozak, K.H., Weisrock, D.W. & Larson, A. (2006) Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). *Proceedings of the Royal Society B: Biological Sciences*, **273**, 539-546.
Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. & Janzen, D.H. (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences, USA*, **102**, 8369-8374.
Kroonenberg, S.B., Bakker, J.G.M. & van der Wiel, A.M. (1990) Late Cenozoic uplift and paleogeography of the Colombian Andes: constraints on the development of high-Andean biota. *Geologie en Mijnouw*, **69**, 279-290.
Kuittinen, H., Aguade, M., Charlesworth, D., Haan, A.D.E., Lauga, B., Mitchell-Olds, T., Oikarinen, S., Ramos-Onsins, S., Stranger, B., Van Tienderen, P. & Savolainen, O. (2002) Primers for 22 candidate genes for ecological adaptations in Brassicaceae. *Molecular Ecology Notes*, **2**, 258-262.
Kunst, L., Browse, J. & Somerville, C. (1988) Altered regulation of lipid biosynthesis in a mutant of *Arabidopsis* deficient in chloroplast glycerol-3-phosphate acyltransferase activity. *Proceedings of the National Academy of Sciences, USA*, **85**, 4143-4147.
Kurlovich, B.S., ed. (2002) *Lupins: Geography, Classification, GeneticResource and Breeding*. Intan, St Petersburg.
Kurlovich, B.S. & Repjev, S. (1994) Legume Systematic Research in the N.I. All-Russian Scientific Institute of Plant Industry (VIR). *Bean Bag*, **40**, 4-5.
Kurlovich, B.S. & Stankevich, A.K. (2002) Classification of Lupins. In *Lupins - Geography, Classification, Genetic Resources and Breeding.* (ed. B.S. Kurlovich), pp. 39-87. Itan, St. Petersburg.
Kurlovich, B.S., Heinänen, J., Kartuzova, L.T., Benken, I.I., Chmeleva, Z.V. & Bernatskaya, M.L. (2003) Diversity of lupin (*Lupinus* L.) based on biochemical composition. *Plant Genetic Resources Newsletter*, **134**, 42-57Ladizinsky, G. (1979) Seed dispersal in relation to the domestication of Middle East legumes. *Economic Botany*, **33**, 284-289.
Lambertini, C., Gustaffson, M.H.G., Frydenberg, J., Lissner, J., Speranza, M. & Brix, H. (2006) A phylogeographic study of the cosmopolitan genus *Phragmites* (Poaceae) based on AFLPs. *Plant Systematics and Evolution*, **258**, 161-182.
Lammers, T.G. (1990) Sequential paedomorphosis among the endemic Hawaiian Lobelioideae (Campanulaceae). *Taxon*, **39**, 206-211.
Lamoreux, J.F., Morrison, J.C., Ricketts, T.H., Olson, D.M., Dinerstein, E., McKnight, M.W. & Shugart, H.H. (2006) Global tests of biodiversity concordance and the importance of endemism. *Nature*, **440**, 212-214.
Lapage, S.P., Sneath, P.H.A., Lessel, E.F., Skerman, V.B.D. & Seeliger, H.P.R., eds.. (1976) *International Code of Nomenclature of Bacteria, and Statues of the International Committee on Systematic Bacteriology, and Statutes of the Bacteriology Section of the International Association of Microbiological Societies: Bacteriological*

Code. American Society for Microbiology, Washington.
Lara, M.C., Patton, J.L. & Da Silva, M.N.F. (1996) The simultaneous diversification of South American echimyid rodents (Hystricognathi) based on complete cytochrome b sequences. *Molecular Phylogenetics and Evolution*, **5**, 403-413.
Larson, G., Dobney, K., Alberella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conway, P., Andersson, L. & Cooper, A. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, **357**, 1618-1621.
Lavin, M.T., Pennington, R.T., Klitgaard, B.B., Sprent, J.I., Cavalcante de Lima, H. & Gasson, P., E. (2001a) The Dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *American Journal of Botany*, **88**, 503-533.
Lavin, M.T., Wojciechowski, M.F., Richman, A., Rotella, J., Sanderson, M.J. & Matos, A.B. (2001b) Identifying tertiary radiations of Fabaceae in the Greater Antilles: alternatives to cladistic vicariance analysis. *International Journal of Plant Science*, **162**, S53-S76.
Lavin, M.T., Wojciechowski, M.F., Gasson, P., Hughes, C.E. & Wheeler, E. (2003) Phylogeny of Robinoid legumes (Fabaceae) revisited: *Coursetia* and *Gliridiidia* recircumscribed, and a biogeographical appraisal of the Caribbean endemics. *Systematic Botany*, **28**, 387-409.
Lavin, M.T., Schrire, B., P., Lewis, G.P., Pennington, R.T., Delgado-Salinas, A., Thulin, M., Hughes, C.E. & Wojciechowski, M.F. (2004) Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 1509-1522.
Lavin, M.T., Herendeen, P.S. & Wojciechowski, M.F. (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification lineages during the Tertiary. *Systematic Biology*, **54**, 575-594.
Le Thierry d'Ennequin, M., Panaud, O., Toupance, B. & Sarr, A. (2000) Assessment of genetic relationships between *Setaria italica* and its wild relative *S. viridis* using AFLP markers. *Theoretical and Applied Genetics*, **100**, 1061-1066.
Lee, C., Kim, S.-C. & Santos-Guerra, A. (2004) The phylogeny of the woody *Sonchus* alliance (Asteraceae: Sonchinae) based on chloroplast and low-copy nuclear gene sequence. In Botany 2004, Vol. 1, Snowbird, Utah.
Lersten, N.R. (1981) Testa topography in Leguminosae subfamily Papilonoideae. *Proceedings of Iowa Academy of Science*, **88**, 180-191.
Lewis, G.P., Schrire, B., Mackinder, B. & Lock, M., eds.. (2005a) *Legumes of the World*, First edn. Royal Botanic Gardens, Kew.
Lewis, P.O., Holder, M.T. & Holsinger, K.E. (2005b) Polytomies and Bayesian phylogenetic inference. *Systematic Biology*, **54**, 241-253.
Li, C., Zhou, A. & Sang, T. (2006) Rice domestication by reduced shattering. *Science*, **311**, 1936-1939.
Li, P. & Johnston, M.O. (2000) Heterochrony in plant evolutionary studies through the twentieth century. *The Botanical Review*, **66**, 57-88.
Lihova, J., Shimizu, K.K. & Marhold, K. (2006) Allopolyploid origin of *Cardamine asarifolia* (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA sequences solved by a single-copy nuclear gene. *Molecular Phylogenetics and Evolution*, **39**, 759-786.
Linder, C.R. & Rieseberg, L.H. (2004) Reconstructing patterns of reticulate evolution in

plants. *American Journal of Botany*, **91**, 1700-1708.
Linder, H.P., Hardy, C.R. & Rutschmann, F. (2005) Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Molecular Phylogenetics and Evolution*, **35**, 569-582.
Linnaeus, C. (1771) *Genera plantarum, ex ed. 12a Systematis naturæ*, Edinburgh.
Liu, B. & Wendel, J.F. (2001) Intersimple sequence repeat (ISSR) polymorphisms as a genetic marker system in cotton. *Molecular Ecology Notes*, **1**, 205-208.
Liu, J.-Q., Gao, T.-G., Chen, Z.-D. & Lu, A.-M. (2002) Molecular phylogeny and biogeography of the Qinghai-Tibet Plateau endemic *Nannoglottis* (Asteraceae). *Molecular Phylogenetics and Evolution*, **23**, 307-325.
Liu, J.-Q., Wang, Y.-J., Wang, A.-L., Hideaki, O. & Abbott, R.J. (2006) Radiation and diversification within the *Ligularia-Cremanthodium-Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan Plateau. *Molecular Phylogenetics and Evolution*, **38**, 31-49.
Lockhart, P.J., McLanachan, P.A., Havell, D., Glenny, D., Huson, D. & Jensen, U. (2001) Phylogeny, radiation, and transoceanic dispersal of New Zealand alpine buttercups: molecular evidence under split decomposition. *Annals of Missouri Botanical Garden*, **88**, 459-477.
Londo, J.P., Chaing, Y.-C., Hung, K.-H., Chiang, T.-Y. & Schaal, B.A. (2006) Phylogeography of Asian wild rice, *Oryza rufipogon,* reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences, USA*, **103**, 9578-9583.
Loo, A.H.B., Dransfield, J., Chase, M.W. & Baker, W.J. (2006) Low-copy nuclear DNA, phylogeny and the evolution of dichogamy in the betel nut palms and their relatives (Arecinae; Arecaceae). *Molecular Phylogenetics and Evolution*, **39**, 598-618.
Losos, J.B., Jackman, T.R., Larson, A., de Queiroz, K. & Rodríguez-Schettino, L. (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, **279**, 2115-2118.
Luckow, M. & Hopkins, H.C.F. (1995) A cladistic analysis of *Parkia* (Leguminosae: Mimosoideae). *American Journal of Botany*, **82**, 1300-1320.
Lukens, L. & Doebley, J. (2001) Molecular evolution of the teosinte branched gene among maize and related grasses. *Molecular Biology and Evolution*, **18**, 627-638.
Lukens, L., Zou, F., Lydiate, D., Parkin, I. & Osborn, T. (2003) Comparison of a *Brassica oleraceae* genetic map with the genome of *Arabidopsis thaliana*. *Genetics*, **164**, 359-372.
Luteyn, J.L. (1999) *Páramos: A Checklist of Plant Diversity, Geographical Distribution, and Botanical Literature* Memoirs of the New York Botanical Garden 84, New York.
Lynch, M. & Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, **290**, 1151-1155.
Lynch, T.F. (1980) *Guitarrero Cave: Early Man in the Andes* Academic Press, New York.
Lynch, T.F. (1983) Camelid pastoralism and the emergence of Tiwananku civilization in the south-central Andes. *World Archaeology*, **15**, 1-14.
Lyons, L.A., Laughlin, T.F., Copeland, N.G., Jenkins, N.A., Womack, J.E. & O'Brien, S.J. (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics*, **15**, 47-56.
Mabberley, D.J. (1997) *The PlantBbook: A Portable Dictionary of the Vascular Plants, Utilizing Kubitzki's The Families and Genera of Vascular Plants (1990-), Cronquist's*

An Integrated System of Classification of Flowering Plants (1981) And Current Botanical Literature Arranged Largely On The Principles of Editions 1-6 (1896/97-1931) of Willis's A Dictionary of the Flowering Plants and Ferns, 2nd edn.
Maciel, H.S. & Schifino-Wittmann, M.T. (2002) First chromosome number determinations in south-eastern South American species of *Lupinus* L. (Leguminosae). *Botanical Journal of the Linnean Society*, **139**, 395-400.
Maddison, W.P. (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523-536.
Maisurian, N.A. & Atabekova, A.I. (1974) *Lupin*, Moscow.
Malcomber, S.T. (2002) Phylogeny of *Gaertnera*. Lam. (Rubiaceae) based on multiple DNA markers: Evidence of a rapid radiation in a widespread, morphologically diverse genus. *Evolution*, **56**, 42-57.
Marhold, K., Lihová, J., Perný, M. & Bleeker, W. (2004) Comparative ITS and AFLP analysis of diploid *Cardamine* (Brassicaceae) taxa from closely related polyploid complexes. *Annals of Botany*, **93**, 507-520.
Marshall, J.A., Knapp, S., Davey, M.R., Power, J.B., Cocking, E.C., Bennett, M.D. & Cox, A.V. (2001) Molecular systematics of *Solanum* section *Lycopersicum* (Lycopersicon) using the nuclear ITS rDNA region. *Theoretical and Applied Genetics*, **103**, 1216-1222.
Martin, A. & Burg, T.M. (2002) Perils of paralogy: using HSP70 genes for inferring organismal phylogenies. *Systematic Biology*, **51**, 570-587.
Mason-Gamer, R.J., Weil, C.F. & Kellogg, E.A. (1998) Granule-bound starch synthase: structure, function, and phylogenetic utility. *Molecular Biology and Evolution*, **15**, 1658-1673.
Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez G., J., Buckler, E. & Doebley, J. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences, USA*, **99**, 6080-6084.
Mayol, M. & Rosselló, J. (2001) Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution*, **19**, 167-176.
McCune, A.R. (1997) How fast is speciation? Molecular, geological, and phylogenetic evidence from adaptive radiations in fishes. In *Molecular Evolution and Adaptive Radiation* (eds. T.J. Givnish & K.J. Sytsma). CUP, Cambridge.
McKinnon, G.E., Vaillancourt, R.E., Steane, D.A. & Potts, B.M. (2004) The rare silver gum, *Eucalyptus cordata*, is leaving its trace in the organellar gene pool of *Eucalyptus globulus*. *Molecular Ecology*, **13**, 3751-3762.
McLenachan, P.A., Stöckler, K., Winkworth, R.C., McBreen, K., Zauner, S. & Lockhart, P.J. (2000) Markers derived from amplified fragment length polymorphism gels for plant ecology and evolution studies. *Molecular Ecology*, **9**, 1899-1903.
McMahon, M. & Hufford, L. (2005) Evolution and development in the amorphoid clade (Amorpheae: Papilionoideae: Leguminosae): petal loss and differentiation. *International Journal of Plant Sciences*, **166**, 383-396.
Megans, H.-J., van Moorsel, C.H.M., Piel, W.H., Pierce, N.E. & de Jong, R. (2004) Tempo of speciation in a butterfly genus from the Southeast Asian tropics, inferred from mitochondrial and nuclear DNA sequence data. *Molecular Phylogenetics and Evolution*, **31**, 1181-1196.
Melotto, M., Afanador, L. & Kelly, J.D. (1996) Development of a SCAR marker linked to the I gene in common bean. *Genome*, **39**, 1216-1219.
Menges, E.S. & Kohfeldt, N. (1995) Life history strategies of Florida scrub plants in relation to fire. *Bulletin of the Torrey Botanic Club*, **122**, 282-297.

Merino, E.F., Planchuelo, A.M. & Wink, M. (1999) Phylogenetic analysis of *Lupinus*. In *Lupin, an Ancient Crop for the New Millennium. Proceedings of the 9th International Lupin Conference*, Vol. 1, pp. 287-290, Klink/Muritz, Germany.
Merrill, E.K. (1979) Comparison of ontogeny of three types of leaf architecture in *Sorbus* L. (Rosaceae). *Botanical Gazette*, **140**, 328-337.
Merrill, E.K. (1986) Heteroblastic seedlings of green ash. II. Early development of simple and compound leaves. *Canadian Journal of Botany*, **64**, 2650-2661.
Meyer, A. (1993) Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in Ecology and Evolution*, **8**, 279-284.
Mickevich, M.F. & Farris, J. (1981) The implications of congruence in *Menidia*. *Systematic Zoology*, **30**, 351-370.
Milinkovitch, M.C. & Farris, J.S. (1998) Finding optimal ingroup topologies and convexitites when the choice of outgroups is not obvious. *Molecular Phylogenetics and Evolution*, **9**, 348-57.
Miller, R.E., Rausher, M.D. & Manos, P.S. (1999) Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and waxy sequence. *Systematic Botany*, **24**, 209-227.
Milne, R.I. (2004) Phylogeny and biogeography of *Rhododendron* subsection Pontica, a group with a tertiary relict distribution. *Molecular Phylogenetics and Evolution*, **33**, 389-401.
Minnis, P.E. (1981) Seeds in archaeological sites: sources and some interpretive problems. *American Antiquity*, **46**, 143-152.
Mittermeier, R.A., Myers, N., Thomsen, J.B., da Fonseca, G.A.B. & Oliveieri, S. (1998) Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. *Conservation Biology*, **12**, 516-520.
Miyamoto, M.M. & Fitch, W.M. (1995) Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology*, **44**, 64–76.
Mols, J.B., Gravendeel, B., Chatrou, L.W., Pirie, M.D., Bygrave, P.C., Chase, M.W. & Kessler, P.J.A. (2004) Identifying clades in Asian Annonaceae: Monophyletic genera in the polyphyletic Miliuseae. *American Journal of Botany*, **91**, 590-600.
Monteiro, R. & Gibbs, P.E. (1986a) A taxonomic revision of the unifoliolate species of *Lupinus* (Leguminosae) in Brasil. *Notes of the Royal Botanic Garden of Edinburgh*, **44**, 71-104.
Monteiro, R. (1986b) Observa aoes sobre a classifa ao tribal de *Lupinus* L. (Leguminosae, Papilionoideae). *Eugeniana*, **11**, 3-7.
Monteiro, R. (1987) Seed testa pattern of unifoliolate species of *Lupinus* L. (Leguminosae). *Salusvita*, **6**, 20-31.
Moore, G. (1998) A comparison of traditional and phylogenetic nomenclature. *Taxon*, **47**, 561-578.
Moore, G. (2003) Should taxon names be explicitly defined? *The Botanical Review*, **69**, 2-22.
Morrone, J.J. & Crisci, J.V. (1995) Historical biogeography: introduction to methods. *Annual Review of Ecology, Evolution, and Systematics*, **26**, 373-401.
Mort, M.E. & Crawford, D.J. (2004) The continuing search: low copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon*, **33**, 257-261.
Mort, M.E., Crawford, D.J. & Fairfield, K.N. (2004) Phylogeny and character evolution in California *Coreopsis* (Asteraceae): insights from morphology and from sequences of the nuclear and plastid genomes. *Systematic Botany*, **29**, 781-789.

Moseley, M.E. (1975) *Maritime Foundations of Andean civilizations.* Cummings.
Muellner, A.N., Tremetsberger, K., Stuessy, T.F. & Baeza, C.M. (2005) Pleistocene refugia and recolonization routes in the southern Andes: insights from *Hypochaeris palustris* (Asteraceae, Lactuceae). *Molecular Ecology*, **14**, 203-212.
Müller, K. (2005) SeqState: primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics*, **4**, 65-69.
Murata, N. & Tasaka, Y. (1997) Glycerol-3-phosphate acyltransferase in plants. *Biochemica et biophysica acta*, **1348**, 10-16.
Myers, N., Mittermeier, R.A., Mittermeier, C., G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotpots for conservation priorities. *Nature*, **403**, 853-858.
Naganowska, B. & Zielinska, A. (2002) Physical mapping of 18S-25S rDNA and 5S rDNA in *Lupinus* via fluorescent in situ hybridization. *Cellular and Molecular Biology Letters*, **7**, 665-670.
Naganowska, B., Wolko, B., Sliwinska, E. & Kaczmarek, Z. (2003) Nuclear DNA content variation and species relationships in the genus *Lupinus* (Fabaceae). *Annals of Botany*, **92**, 349-355.
Naganowska, B., Wolko, B., Sliwinska, E., Kaczmarek, Z. & Schifino-Wittmann, M.T. (2006) 2C DNA variation and relationships among New World species of the genus *Lupinus* (Fabaceae). *Plant Systematics and Evolution*, **256**, 147-157.
National Research Council, ed. (1989a) *Lost Crops of the Incas, Little-known Plants of the Andes with Promise for Worldwide Cultivation*. Report of an Ad Hoc Panel of the Advisory Committee on Technology Innovation Board on Science and Technology for International Development National Research Council.
National Research Council (1989b) Tarwi. In *Lost Crops of the Incas, Little-known Plants of the Andes with Promise for Worldwide Cultivation.*, pp. 181-189. Report of an Ad Hoc Panel of the Advisory Committee on Technology Innovation Board on Science and Technology for International Development National Research Council.
Nelson, G.J., Murphy, D.J. & Ladiges, P.Y. (2003) Brummitt on paraphyly: a response. *Taxon*, **52**, 295-298.
Nelson, G.J. & Platnick, N.I. (1981) *Systematics and biogeography: Cladistics and Vicariance* Columbia University Press, New York.
Nesbitt, T.C. & Tanksley, S.D. (2002) Comparative sequencing in the genus *Lycopersicon*: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics*, **162**, 635-379.
Nicholls, K.W. & Bohm, B.A. (1983) Flavonoids and affinities of some North American Lupines. *Canadian Journal of Botany*, **61**, 708-730.
Nichols, R. (2001) Gene trees and species trees are not the same. *Trends in Ecology and Evolution*, **16**, 358-364.
Nishida, I., Tasaka, Y., Shiraisi, H. & Murata, N. (1993) The gene and the RNA for the precursor to the plastid-located glycerol-3-phosphate acyltransferase of *Arabidopsis thaliana*. *Plant Molecular Biology*, **21**, 267-277.
Nixon, K.C. (1999) The parsimony ratchet, a new method for a rapid parsimony analysis. *Cladistics*, **15**, 407-414.
Nixon, K.C. (2002) WinClada. Nixon, K. C., Ithaca, NY, USA.
Nixon, K.C. & Wheeler, Q.D. (1990) An amplification of the phylogenetic species concept. *Cladistics*, **6**, 211-223.
Nixon, K.C. & Carpenter, J.J. (1993) On Outgroups. *Cladistics*, **9**, 413-426.

Nixon, K.C. & Carpenter, J.J. (1994) On outgroups. *Cladistics*, **15**, 413-426.
Nixon, K.C. & Carpenter, J.M. (1996) On simultaneous analysis. *Cladistics*, **12**, 221-241.
Nixon, K.C., Carpenter, J.M. & Stevenson, D.W. (2003) The PhyloCode is fatally flawed, and the "Linnaean" system can easily be fixed. *The Botanical Review*, **69**, 111-120.
Nowacki, E. & Jaworski, A. (1978) A study of the evolution of the genus *Lupinus* by serological methods. *Genetica Polonica*, **19**, 153-163.
Nylander, J.A.A. (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
Oh, S.-H. & Potter, D. (2003) Phylogenetic utility of the second intron of LEAFY in *Neillia* and *Stephanandra* (Rosaceae) and implications for the origin of *Stephanandra*. *Molecular Phylogenetics and Evolution*, **29**, 203-215.
Oh, S.-H. & Potter, D. (2005) Molecular phylogenetic systematics and biogeography of tribe Neillieae (Rosaceae) using DNA sequences of cpDNA, rDNA, and LEAFY. *American Journal of Botany*, **92**, 179-192.
Ohri, D. (1998) Genome size variation and plant systematics. *Annals of Botany*, **82 (Supplement A)**, 75-83.
Olmstead, R.G. (1989) Phylogeny, phenotypic evolution, and biogeography of the *Scutellaria angustifolia* complex (Lamiaceae): inference from morphological and molecular data. *Systematic Botany*, **14**, 320-338.
Olmstead, R.G. & Palmer, J.D. (1994) Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany*, **81**, 1205-1224.
Olsen, K.M. & Schaal, B.A. (1999) Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. *Proceedings of the National Academy of Sciences, USA*, **96**, 5586.
Orthia, L.A., Cook, L.G. & Crisp, M.D. (2005) Generic delimitation and phylogenetic uncertainty: an example from a group that has undergone an explosive radiation. *Australian Systematic Botany*, **18**, 41-47.
Owen, R. (1843) Lectures on the comparative anatomy and physiology of the invertebrate animals: delivered at the Royal college of surgeons, in 1843 / By Richard Owen ... From notes taken by William White Cooper and revised by Professor Owen.
Padolina, J., Timme, R., Linder, R., Briggs, W., Xu, W., Liu, W. & Miranker, D. (2004) Identification of broadly applicable nuclear DNA regions for phylogenetic reconstruction in angiosperms. *Botanical Society of America*, 528.
Palmer, J.D. (1992) Mitochondrial DNA in plant systematics: applications and limitations. In *Molecular Systematics of Plants.* (eds. P.S. Soltis, D.E. Soltis & J.J. Doyle), pp. 36-49. Chapman and Hall, New York.
Palmer, J.D., Soltis, D.E. & Chase, M.W. (2004) The Plant Tree of Life: an overview and some points of view. *American Journal of Botany*, **91**, 1437-1445.
Pamilo, P. & Nei, M. (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568-583.
Pardo, C., Cubas, P. & Tahiri, H. (2004) Molecular phylogeny and systematics of *Genista* (Leguminosae) and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (*trn*L-*trn*F intergenic spacer). *Plant Systematics and Evolution*, **244**, 93-119.
Paris, H.S., Yonash, N., Portnoy, V., Mozes-Daube, N., Tzuri, G. & Katzir, N. (2003) Assessment of genetic relationships in *Cucurbita pepo* (Cucurbitaceae) using DNA

markers. *Theoretical and Applied Genetics*, **106**, 971–978.
Parkin, D.T. (1953) The Durian theory - a criticism. *Phytomorphology*, **3**.
Pascual, H. (2004) *Lupinus marie-josephi* (Fabaceae), nueva y sorprendente especie descubierta en España. *Anales del Jardín Botanico de Madrid*, **61**, 69-72.
Patterson, C. (1982a) Cladistics and classification. *New Scientist*, **94**, 303-306.
Patterson, C. (1982b) Morphological Characters and Homology. In *Problems of Phylogenetic Reconstruction.* (eds. K.A. Joysey & A.E. Friday), Vol. 21, pp. 21-74. Academic Press, London.
Patterson, C. (1988) Homology in classical and molecular biology. *Molecular Biology and Evolution*, **5**, 603-625.
Patterson, C., Williams, D.M. & Humphries, C.J. (1993) Congruence between molecular and morphological phylogenies. *Annual Reviews in Ecology and Systematics*, **24**, 153-188.
Pearsall, D.M. (1980) Pachamachay ethnobotanical report: Plant utilization at a hunting base camp. In *Prehistoric Hunters of the High Andes.* (ed. J.W. Rick), pp. 191-231. Academic Press.
Pearsall, D.M. (1988) Interpreting the meaning of macroremain abundance: the impact of source and context. In *Current Paleoethnobotany: Analytical Methods and Cultural Interpretations of Archaeological Plant Remains* (eds. C.A. Hastorf & V.S. Popper), pp. 97-118. The Chicago University Press, Chicago and London.
Pearsall, D.M. (1992) The origins of plant cultivation in South America. In *The Origins of Agriculture: An International Perspective* (eds. C.W. Cowan, J. Watson & N.L. Benco), pp. 173-205. Smithsonian Institution Press, Washington D.C., London.
Pearsall, D.M. (2003) Plant food resources of the Ecuadorian formative: An overview and comparison to the central Andes. In *Archaeology of Formative Ecuador.* (eds. J.S. Raymond & R.L. Burger), pp. 213-257. Dumbarton Oaks, Washington D.C.
Pelser, P.B., Gravendeel, B. & van der Meijden, R. (2002) Tackling speciose genera: species composition and phylogenetic position of *Senecio* sect. Jacobaea (Asteraceae) based on plastid and nrDNA sequences. *American Journal of Botany*, **89**, 929-939.
Pelser, P.B., Gravendeel, B. & van der Meijden, R. (2003) Phylogeny reconstruction in the gap between too little and too much divergence: the closest relatives of *Senecio jacobaea* (Asteraceae) according to DNA sequences and AFLPs. *Molecular Phylogenetics and Evolution*, **29**, 613-628.
Penner, G.A., Bush, A., Wise, R., Kim, W., Domier, L., Kasha, K., Laroche, A., Scoles, G., Molnar, S.J. & Fedak, G. (1993) Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods and Applications*, **2**, 341-345.
Pennington, R.T. (1996) Molecular and morphological data provide phylogenetic resolution at different hierarchical levels in *Andira*. *Systematic Biology*, **45**, 496-515.
Pennington, R.T. & Dick, C.W. (2004) The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 1611-1622.
Perry, L., Sandweiss, D.H., Piperno, D.R., Rademaker, K., Malpass, M.A., Umire, A. & de la Vera, P. (2006) Early maize agriculture and interzonal interaction in southern Peru. *Nature*, **440**, 76-79.
Peters, J.L., McCraken, K.G., Zhuravlev, Y.N., Wilson, R.E., Johnson, K.P. & Omland,

K.E. (2005) Phylogenetics of wigeons and allies (Anatidae: Aves): the importance of sampling multiple loci and multiple individuals. *Molecular Phylogenetics and Evolution*, **35**, 209-224.
Petersen, G., Seberg, O., Yde, M. & Berthelsen, K. (2006) Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution*, **39**, 70-82.
Pfeil, B.E. & Crisp, M.D. (2005) What to do with *Hibiscus?* A proposed nomenclatural resolution for a large and well known genus of Malvaceae and comments on paraphyly. *Australian Systematic Botany*, **18**, 49-60.
Pfeil, B.E., Schlueter, J.A., Shoemaker, R.C. & Doyle, J.J. (2005) Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis on legumes using 39 gene families. *Systematic Biology*, **54**, 442-454.
Phillips, L.L. (1957) Chromosome numbers in *Lupinus*. *Madroño*, **14**, 30-36.
Piazza, P., Jasinski, S. & Tsiantis, M. (2005) Evolution of leaf developmental mechanisms. *New Phytologist*, **167**, 693-710.
Pimental, R. & Riggins, R. (1987) The nature of cladistic data. *Cladistics*, **6**, 201-209.
Pinheiro, M. & Miotto, S.T.S. (2001) Leguminosae-Fadoideae, Gènero *Lupinus* L. *Boletim do Instituto de Biociències UFRGS*, **60**, 1-100.
Piper, C.V. (1906) *Flora of the state of Washington*, Wash.
Piper, C.V. & Robinson, B.L. (1906) *Lupinus*. In *Flora of the state of Washington* (ed. C.V. Piper), pp. 350-358.
Piperno, D.R. & Pearsall, D.M. (1998) *The Origins of Agriculture in the Lowland Neotropics.* Academic Press, San Diego.
Planchuelo, A.M. (1984) Taxonomic studies of *Lupinus* in South America. In *Proceedings of the 3rd International Lupin Conference*, pp. 39-54, La Rochelle, France.
Planchuelo, A.M. (1996a) Biodiversity of lupins in South America. In Towards the 21st Century. *Proceedings of the 8th International Lupin Conference*. (ed. G. Hill, D.), pp. 394-399, Asilomar, California.
Planchuelo, A.M. (1996b) Relationship between South American and European species of *Lupinus*. In *Advances in Legume Systematics 8: Legumes of Economic Importance.* (eds. B. Pickersgill & J.M. Lock), pp. 109-116. Royal Botanic Gardens, Kew.
Planchuelo, A.M. & Dunn, B.D. (1984) The simple leaved lupines and their relatives in Argentina. *Annals of Missouri Botanical Garden*, **71**, 92-103.
Planchuelo, A.M. & Dunn, B.D. (1989) Two new lupin species of the *Lupinus lanatus* complex. *Annals of Missouri Botanical Garden*, **76**, 303-309.
Platnick, N.I. (1979) Philosophy and the transformation of cladistics. *Systematic Zoology*, **28**, 537-546.
Plitmann, U. (1981) Evolutionary history of the old world lupines. *Taxon*, **30**, 430-437.
Plitmann, U. & Heyn, C.C. (1984) Old World *Lupinus*: taxonomy, evolutionary relationships and links with New World species. In 3rd International Lupin Conference, pp. 55-66.
Plitmann, U. & Pazy, B. (1984) Cytogeographical distribution of the Old World *Lupinus*. *Webbia*, **38**, 531-539.
Poe, S. & Wiens, J.J. (2000) Character selection and the methodology of morphological phylogenetics. In *Phylogenetic Analysis of Morphological Data* (ed. J.J. Wiens), pp. 20-36. Smithsonian Institution Press, Washington DC.
Poke, F.S., Martin, D.P., Steane, D.A., Vaillancourt, R.E. & Reid, J.B. (2006) The impact

of intragenic recombination on phylogenetic reconstruction at the sectional level in *Eucalyptus* when using a single copy nuclear gene (cinnamoyl CoA reductase). *Molecular Phylogenetics and Evolution*, **39**, 160-170.
Polhill, R.M. (1976) Genisteae (Adans.) Benth. and Related Tribes (Leguminosae). *Botancial Systematics*, **1**, 143-368.
Polhill, R.M. (1982) *Crotalaria in Africa and Madagascar* AA Balkema, Rotterdam.
Polhill, R.M. (1994) Complete synopsis of legume genera. In *Phytochemical dictionary of the Leguminosae* (eds. F.A. Bisby, J. Buckingham, J.B. Harborne, I.W. Southon & J.L. Zarucchi), Vol. 1, pp. xlix-lvii. Chapman & Hall, London.
Posada, D. & Crandall, K.A. (2002) The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution*, **54**, 396-402.
Posada, D. & Buckley, T. (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**, 793-808.
Potter, D. & Freudenstein, J.V. (2005) Character-based phylogenetic Linnaean classification: taxa should be ranked and monophyletic. *Taxon*, **54**, 1033-1035.
Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. & Rafalski, A. (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, **2**, 225-238.
Prance, G.T. (1982) A review of the phytogeographic evidence for Pleistocene climate changes in the neotropics. *Annals of Missouri Botanical Garden*, **69**, 594-624.
Prance, G.T. & Sandved, K.B. (1985) *Leaves: The Formation, Characteristics And Uses of Hundreds of Leaves Found in all Parts of the World.* Thames and Hudson, London.
Rambaut, A. & Drummond, A. (2003-2005) Tracer, pp. MCMC Trace Analysis Tool. University of Oxford.
Rameau, C., Dénoue, D., Fraval, F., Haurogné, K., Josserand, J., Laucou, V., Batge, S. & Murfet, I.C. (1998) Genetic mapping in pea. 2. identification of RAPD and SCAR markers lined to genes affecting plant architecture. *Theoretical and Applied Genetics*, **97**, 916-928.
Raven, P.H. (1963) Amphitropical relationships in the floras of North and South America. *The Quarterly Review of Biology*, **38**, 151-177.
Ravines, R. (1972) Secuencia y cambios en los artifactos líticos del sur del Perú. *Revista del Museo Nacional*, **38**, 133-84.
Raymond, O., Piola, F. & Sanlaville-Boisson, C. (2002) Inferences of reticulation in outcrossing allopolyploid taxa: caveats, likelihood and perspectives. *Trends in Ecology and Evolution*, **17**, 3-6.
Ree, R.H., Citerne, H.L., Lavin, M.T. & Cronk, Q.B. (2004) Heterogeneous selection on LEGCYC paralogs in relation to flower morphology and the phylogeny of *Lupinus* (Leguminosae). *Molecular Biology and Evolution*, **21**, 321-331.
Ricardi, M. (1996a) Morfología de los cotiledones de plántulas de algunas familias o gèneros presentes en Venezuela come fuente de caracteres para su determinación. *Plantula*, **1**, 1-11.
Ricardi, M. (1996b) Morfologia de plátulas de bejucos de un bosque montano bajo. *Plantula*, **1**, 13-54.
Ricardi, M. (1999) Morfología de plántulas de árboles de bosques tropófilos de los estados Apure y Amazonas, Venezuela. *Plantula*, **2**, 103-118.
Richardson, J.E., Fay, M.F., Cronk, Q.B., Bowman, D. & Chase, M.W. (2000) A phylogenetic analysis of Rhamnaceae using *rbcL* and *trnL-F* plastid DNA

sequences. *American Journal of Botany*, **87**, 1309-1324.
Richardson, J.E., Pennington, R.T., Pennington, T.D. & Hollingsworth, P.M. (2001a) Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science*, **293**, 2242-2245.
Richardson, J.E., Weitz, F.M., Fay, M.F., Cronk, Q., Linder, H.P., Reeves, G. & Chase, M.W. (2001b) Rapid and recent origin of species richness in the Cape Flora of South Africa. *Nature*, **412**, 181-183.
Rick, J.W. (1980) *Prehistoric Hunters of the High Andes.* Academic Press, New York.
Rick, J.W. (1988) The Character and Context of Highland Preceramic Society. In *Peruvian Prehistory.* (ed. R.W. Keatinge), pp. 3-40. Cambridge University Press, Cambridge.
Rieppel, O. & Kearney, M. (2002) Similarity. *Biological Journal of the Linnean Society*, **75**, 59-82.
Rieseberg, L.H. (1996) The homology among RAPD fragments in interspecific comparisons. *Molecular Ecology*, **5**, 99-105.
Rieseberg, L.H. & Carney, S.E. (1998) Plant hybridization. *New Phytologist*, **140**, 599-624.
Riggins, R. & Sholars, T.A. (1993) *Lupinus* L. In *A manual of the flowering plants of California*, pp. 633-636. University of California Press, Berkeley.
Rokas, A., Krüger, D. & Carroll, S.B. (2005) Animal evolution and the molecular signature of radiations compressed in time. *Science*, **310**, 1933-1938.
Rokas, A., Williams, B.L., King, N. & Caroll, S.B. (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, **425**, 798-804.
Roman, B., Alfaro, C., Torres, A.M., Moreno, M.T., Satovic, Z., Pujadas, A. & Rubiales, D. (2003) Genetic relationships among *Orobanche* species as revealed by RAPD analysis. *Annals of Botany*, **91**, 637-642.
Ronquist, F. (1996) DIVA version 1.1. Computer program and manual available by anonymous FTP from Uppsala University (ftp:uu.se or ftp:systbot.uu.se).
Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.
Ronsted, N., Law, S., Thornton, H., Fay, M.F. & Chase, M.W. (2005) Molecular phylogenetic evidence for the monophyly of *Fritillaria* and *Lilium* (Liliaceae; Liliales) and the infrageneric classification of *Fritillaria*. *Molecular Phylogenetics and Evolution*, **35**, 509-527.
Rossetto, M., Jackes, B.R., Scott, K.D. & Henry, R.J. (2002a) Is the genus *Cissus* (Vitaceae) monophyletic? Evidence from plastid and nuclear ribosomal DNA. *Systematic Botany*, **27**, 522-533.
Rossetto, M., McNally, J. & Henry, R.J. (2002b) Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae. *Theoretical and Applied Genetics*, **104**, 61-66.
Rouhan, G., Dubuisson, J.-Y., Rakotondrainibe, F., Motley, T.J., Mickel, J.T., Labat, J.-N. & Moran, R.C. (2004) Molecular phylogeny of the fern genus *Elaphoglossum* (Elaphoglossaceae) based on chloroplast non-coding DNA sequences: contributions of species from the Indian Ocean area. *Molecular Phylogenetics and Evolution*, **33**, 745-763.
Rouy, G., Foucaud, J. & Camus, E.G. (1893) *Flore de France: ou, Description des plantes qui croissent spontanément en France, en Corse et en Alsace Lorraine* Société des Sciences naturelles de la Charente-Inférieure, La Rochelle.

Rutschmann, F. (2006) Molecular dating of phylogenetic trees: a brief review of current methods that estimate divergence times. *Diversity and Distributions*, **12**, 35-48.
Rydberg, P.A. (1917) *Flora of the Rocky Mountains and adjacent plains: Colorado, Utah, Wyoming, Idaho, Montana, Saskatchewan, Alberta, and neighboring parts of Nebraska, South Dakota, North Dakota, and British Columbia*, 1st edn.
Saint-Martin, M. (1978) Observations séminologiques au microscope électronique á balayage de diverses espécies de Papilionacées. *C. R. Academy Science, Paris, Series D*, **287**, 927-930.
Saint-Martin, M. (1986) Micromorphologie tégumentaire des graines de Papilionaceae. *Bulletin de la societé Botanique de France, lettres Botanique*, **133**, 137-153.
Sakai, M., Kanazawa, A., Fujii, A., Thseng, F.S., Abe, J. & Shimamoto, Y. (2003) Phylogenetic relationships of the chloroplast genomes in the genus *Glycine* inferred from four intergenic spacer regions. *Plant Systematics and Evolution*, **329**, 29-54.
Samuel, R., Bachmair, A., Jobst, J. & Ehrendorfer, F. (1998) ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American taxa of *Quercus* (Fagaceae). *Plant Systematics and Evolution*, **211**, 129-139.
Sanderson, M.J. (2002) r8s. Published privately, University California, Davis, California http://ginger.ucdavis.edu/r8s/.
Sanderson, M.J. & Wojciechowski, M.F. (1996) Diversification rates in a temperate legume clade: are there "so many species" of *Astragalus* (Fabaceae)? *American Journal of Botany*, **83**, 1488-1502.
Sanderson, M.J. & Shaffer, H.B. (2002) Troubleshooting molecular phylogenetic analyses. *Annual Review of Ecology and Systematics*, **33**, 49-72.
Sanderson, M.J., Thorne, J.L., Wikstrom, N. & Bremer, K. (2004) Molecular evidence on plant divergence times. *American Journal of Botany*, **91**, 1656-1665.
Sang, T. (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology*, **37**, 121-147.
Sang, T., Crawford, D.J. & Stuessy, T.F. (1995) Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proceedings of the National Academy of Sciences, USA*, **92**, 6813-6817.
Sang, T., Donoghue, M.J. & Zhang, D. (1997) Evolution of alcohol dehydrogenase genes in peonies (*Paeonia*): phylogenetic relationships of putative nonhybrid species. *Molecular Biology and Evolution*, **14**, 994-1007.
Sang, T. & Zhang, D. (1999) Reconstructing hybrid speciation using sequences of low copy nuclear genes: hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. *Systematic Botany*, **24**, 148-163.
Sanjur, O.I., Piperno, D.R., Andres, T.C. & Wessel-Beaver, L. (2002) Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: implications for crop plant evolution and areas of origin. *Proceedings of the National Academy of Sciences, USA*, **99**, 535-540.
Sattler, R. & Rutishauser, R. (1992) Partial homology of pinnate leaves and shoots: orientation of leaflet inception. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie*, **114**, 61-79.
Sauquet, H., Doyle, J.A., Scharaschkin, T., Borsch, T., Hilu, K.W., Chatrou, L.W. & Le Thomas, A. (2003) Phylogenetic analysis of Magnoliales and Myristicaeae based

on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society*, **142**, 125-186.

Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K. & Lane, R. (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1805-1811.

Sawickia-Sienkiewicz, E.J. & Brejdak, E. (1996) Interspecific crossibility of the Andean lupin (*Lupinus mutabilis* Sweet). In *Towards the 21st Century. Proceedings of the 8th International Lupin Conference*. (ed. G. Hill, D.), pp. 357-360, Asilomar, California.

Scherson, R.A., Choi, H.-K., Cook, D.R. & Sanderson, M.J. (2005) Phylogenetics of New World *Astragalus*: screening of novel nuclear loci for the reconstruction of phylogenies at low taxonomic levels. *Brittonia*, **57**, 354-366.

Schilling, E.E., Linder, C.R., Noyes, R.D. & Rieseberg, L.H. (1998) Phylogenetic relationships in *Helianthus* (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. *Systematic Botany*, **23**, 177-187.

Schluter, D. (2000) *The Ecology of Adaptive Radiation.* Oxford University Press, Oxford.

Schluter, D. (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372-380.

Schuh, R., T. (2003) The Linnean system and its 250-year persistence. *The Botanical Review*, **69**, 59-78.

Scotland, R.W., Olmstead, R.G. & Bennett, J.R. (2003) Phylogeny reconstruction: The role of morphology. *Systematic Biology*, **52**, 539-548.

Scotland, R.W. & Sanderson, M.J. (2004) The significance of few versus many in the tree of life. *Science*, **303**, 643.

Seehausen, O. (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the National Academy of Sciences, USA*, **273**, 1987-1998.

Selinger, D.A. & Chandler, V.L. (1999) Major recent and independent changes in levels and patterns of expression have occurred at the b gene, a regulatory locus in maize. *Proceedings of the National Academy of Sciences, USA*, **96**, 15007-15012.

Semerikov, V.L., Zhang, H., Sun, M. & Lascoux, M. (2003) Conflicting phylogenies of *Larix* (Pinaceae) based on cytoplasmic and nuclear DNA. *Molecular Phylogenetics and Evolution*, **27**, 173-184.

Senchina, D.S., Alvarez, I., Cronn, R.C., Liu, B., Rong, J., Noyes, R.D., Paterson, A.H., Wing, R.A., Wilkins, T.A. & Wendel, J.F. (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution*, **20**, 633-643.

Shaw, J., Cox, C.J. & Boles, S.B. (2003) Polarity of peatmoss (*Sphagnum*) evolution: who says bryophytes have no roots? *American Journal of Botany*, **90**, 1777-1787.

Shaw, J. & Small, R.L. (2005) Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). *American Journal of Botany*, **92**, 2011-2030.

Shaw, J., Cox, C.J. & Boles, S.B. (2005a) Phylogeny, species delimitation, and recombination in *Sphagnum* Section *Acutifolia*. *Systematic Botany*, **30**, 16-33.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E. & Small, R.L. (2005b) The tortoise and the hare II: relative utility of 21 non-coding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142-166.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114-1116.
Shoemaker, R.C., Schlueter, J. & Doyle, J.J. (2006) Paleopolyploidy and gene duplication in soybean and other legumes. *Current Opinion in Plant Biology*, **9**, 104-109.
Sholars, T.A. (1996) The systematics of perennial *Lupinus* in North America. In *Towards the 21st Century. Proceedings of the 8th International Lupin Conference*. (ed. G. Hill, D.), pp. 401-413, Asilomar, California.
Siddall, M.E. (1997) Prior agreement: arbitration or arbitrary? *Systematic Biology*, **46**, 765-769.
Sieburth, L.E. & Meyerowitz, E.M. (1997) Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. *Plant Cell*, **9**, 355-365.
Simmonds, N.W., ed. (1986) *Evolution of Crop Plants.*, 2nd edn. Vol. 1, pp 339. Longman Scientific and Technical.
Simmons, M.P. & Ochoterena, H. (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, **49**, 369-381.
Simmons, M.P., Pickett, K.M. & Miya, M. (2004) How meaningful are Bayesian support values? *Molecular Biology and Evolution*, **21**, 188-199.
Simpson, B.B. (1975) Pleistocene change in the flora of the high tropical Andes. *Paleobiology*, **1**, 273-294.
Simpson, B.B. & Neff, J.L. (1985) Plants, their pollinating bees, and the Great American Interchange. In *The Great American Biotic Interchange.* (eds. F.G. Stehli & S.D. Webb), pp. 427-453.
Simpson, B.B., Tate, J.A. & Weeks, A. (2004) Phylogeny and character evolution of *Hoffmannseggia* (Caesalpinieae: Caesalpinioideae: Leguminosae). *Systematic Botany*, **29**, 933-946.
Simpson, G.G. (1980) *Splendid Isolation: The Curious History of South American Mammals.* Yale University Press, New Haven.
Sinnott, E.W. & Bailey, I.W. (1915) Investigations on the phylogeny of angiosperms 5. Foliar evidence as to the ancestry and early climatic environment of the angiosperms. *American Journal of Botany*, **2**, 1-22.
Slade, R.W., Moritz, C. & Heideman, A. (1994) Multiple nuclear-gene phylogenies: applications to Pinnipeds and comparison with a mitochondrial DNA gene phylogeny. *Molecular Biology and Evolution*, **11**, 341-356.
Small, R.L. (2004) Phylogeny of *Hibiscus* sect. *Muenchhusia* (Malvaceae) based on chloroplast *rpl16* and *ndhF*, and nuclear ITS and GBSSI sequences. *Systematic Botany*, **29**, 385-392.
Small, R.L., Cronn, R.C. & Wendel, J.F. (2004) Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany*, **17**, 145-170.
Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T. & Wendel, J.F. (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany*, **85**, 1301-1315.
Smartt, J. (1980) Evolution and evolutionary problems in food legumes. *Economic Botany*, **34**, 219-235.
Smartt, J. (1990) *Grain Legumes: Evolution and Genetic Resources.* Cambridge University Press, Cambridge.

Smartt, J. & Simmonds, N.W. (1995) *Evolution of Crop Plants.*, 2nd edn. Longman Scientific & Technical, Harlow.
Smith, B.D. (1998) *The Emergence of Agriculture.* Freeman, New York.
Smith, B.D. (2001) Documenting plant domestication: the consilience of biological and archaeological approaches. *Proceedings of the National Academy of Sciences, USA*, **98**, 1324-1326.
Smith, C.P. (1918) Studies in the genus *Lupinus* - II. The Microcarpi, exclusive of *Lupinus densiflorus*. *Bulletin of the Torrey Botanic Club*, **45**, 1-22.
Smith, C.P. (1919) Studies in the genus *Lupinus* - IV. The Pusilli. *Bulletin of the Torrey Botanic Club*, **46**, 337-388.
Smith, C.P. (1938-1953) *Species Lupinorum.* Privately published, Saratoga, California.
Smith, C., Piper (1944a) Signature twenty-five, paper thirty-three concluded (*Lupinus* in Ecuador). In *Species Lupinorum.* (ed. C. Smith, Piper), pp. 385-390. privately published.
Smith, C.P. (1944b) 3 *Lupinus* L. In *Illustrated Flora of the Pacific States* (ed. L. Abrams), Vol. 2, pp. 485-519. Stanford University Press, Stanford, California.
Smith, C.P. (1946) *Lupinus* in North America - basic taxonomy. In *Species Lupinorum.*, pp. 545-576.
Smith, C.P. (1948) Paper sixty-two, Peru seven, Cuzco three. In *Species Lupinorum.* (ed. C. Smith, Piper), pp. 584-593. Smith, Charles. Piper.
Smith, C.P. (1951) 11 *Lupinus* L. Lupine. In *A Manual of the Flowering Plants of California.* (ed. W.L. Jepson), pp. 516-533, Berkeley, California.
Smith, C.P. (1953) Signature forty-four, paper one hundred and five. In *Species Lupinorum.*, pp. 753-768.
Smith, J.J., Scott-Craig, J.S., Leadbetter, J.R., Bush, G.L., Roberts, D.L. & Fulbright, D.W. (1994) Characterisation of random amplified polymorphic DNA (RAPD) products from *Xanthomonas campestris* and some comments on the use of RAPD products in phylogenetic analysis. *Molecular Phylogenetics and Evolution*, **3**, 135-145.
Smith, J.M.B. & Cleef, A.M. (1988) Composition and origins of the world's tropicalpine floras. *Journal of Biogeography*, **15**, 631-645.
Smith Jr., C.E. (1980) Plant Remains from Guitarrero Cave. In *Guitarrero Cave Early Man in the Andes.* (ed. T.F. Lynch), pp. 87-118. Academic Press, New York.
Smith, P.T., Kambhampati, S. & Armstrong, K.A. (2005) Phylogenetic relationships and character evolution among selected species of Bactrocera (Diptera: Tephritidae) based on multiple mitochondrial genes. *Insect systematics and evolution*, **36**, 343-359.
Smouse, P.E. (2000) Reticulation within the species boundary. *Journal of Classification*, **17**, 165-173.
Soltis, D.E., Tago-Nakazawa, M., Xiang, J.Q.-Y., Kawano, S., Murata, J., Wakabayashi, M. & Hibsch-Jetter, C. (2001) Phylogenetic relationships and evolution in *Chrysosplenium* (Saxifragaceae) based on matK sequence data. *American Journal of Botany*, **88**, 883-893.
Soltis, D.E., Soltis, P.S., Endress, P.K. & Chase, M.W. (2005) *Phylogeny and Evolution of Angiosperms.* Sinauer, Sunderland MA.
Sonnante, G., Galasso, I. & Pignone, D. (2003) ITS sequence analysis and phylogenetic inference in the genus *Lens* Mill. *Annals of Botany*, **91**, 49-54.
Sosef, M.S.M. (1997) Hierarchical models, reticulate evolution and the inevitability of

paraphyletic supraspecific taxa. *Taxon*, **46**, 75-85.
Spooner, D.M., Peralta, I.E. & Knapp, S. (2005) Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes [*Solanum* L. section Lycopersicon (Mill.) Wettst.]. *Taxon*, **54**, 43-61.
Sprent, J.I. (2001) *Nodulation in Legumes.* Royal Botanic Gardens, Kew.
Stace, C.A. (1980) *Plant Taxonomy and Biosystematics.* Edward Arnold, London.
Stafleu, F.A. & Cowan, R.S. (1976) *Taxonomic Literature: A Selective Guide to Botanical Publications and Collections with Dates, Commentaries and Types.*, 2nd edn. Bohn Scheltema & Holkema, Utrecht.
Standley, P.C. & Steyermark, J.A. (1946) Flora of Guatemala: Families Leguminosae, Geraniaceae, Oxalidaceae, Tropaeolaceae, Linaceae, Erythroxylaceae, Zygophyllaceae, Rutaceae, Simaroubaceae, Burseraceae, Meliaceae, and Malpighiaceae. *Fieldiana Botany*, **24**, 499.
Stehli, F.G. & Webb, S.D. (1985) *The Great American biotic interchange* Plenum, New York.
Stevens, P.F. (1991) Character states, morphological variation, and phylogenetic analysis: a review. *Systematic Botany*, **16**, 553-583.
Stevens, P.F. (2001 onwards) Angiosperm phylogeny website. Version 6, May 2005, pp. APG. www.mobot.org/MOBOT/Research/APweb/welcome.html.
Stevens, P.F. (2006) Book review: phylogeny and evolution of angiosperms. *International Journal of Plant Sciences*, **167**, 607-611.
Strand, A.E., Leebens-Mack, J. & Milligan, B.G. (1997) Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology*, **6**, 113-118.
Straub, S.C.K., Pfeil, B.E. & Doyle, J.J. (2006) Testing the polyploid past of soybean using a low-copy nuclear gene-Is *Glycine* (Fabaceae: Papilionoideae) an auto- or allopolyploid? *Molecular Phylogenetics and Evolution*, **39**, 580-584.
Sun, Y., Skinner, D.Z., Liang, G.H. & Hulbert, S.H. (1994) Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics*, **89**, 26-32.
Swiecicki, W., Wolko, B. & Jach, K. (1994) *Lupinus anatolicus* - an unknown species of the Old World. In *Advances in lupin research, Proceedings of the 7th International Lupin Conference*. (eds. J. Neves-Martins & M.L. Beirao de Costa), pp. 63-63. I. S. de Agronomia Press, Lisbon, Evora, Portugal.
Swofford, D.L. & Selander, R.B. (1981) BIOSYS-1: a FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *Journal of Heredity*, **72**.
Swofford, D.L. & Maddison, D.R. (1987) Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences*, **87**, 199-229.
Syring, J., Willyard, A., Cronn, R. & Liston, A. (2005) Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *American Journal of Botany*, **92**, 2086-2100.
Systematics Association committee for descriptive biological terminology (1962) II. Terminology of simple symmetrical plane shapes (Chart 1). *Taxon*, **11**, 145-156.
Sytsma, K.J. (1990) DNA and morphology: inference of plant phylogeny. *Trends in Ecology and Evolution*, **5**, 104-110.
Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105-1110.

Takhtajan, A.L. (1991) *Evolutionary trends in flowering plants* Columbia University Press, New York ; Oxford.
Tank, D.C. & Sang, T. (2001) Phylogenetic utility of the glycerol-3-phosphate acyltransferase gene: evolution and implications in *Paeonia* (Paeoniaceae). *Molecular Phylogenetics and Evolution*, **19**, 421-429.
Tanno, K.-I. & Willcox, G. (2006) How fast was wild wheat domesticated? *Science*, **311**, 1886.
Tapia, M. & Vargas, C. (1982) Wild lupines of the Andes of southern Peru. In *Proceedings of the 1st International Lupin Conference*, pp. 23-32, Lima/Cuzco, Peru.
Tapia, M., Riva, A. & Hernàndez-Bravo, G. (1988) Cool season food legumes in South and Central America. In *World Crops: Cool Season Food Legumes* (ed. R.J. Summerfield), pp. 1051-1057. Kluwer Academic Press, The Netherlands.
Tapia, M. & De la Torre, A. (1998) *Women Farmers and Andean Seeds*, 2nd edn. International Plant Genetic Resources Institute, Rome.
Tarrio, R., Rodriguez-Trelles, F. & Ayala, F.J. (2000) Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: the *Drosophila saltans* and *willistoni* groups, a case study. *Molecular Phylogenetics and Evolution*, **16**, 344-49.
Taubert, A. (1894) *Lupinus*. In *Die Natürlichen Pflanzenfamilien* (eds. H.G. Engler & K.A. Prantl), Vol. 3(3), pp. 231-232, Berlin.
Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. (2002) DNA points the way ahead in taxonomy. *Nature*, **418**, 479.
Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. (2003) A plea for DNA taxonomy. *Trends in Ecology & Evolution*, **18**, 70-74.
Thomas, J.H. (1974) New names etc. to be published in a single journal. *Taxon*, **23**, 833-835.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876-4882.
Thormann, C.E., Ferreira, M.E., Camargo, L.E.A., Tivang, J.G. & Osborn, T.C. (1994) Comparison of RFLP and RAPD markers to estimate genetic relationships within and among cruciferous species. *Theoretical and Applied Genetics*, **88**, 973-980.
Tomlinson, P.B. (1960) Seedling leaves in palms and their morphological significance. *Journal of the Arnold Arboretum*, **41**, 414-428.
Torres Tello, F. (1976) *Lupinus mutabilis* - a potent food source from the Andean regions. *The American Journal of Clinical Nutrition*, **29**, 933.
Towle, M.A. (1961) *The Ethnobotany of Pre-columbian Peru* Aldine Publishing Company, Chicago.
Tucker, S.C. & Hodges, S.A. (2005) Floral ontogeny of *Aquilegia*, *Semiaquilegia* and *Enemion* (Ranunculaceae). *International Journal of Plant Science*, **166**, 557-574.
Turner, B.L. (1995) A new species of *Lupinus* (Fabaceae) from Oaxaca, Mexico; a shrub or tree mostly three to eight meters high. *Phytologia*, **79**, 102-107.
Valenti, G.S., Melone, L., Ferro, M. & Bozzini, A. (1989) Comparative studies on testa structure of 'hard seeded' and 'soft seeded' varieties of *L. angustifolius* L. (Leguminosae) and on mechanism of water entry. *Seed Science and Technology*, **17**, 563-581.
Van den Berg, R.G., Miller, J.T., Ugarte, M.L., Kardolus, J.P., Villand, J., Nienhuis, J. & Spooner, D.M. (1998) Collapse of morphological species in the wild potato

Solanum brevicaule complex (Solanaceae: Sect. *Petota*). *American Journal of Botany*, **85**, 92-109.
Van der Hammen, T. & Cleef, A. (1986) Development of the High Andean Páramo Flora and Vegetation. In *High Altitude Tropical Biogeography.* (eds. F. Vuilleumier & M. Monasterio), pp. 153-201. Oxford University Press, New York.
van Welzen, P.C. (1998) Phylogenetic versus Linnaean taxonomy, the continuing story. *Taxon*, **47**, 413-423.
VandenBosch, K.A. & Stacey, G. (2003) Summaries of legume genomics projects from around the globe. Community resources for crops and models. *Plant Physiology*, **131**, 840-865.
Vargas, P., Baldwin, B. & Constance, L. (1998) Nuclear ribosomal DNA evidence for a western North American origin of Hawaiian and South American species of *Sanicula* (Apiaceae). *Proceedings of the National Academy of Sciences, USA*, **95**, 235-240.
Vaughn, P. & Dunn, D., B. (1977) The *Lupinus latifolius* Agardh complex. *Transactions, Missouri Academy of Science*, **10 + 11**, 90-106.
Vavilov, N.I. (1926) Centres of origin of cultivated plants. *Bulletin of applied botany and plant-breeding*, **16**, 139-248.
Vavilov, N.I. (1951) *The Origin, Variation, Immunity and Breeding of Cultivated Plants: Selected Writings.* [Chronica Botanica Co.], [Waltham, Mass.].
Vavilov, N.I. (1957) *World Resources of Cereals, Leguminous Seed Crops and Flax, and their Utilization in Plant Breeding.* Academy of Sciences of the USSR, Moscow.
Vavilov, N.I. & Dorofeev, V.F. (1992) *Origin and Geography of Cultivated Plants.* Cambridge University Press, Cambridge.
Veasey, E.A., Negrão, M.O., Martins, P.S. & Bandel, G. (1999) Early growth and seedling morphology of species of *Sesbania* Scop. (Leguminosae, Robinieae). *Scientia Agricola*, **56**.
Verheyen, E., Salzburger, W., Snoeks, J. & Meyer, A. (2003) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science*, **300**, 325-329.
Von Hagen, K.B. & Kadereit, J.W. (2001) The phylogeny of *Gentianella* (Gentianaceae) and its colonization of the southern hemisphere as revealed by nuclear and chloroplast DNA sequence variation. *Organisms, Diversity and Evolution*, **1**, 61-79.
Von Hagen, K.B. & Kadereit, J.W. (2002) Phylogeny and flower evolution of the Swertiinae (Gentianaceae-Gentianeae): homoplasy and the principle of variable proportions. *Systematic Botany*, **27**, 548-572.
von Humboldt, A. & A., B. (1807) *Ideen zueiner Geographie der pfhanzen*, Paris.
Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407-4414.
Vuilleumier, F. & Monasterio, M. (1986) Diversification and Adaptive Radiation. In *High Altitude Tropical Biogeography.* (eds. F. Vuilleumier & M. Monasterio), pp. 265-266.
Wainwright, C.M. (1977) Sun-tracking and related leaf movements in a desert lupine (*Lupinus arizonicus*). *American Journal of Botany*, **64**, 1032-1041.
Walsh, H.E. & Friesen, V.L. (2001) Power and stochasticity in the resolution of soft polytomies: a reply to Braun *et al. Evolution*, **55**, 1264-1266.
Walsh, H.E., Kidd, M.G., Moum, T. & Friesen, V.L. (1999) Polytomies and the power of

phylogenetic inference. *Evolution*, **53**, 932-937.
Walters, S.M. (1961) The shaping of angiosperm taxonomy. *New Phytologist*, **60**, 74-84.
Wang, R.-L., Stec, A., Hey, J., Lukens, L. & Doebley, J. (1999) The limits of selection during maize domestication. *Nature*, **398**, 236-239.
Wang, Y.G., Fritsch, P.W., Shi, S.H., Almeda, F., Cruz, B.C. & Kelly, L.M. (2004) Phylogeny and infrageneric classification of *Symplocos* (Symplocaceae) inferred from DNA sequence data. *American Journal of Botany*, **91**, 1901-1914.
Watson, S. (1873) Revision of the extra-tropical North American species of the genus *Lupinus*. *Proceedings of the American Academy of Arts and Science*, **8**, 517-548.
Watson, S. (1882) Contributions to American Botany. *Proceedings of the American Academy of Arts and Science*, **XVII**, 316-382.
Weber, S., Wolter, F.P., Buck, F., Frentzen, M. & Heinz, E. (1991) Purification and cDNA sequencing of an oleate-selective acyl-ACO:sn-glycerol-3-phosphate acyltransferase from pea chloroplasts. *Plant Molecular Biology*, **17**, 1067-1076.
Weeden, N.F., Wolko, B. & Vaillancourt, R. (1996) Contrasting patterns of partitioning genetic diversity in cultivated legumes. In *Advances in Legume Systematics 8: Legumes of Economic Importance.* (eds. B. Pickersgill & J.M. Lock), pp. 1-9. Royal Botanic Garden, Kew.
Weeden, N.F., Brauner, S. & Przyborowski, J.A. (2002) Genetic analysis of pod dehiscence in Pea (*Pisum sativum* L.). *Cellular and Molecular Biology Letters*, **7**, 657-663.
Weeks, A. & Simpson, B.B. (2004) Molecular genetic evidence for interspecific hybridization among endemic Hispaniolan *Bursera* (Burseraceae). *American Journal of Botany*, **91**, 976-984.
Weins, J.J. (1995) Polymorphic characters in phylogenetic systematics. *Systematic Biology*, **44**, 482-500.
Weins, J.J. (1998) Does adding characters with missing data increase or decrease phylogenetic accuracy. *Systematic Biology*, **47**, 625-640.
Weins, J.J. (2006) Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics*, **39**, 34-42.
Weins, J.J. & Reeder, T.W. (1995) Combining data sets with different numbers of taxa for phylogenetic analysis. *Systematic Biology*, **44**, 548-558.
Weins, J.J., Brandley, M.C. & Reeder, T.W. (2006) Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution*, **60**, 123-141.
Weising, K., Nybom, H., Wolff, K. & Meyer, W. (1995) *DNA Fingerprinting in Plants and Fungi* CRC Press, Boca Raton.
Weiss, E., Wetterstrom, W., Nadel, D. & Bar-Yosef, O. (2004) The broad spectrum revisited: evidence from plant remains. *Proceedings of the National Academy of Sciences, USA*, **101**, 9551-9555.
Wendel, J.F. (2000) Genome evolution in polyploids. *Plant Molecular Biology*, **42**, 225-249.
Wendel, J.F., Schnabel, A. & Seelanan, T. (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences, USA*, **92**, 280-284.
Westoby, M. & Swanborough, P. (1996) Seedling relative growth rate and its components in relation to seed size: phylogenetically independent contrasts. *Functional Ecology*, **10**, 176-184.

Weston, P.H. (2000) Process morphology from a cladistic perspective. In *Homology and Systematics.* (eds. R.W. Scotland & R.T. Pennington), pp. 124-144. Taylor and Francis, London and New York.
Wheeler, Q.D., ed. (in press) *Phylo-Code v. Linnaean Code - A Debate*.
White, T.J., Bruns, Y., Lee, S. & Taylor, J. (1990) Amplification and direct sequencing of fungal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications.* (eds. M. Innis, D. Gelfand, J. Sninsky & T. White), pp. 315-322. Academic Press, San Diego.
Whittall, J.B., Medina-Marino, A., Zimmer, E.A. & Hodges, S.A. (2006) Generating single-copy nuclear gene data for a recent adaptive radiation. *Molecular Phylogenetics and Evolution*, **39**, 124-134.
Wiley, E.O. (1981) *Phylogenetics: The Theory and Practice of Phylogenetic Systematics.* Wiley, New York ; Chichester.
Wilkin, P., Schols, P., Chase, M.W., Chayamarit, K., Furness, C.A., Huysmans, S., Rakotonasolo, F., Smets, E. & Thapyai, C. (2006) A plastid gene phylogeny of the yam genus, *Dioscorea*: roots, fruits and Madagascar. *Systematic Botany*, **30**, 736-749.
Williams, D.M., Ebach, M.C. & Wheeler, Q.D. (2005) 150 reasons for paraphyly: a response. *Taxon*, **54**, 858.
Williams, K.J., Kress, W.J. & Manos, P.S. (2004) The phylogeny, evolution, and classification of the genus *Globba* and tribe Globbeae (Zingiberaceae): appendages do matter. *American Journal of Botany*, **91**, 100-114.
Williams, P.A., Wilton, A. & Spencer, N. (2002) A proposed conservation weed risk assessment system for the New Zealand border. *Science for conservation*, **208**, 1-47.
Willis, J.C. & Yule, G.U. (1922) Some statistics of evolution and geographical distribution in plants and animals, and their significance. *Nature*, **109**, 177-179.
Wilson, C.A. & Calvin, C.L. (2006) An origin of aerial branch parasitism in the mistletoe family, Loranthaceae. *American Journal of Botany*, **93**, 787-796.
Wiltshire, R.J.E., Murfet, I.C. & Reid, J.B. (1994) The genetic control of heterochrony: evidence from developmental mutants of *Pisum sativum* L. *Journal of Evolutionary Biology*, **7**, 447-465.
Wing, E.S. (1986) Domestication of Andean mammals. In *High Altitude Tropical Biogeography.* (eds. F. Vuilleumier & M. Monasterio), pp. 246-264.
Wink, M., Merino, F. & Käss, E. (1999) Molecular evolution of Lupins (Leguminosae: *Lupinus*). In *Lupin, an Ancient Crop for the New Millennium. Proceedings of the 9th International Lupin Conference*, Vol. 1, pp. 278-286, Klink/Muritz, Germany.
Wink, M. & Mohamed, G.I.A. (2003) Evolution of chemical defence traits in the Leguminosae: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the rbcL gene. *Biochemical Systematics and Ecology*, **31**, 897-917.
Winkworth, R.C. & Donoghue, M.J. (2004) Viburnum phylogeny: evidence from the duplicated nuclear gene GBSSI. *Molecular Phylogenetics and Evolution*, **33**, 109-126.
Wojciechowski, M.F. (2005) *Astragalus* (Fabaceae): A molecular phylogenetic perspective. *Brittonia*, **57**, 382-396.
Wojciechowski, M.F., Sanderson, M.J., Baldwin, B.G. & Donoghue, M.J. (1993) Monophyly of aneuploid *Astragalus* (Fabaceae): evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany*, **80**, 711-

721.
Wojciechowski, M.F., Sanderson, M.J. & Hu, J.-M. (1999) Evidence of the monophyly of *Astragalus* (Fabaceae) and its major subgroups based on nuclear ribosomal DNA ITS and chloroplast DNA *trnL* intron data. *Systematic Botany*, **24**, 409-437.
Wojciechowski, M.F., Lavin, M.T. & Sanderson, M.J. (2004) A phylogeny of legumes (Leguminosae) based on analysis of the plastid *MATK* gene resolves many well-supported subclades within the family. *American Journal of Botany*, **91**, 1846-1862.
Wolfe, K.H., Li, W.-H. & Sharp, P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA*, **84**, 9054-9058.
Wolko, B. & Weeden, N.F. (1989) Estimation of *Lupinus* genome polyploidy on the bases of isozymic loci number. *Genetica Polonica*, **30**, 165-171.
Wolko, B. & Weeden, N.F. (1990a) Isozyme number as an indication of phylogeny in *Lupinus*. *Genetica Polonica*, **31**, 179-187.
Wolko, B. & Weeden, N.F. (1990b) Relationships among lupin species as reflected by isozyme phenotype. *Genetica Polonica*, **31**, 190-197.
Wortley, A.H., Rudall, P., Harris, D. & Scotland, R.W. (2005) How much data are needed to resolve a difficult phylogeny? Case study in the Lamiales. *Systematic Biology*, **54**, 697-709.
Xu, F. & Sun, M. (2001) Comparative analysis of phylogenetic relationships of grain Amaranths and their wild relatives (*Amaranthus*; Amaranthaceae) using internal transcribed spacer, amplified fragment length polymorphism, and double-primer fluorescent intersimple sequence repeat markers. *Molecular Phylogenetics and Evolution*, **21**, 372-387.
Xu, W., Briggs, W.J., Padolina, J., Timme, R., Liu, W., Linder, C.R. & Miranker, D.P. (2004) Using MoBIoS scalable genome join to find conserved primer pair candidates between two genomes. *Bioinformatics*, **20**, i355-i362.
Yacovleff, E. & Herrera, F.L. (1934) El mundo vegetal de los antiguos Peruanos. *Revista del Museo Nacional (Lima)*, **3**, 241-322.
Yang, H., Sweetingham, M.W., Cowling, W.A. & Smith, P.M.C. (2001) DNA fingerprinting based on microsatellite-anchored fragment length polymorphisms, and isolation of sequence-specific PRC markers in lupin (*Lupinus angustifolius* L.). *Molecular Breeding*, **7**, 203-209.
Yi, T., Miller, A.J. & Wen, J. (2004) Phylogenetic and biogeographic diversification of *Rhus* (Anacardiaceae) in the northern Hemisphere. *Molecular Phylogenetics and Evolution*, **33**, 861-879.
Yockteng, R. & Nadot, S. (2004a) Infrageneric phylogenies: a comparison of chloroplast-expressed glutamine synthetase, cytosol-expressed glutamine synthetase and cpDNA maturase K in *Passiflora*. *Molecular Phylogenetics and Evolution*, **31**, 397-402.
Yockteng, R. & Nadot, S. (2004b) Phylogenetic relationships among *Passiflora* species based on the glutamine synthetase nuclear gene expressed in chloroplast (ncpGS). *Molecular Phylogenetics and Evolution*, **31**, 379-396.
Young, K.R., Ulloa, C.U., Luteyn, J.L. & Knapp, S. (2002) Plant evolution and endemism in Andean South America: an introduction. *The Botanical Review*, **68**, 4-21.
Young, N.D., Mudge, J. & Ellis, T.N. (2003) Legume genomes: more than peas in a pod.

Current Opinion in Plant Biology, **6**, 199-204.
Zeder, M.A., Bradley, D.G., Emshwiller, E. & Smith, B.D. (2006) *Documenting Domestication, New Genetic and Archaeological Paradigms* University of California Press, Berkeley.
Zeven, A.C. & Brandenburg, W.A. (1986) Use of paintings from the 16th to 19th centuries to study the history of domesticated plants. *Economic Botany*, **40**, 397-408.
Zhang, L.-B., Comes, H.P. & Kadereit, J.W. (2001) Phylogeny and quaternary history of the European montane/alpine endemic *Soldanella* (Primulaceae) based on ITS and AFLP variation. *American Journal of Botany*, **88**, 2331-2345.
Zhang, L.-B., Simmons, M.P., Kocyan, A. & Renner, S.S. (2006) Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: implications for morphological and sexual system evolution. *Molecular Phylogenetics and Evolution*, **39**, 305-322.
Zhang, W., Wendel, J.F. & Clark, L.G. (1997) Bamboozled again: inadvertent isolation of fungal sequences from Bamboo (Poaceae: Bambusoideae). *Molecular Phylogenetics and Evolution*, **8**, 205-217.
Zhou, J. & Kleinhofs, A. (1996) Molecular evolution of nitrate reductase genes. *Journal of Molecular Evolution*, **42**, 432-442.
Zhou, S., Renner, S.S. & Wen, J. (2006) Molecular phylogeny and intra- and intercontinental biogeography of Calycanthaceae. *Molecular Phylogenetics and Evolution*, **39**, 1-15.
Zohary, D. (1989) Pulse domestication and cereal domestication: how different are they? *Economic Botany*, **43**, 31-34.
Zohary, D. & Hopf, M. (1973) Domestication of pulses in the Old World. *Science*, **182**, 887-894.
Zohary, D. & Hopf, M. (1994) Sources of evidence for the origin and spread of cultivated plants. In *Domestication of Plants in the Old World.* Oxford University Press, Oxford.

Abbreviations

AFLP	Amplified Fragment Length Polymorphism
BA	Bayesian Analysis
BLAST	Basic Local Alignment Search Tool
bp	Base Pairs
BP	Before Present
CATS	Comparative Anchor Tagged Sequences
CEH	Colin E. Hughes
CFI	Consensus Fork Index
CI	Consistency Index
COS	Conservered Ortholog Set
cpDNA	Chloroplast DNA
DNA	Deoxyribonucleic Acid
EST	Expressed Sequence Tag
ETS	External-Transcribed Spacer
GPAT	Glyceraldehyde Phosphate Acetyl Transferase
ILD	Incongruence Length Difference
IPNI	International Plant Names Index
ISSR	Inter-Simple Sequence Repeat
ITS	Internal-Transcribed Spacer
kb	Kilo Bases
L	Tree Length
LEGCYC	Legume Cycloidea
LCNG	Low Copy Nuclear Gene
PA	Parsimony Analysis
PCR	Polymerase Chain Reaction
PIC	Parsimony Informative Character
m	Metres
mm	Millimetres
mRNA	Messenger Ribonucleic Acid
mtDNA	Mitochrondrial DNA
Mya	Million Years Ago
Myr	Million Years
n	Number
N	Number
nDNA	Nuclear DNA
nrDNA	Nuclear Ribosomal DNA
PP	Posterior Probability
NPRS	Non-Parametric Rate Smoothing
RAPD	Randomly Aamplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RI	Retention Index
RJE	Ruth J. Eastwood
SCAR	Sequence Characterised Amplified Region
SLCNG	Single or Low Copy Nuclear Gene
SSR	Simple Sequence Repeat
QTL	Quantitative Trait Locus
TBR	Tree Bisection-Reconnection
UTR	Untranscribed Terminal Region

Appendix 1 Taxonomic account of *L. mutabilis* and allies

Introduction

The primary aim of this taxonomic study is to investigate the close relatives of the domesticated Andean lupin *L. mutabilis.* A morphological survey of all ca. 85 Andean lupin species based on herbarium specimens, field observations and greenhouse-grown material has provided the basis for delimitation of three species (*L. piurensis* C.P. Sm., *L. praestabilis* C. P. Sm. and *L. semperflorens* Hartw. ex Benth.) which show particular morphological similarities to *L. mutabilis* and an additional species (*L. ellsworthianus* C.P. Sm.) which shares certain characters with *L. piurensis* and *L. semperflorens*. *Lupinus praestabilis* has been suggested as a potential progenitor of the domesticate *L. mutabilis* (Tapia & Vargas, 1982). On the basis of this study *L. piurensis* is proposed as a more likely progenitor. Species accounts for these five species are presented.

Lupinus ellsworthianus, L. mutabilis, L. piurensis, L. praestabilis and *L. semperflorens* may form a natural group and there is some support for this from phylogenetic analysis of DNA sequences (Chapter 3). However, the only morphological character states supporting a close relationship between all five species are erect habit and glabrous adaxial leaflet surface. This combination of character states is also seen in a large number of other Andean species. Three unique character state combinations support relationships between overlapping subsets of this group and form the basis for their recognition as a group of putatively closely related species. *Lupinus ellsworthianus, L. mutabilis, L. piurensis,* and *L. semperflorens* share the character states of thin and oblanceolate leaflets while *L. mutabilis, L. piurensis, L. praestabilis* and *L. semperflorens* share glaucous stems and large flowers. *Lupinus mutabilis, L. piurensis, L. praestabilis* share the rare indumentum type of sparse short white appressed hairs. Stem and leaflet indumentum type and flower size distinguishes this group from the widespread and complex blue-flowered woody shrub alliance.

Although a phylogenetic species concept (Nixon & Wheeler, 1990) was followed, some species boundaries are inevitably somewhat arbitrary due to overlapping distributions of characters. The majority of morphological diversity found in *Lupinus* is quantitative and continuous and not readily amenable to objective partitioning into clearly delimited character states. Groups of specimens with multiple fixed characters can be identified, but, these often correspond to narrow geographical variants. Each such group shares characters with at least one other, and more commonly with several other geographically based morphological groups. Thus, although distinct entities can be identified, species boundaries are often unclear. To facilitate identification I have attempted to draw the boundaries around groups sharing the greatest number of fixed characters yet allowing for variation between populations. In some cases this approach has resulted in broadly circumscribed species. Varieties have been designated where a particularly distinct partition based on several minor fixed characters can be identified.

One of the major challenges faced during species delimitation is distinguishing environmental from genetic variation. This is particularly an issue in the Andes where steep altitudinal gradients can result in marked phenotypic differences in plant form over short distances. Field observations are particularly useful to observe variation with respect to the environmental conditions although trends can sometimes be observed from herbarium specimens when locality data is sufficiently detailed.

There may be more species which share characteristics with *L. mutabilis* which are as yet uncollected since extensive regions of the Andes remain poorly collected. Approximately 80 other species were dismissed as close relatives to *L. mutabilis* during this study. For example a large podded species *L. mirabilis*, from Colombia, was discarded since pod size was the only shared character. I have seen additional interesting material (herbarium sheets and photographs) which show relatively large flowered herbaceous plants from central and southern Peru. Unfortunately there is insufficient material to be certain of the status of this material. The specimens from Carpish (Huánuco), including the type specimen of *L. eanophyllus* C. P. Smith, (*Ochoa 1060*) would particularly merit further investigation.

Putative hybrid material has not been dealt with in this revision. This is pending further revision of the Andean species in order to confirm hybrid status.

Distribution and Habitat

It is noticeable that *L. piurensis* and *L. semperflorens* are similar and broadly overlapping in their distributions, predominately in northern Peru and southern Ecuador, but rarely grow together. *Lupinus ellsworthianus* and *L. praestabilis* are found in central southern Peru and just into northern Bolivia. However, their distributions are distinct and correspond to particular habitat types. *Lupinus ellsworthianus* is found in very humid transitional montane forest between dry high altitude jalca and wet tropical forest. *Lupinus praestabilis* is restricted to a narrow band of thorn scrub found on either side of the Andes. Habitat and altitude restrictions are consistent with diversification of Andean lupins due to ecological opportunity and environmental heterogeneity (Hughes & Eastwood, 2006).

Key characters

These five species can be distinguished from each other and other Andean species by unique combinations of characters. These are highlighted in the notes at the end of each species account and shown in Table A1. Characters of particular importance are stem indumentum, stipule size, leaflet thickness, shape and indumentum, flower size and shape, pod dehiscence and seed size.

Table A1 Distinguishing characters

Character	*L. mutabilis*	*L. piurensis*	*L. semperflorens*	*L. ellsworthianus*	*L. praestabilis*
Status	cultivated	wild	wild	wild	wild
Habit	annual/ biennial herb or small shrub	annual/short-lived perennial herb	perennial shrub or small tree	annual/short-lived perennial herb or shrublet	perennial shrub or small tree
Herbaceous/ woody	herbaceous, weakly woody at base	herbaceous, weakly woody at base	woody	herbaceous	woody
Stem	glabrous, glaucous	subsp. *piurensus* glabrous, glaucous, subsp. *cotumaza* sparsely pubescent	glaucous on very young plants - dense cover of white hairs	pubescent	glabrous, glaucous
Leaflet shape	oblanceolate, rarely narrow-linear or elliptic	oblanceolate, rarely narrow-oblanceolate, narrow-elliptic	oblanceolate, narrow-oblanceolate, narrow-oblong or narrow-elliptic	narrow oblanceolate, narrow-elliptical	narrow-elliptic to lorate
Abaxial leaflet indumentum	glabrous or rarely sparsely covered in short white appressed hairs	sparsely hairy or rarely glabrous or with long white appressed hairs	densely covered with long white silky hairs	glabrous or pubescent	sparsely covered in short white appressed hairs
Stipule length (mm)	3 - 20	2.4 - 4 (-6)	(7-) 12 - 22 (-32)	(3-) 5 - 10 (-12)	40 - 70
Stipule duration	often caducous	often caducous	persistent	persistent	persistent
Stipule indumentum	silky, often only on margins	glabrous	silky	pubescent	glabrous and papery with age
Leaf habit	leaflets held horizontal	leaflets held horizontal	leaflets often held drooping	leaflets held horizontal	leaflets held horizontal
Flower	large	large	large	small, 'button-like'	large
Floral bracts (mm)	4 - 10	3 - 5	10 - 14	8 - 9	15 - 20
Bracteoles (mm)	0.8 - 3.5	0.5 - 0.8	3.5 - 9	0.5 – 1	5 - 6
Pod length (mm)	52-120	30-53	40-80	(20-) 30-45	40-50
Pod width (mm)	6 - 9	3 - 6	3.5 - 5	3	3 - 4
Pod dehiscent	never	always	always	always	always
Seed length (mm)	(6-) 8-11	(4.6-) 5-5.6	4.5-7.5	4-5	4-5
Seed thickness (mm)	4 - 6	2.5 - 3.4	2 - 3.5	3	2 - 3
Seed colour	cream, white, black or grey or combination	marbled greys and browns	marbled greys and browns	marbled cream and browns	marbled greys and browns

Key to *L. mutabilis* and allies

Annual or perennial. Erect or rarely scrambling herbs or herbaceous shrubs or woody shrubs or small trees. (0.2-) 0.6 – 6 (-10) m in height. Glabrous and often glaucous or pubescent stems. Leaves palmately digitate 2 – 24 cm. Leaflets oblanceolate or rarely narrow-oblanceolate or narrow-linear or elliptic. Adaxial leaflet surface glabrous. Abaxial leaflet surface glabrous or sparsely covered in appressed white hairs or densely covered with long white silky hairs or pubescent. Erect terminal inflorescence held above foliage. Banner orbicular greater than 6 mm high and 5 mm wide.

1) Stipules 40-70 mm long. Stipules papery and sheathing the petiole and shoot. *L. praestabilis*
1) Stipules 2.5-32 mm long. Stipules otherwise. 2)

2) Stems glabrous and often glaucous or with very sparsely pubescent stems. Abaxial leaflet surface glabrous or sparsely covered in appressed white hairs. No noticeable difference between abaxial and adaxial leaf surfaces after drying. Stipules often caducous. Erect herb or barely woody small shrub. 3)
2) Stems with dense cover of silky white hairs or pubescent although stems of young plants maybe glabrous and glaucous. Abaxial leaflet surface densely covered with long white silky hairs or pubescent. Leaves often drying discolourous. Stipules persistent. Erect or sprawling herb, shrub or small tree. 4)

3) Pods indehiscent. Pods fleshy when immature, greater than 52 mm in length at maturity. Seeds white, grey or rarely black, (6-) 8-11 mm in length. *L. mutabilis*
3) Pods dehiscent. Pods non-fleshy when immature, mature length (25-) 30-52 mm. Seeds marbled brown and greys, (4.6-) 5-5.6 mm in length. *L. piurensis*

4) Perennial shrub or small treelet with one principal stem (0.8-) 1.5-6 (-6) m in height. Stipules (7-) 12 - 22 (-32) mm long, silky. Banner petal (13-) 15-18 mm high. *L. semperflorens*
4) Sprawling or erect annual or short-lived perennial herb or shrublet branching from base, (0.2-) 0.6-1.0 (1.8) m in height. Stipules (3-) 5 - 10 (-12) mm long, pubescent. Flowers often 'button-like'. Banner petal (6-) 9-11 (-14) mm high. *L. ellsworthianus*

Abbreviations

Departments of Bolivia

BE	Beni
CH	Chuquisaca
CO	Cochabamba
LP	La Paz
OR	Oruro
PA	Pando
PO	Potosí
SC	Santa Cruz
TA	Tarija

Departments of Colombia

AM	Amazonas
AN	Antioquia
AR	Araua
AT	Atlantico
DC	Distrito Capital de Bogota
BOL	Bolivar
BOY	Boyaca
CAL	Caldas
CAQ	Caqueta
CAS	Casanare
CAU	Cauca
CE	Cesar
CH	Choco
CO	Cordoba
CU	Cundinamarca
GN	Guainia
GV	Guaviare
HU	Huila
LG	La Guajira
MA	Magdalena
ME	Meta
NA	Naríno
NS	Norte de Santander
PU	Putamayo
QU	Quindio
RI	Risaralda
SAP	San Andres y Providencia
SAN	Santander
SU	Sucre
TO	Tolima
VC	Valle del Cauca
VA	Vaupes
VI	Vichada

Provinces of Ecuador

AZ	Azuay
BO	Bolívar
CAN	Cañar
CAR	Carchi
CH	Chimborazo
CO	Cotopaxi
EO	El Oro
ES	Esmeraldas
GA	Galápagos
GU	Guayas
IM	Imbabura
LO	Loja
LR	Los Ríos
MA	Manabí
MO	Morona-Santiago
NA	Napo
OR	Orellana
PA	Pastaza
PI	Pichincha
SU	Sucumbíos
TU	Tungurahua
ZC	Zamora-Chinchipe

Departments of Peru

AM	Amazonas
AN	Ancash
AP	Apurímac
AR	Arequipa
AY	Ayacucho
CAJ	Cajamarca
CU	Cusco
HAV	Huancavelica
HUC	Huánuco
IC	Ica
JU	Junín
LL	La Libertad
LA	Lambayeque
LI	Lima
LO	Loreto
MD	Madre de Dios
MO	Moquegua
PA	Pasco
PI	Piura
PU	Puno
SM	San Martín
TA	Tacna
TU	Tumbes
UC	Ucayali

States of Venezuela

AM	Amazonas
AN	Anzoátegui
AP	Apure
AR	Aragua
BA	Barinas
BO	Bolívar
CA	Carabobo
CO	Cojedes
DE	Delta Amacuro
DF	Distrito Federal
FA	Falcón
GU	Guárico
LA	Lara
ME	Mérida
MI	Miranda
MO	Monagas
NE	Nueva Esparta
PO	Portuguesa
SU	Sucre
TA	Táchira
TR	Trujillo
VA	Vargas
YA	Yaracuy
ZU	Zulia

1 Lupinus mutabilis Sweet, Brit. Fl. Gard. Ser. I. ii: t. 130 (1825). – Type: From Bogota, Mr. Barclay's Bury Hill, Jul 1826, *Smith 1201.15 1,* (Lectotype (here designated): LINN!).

Lupinus cruckshanksii Hook., Bot. Mag. t. 3056 (1832). *Lupinus mutabilis* var. *cruckshanksii* (Hook.) Sweet, Brit. Fl. Gard. Ser. 2. 3: t. 203 (1833). – Type: PERU. Pasco: Valley of Cainus, nr Pasco, Andes of Peru. *Cruckshanks s.n.* (Lectotype (here designated): K!, F-photograph!).

Lupinus mutabilis var. *roseus* C. P. Sm., Sp. Lupinorum 588 (1948). – Type: PERU. Cuzco: Pisac, Huancalle, Between Huancalle and Pisac. 13°27' S, 71°45' W alt. 3400 m, 7 May 1944, *Vargas C 4254* (Holotype: CAS!).

Annual or biennial herb or shrub with erect habit, plants with one principal stem or branching from near the base, (0.4-) 1-2 (-3) m in height commonly becoming top-heavy as plants mature. Stems hollow or pithy, sometimes weakly woody at base, glaucous, glabrous or with occasional white hairs to 0.5 mm. Stipules 3-20 mm long and (0.25-) 0.75-2 mm wide, adnate length 1-6.5 (-10) mm ~50% of length, adnate region linear, free tip lanceolate sometimes reflexed, often caducous, densely covered with silky hairs sometimes confined only to the margins. Leaves palmately digitate, (26-) 45-125 (-140) mm long with 5-9 (-10) leaflets per leaf; petioles (12-) 20-75 (-105) mm long and 0.5-2 mm in diameter, often longest towards base of plant, covered with sparse cover of short white appressed hairs; leaflets (14-) 20-55 (-80) mm long and (4-) 7-25 mm wide although leaflet length and width can vary by up to 20mm and 10mm respectively within one leaf, leaflet apex mucronate, blade oblanceolate rarely narrow-linear or elliptic, base attenuate, immature leaflets conduplicate, mature leaflets held flat, thin, adaxial surface glabrous or with occasional white hairs, abaxial surface glabrous or with sparse white appressed hairs to 1 mm. Erect terminal inflorescence (4.2-) 10.0 - >40.0 cm tall held above foliage, racemes markedly or loosely verticillate with (2-) 7-30 (->52) flowers in (1-) 2-6 (-9) whorls each with 1-8 flowers; peduncle diameter (1-) 2-6 (-10.5) mm, hollow, glabrous, or with sparse cover of short white appressed hairs; pedicels at anthesis (3-) 5-11 (-15) mm long, 0.5-2.5 (-5.5) mm in diameter with a sparse indumentum of white silky appressed hairs; floral bracts 4-10 mm long and 0.5-2 mm wide, lanceolate moderately-densely covered in white silky hairs, present in immature bud but rarely persisting at anthesis; bracteoles 0.8-3.5 mm long and 0.2-0.5 (-1) mm wide, lanceolate with apex attenuate, moderately-densely covered in white silky hairs; calyx upper lip 6-10 mm long and 2-3 mm high, bidentate the notch 0.5-1 mm, lower lip 7-12 (-14) mm long and 2-2.5 mm high, concave, green sometimes tinged inky blue, moderately-densely covered in silky white hairs; banner petal orbicular 11-20 mm high by 16-23 mm wide, the common colour form bright blue with a central white stripe and bright yellow spot which changes to rusty orange as the flower fades and the central stripe darkens to blue/purple, glabrous or with a sparse cover of short white silky hairs along back of banner in centre fold only; wing petals (10-) 15-24 mm wide and (6-) 9-15 mm high, lobe height 2-4 mm, claw length 2-4 mm, united, bright blue, glabrous with 7-9 rows of corrugated petal sculpturing at base near lobe; keel petals 10-11 mm long, 5-6 mm high, white/cream with blue apical region, margin of apex and basal half glabrous, sub-apical region ciliate; flowers often sweetly scented, all floral parts frequently persistent through to pod maturity, occasionally white, lilac, pale blue or pink-flowered individuals are seen. Pods angled upwards at 20-30° clasping stem, on stout, often hollow, peduncles, (1-) 5-8 (-18) pods per infructscence, 52-120 mm long, 7-20 mm wide and 6-9 mm thick, laterally compressed, narrow-oblong, apex mucronate occasionally with persistent style,

base cuneate asymmetrical towards ventral suture, dorsal suture undulate, ventral suture entire, 1 mm thickening surrounding sutures, lateral constrictions visible, venation often visible and prominent on valves near ventral suture, immature pods fleshy, mid-green, often tinged inky blue at sutures, mature pods straw yellow to dark brown, immature pods densely covered in white/grey pubescence, mature pods sparsely hairy or almost glabrous, (1-) 3-7 (-9) seeds per pod, indehiscent. Seeds broadly oblong, oblong, elliptical or narrowly rhombic, (6-) 8-11 mm long, 6-9 mm wide and 4-6 mm thick, 0.14-0.39 g weight, light grey venation visible on paler seeds, testa smooth to touch, white/cream, brown, light grey, black or a combination of these. Cotyledons petiolar, fleshy. First seedling leaves with 4-7 leaflets. Chromosome number: 2n=48 (unpublished data, Maria Schifino-Wittmann) (Darlington & Wylie, 1955). (Figs. A1.1, A1.5-A1.11).

Phenology. The main flowering and fruiting period is April to July but individuals of this species may be found in flower throughout most of the year.

Distribution. *Lupinus mutabilis* is found almost exclusively in cultivation and rarely on waste ground and roadsides in the mountains and valleys of the Andes from Venezuela (State: ME) through Colombia (Departments: CU, SAN,) Ecuador (Provinces: CAR, CH, CO, IM, LO, NA, PI), Peru (Departments: AM, AN, AR, AY, CAJ, CU, HUC, JU, LL, LA, LI, PA, PI, PU), Bolivia (Departments: CH, CO, LP, PO), northern Chile and north western Argentina (Fig. A1.2). The altitudinal range in which *L. mutabilis* is cultivated in South America is (1000-) 2000-4100 m although it has been grown successfully at lower altitudes elsewhere. This species has been widely introduced to Australia, Germany, New Zealand, Poland and the United Kingdom for agricultural and ornamental uses.

Vernacular names. chocho (Colombia, Ecuador, N. Peru), chochito, chuchus, chuchus multi (Bolivia), tarwi, kita tarwi, taurie or tarhui (Central Peru), tauri (Aymara language S. Peru, Bolivia), talhuie, ccequela (Azangaro), kirku, ullus (S. Peru, Bolivia), ullush (Tarma), altramuz, lupino, Andean lupin, changeable lupin, changeable-coloured lupin, pearl lupin, Mr Crunkshanks' Andean lupin.

Ethnobotany. *Lupinus mutabilis* is widely cultivated in the Andes for its seed that is harvested as a pulse crop for human consumption. It is the only lupin domesticated in the New World, although there have been reports of a North American species, *L. douglasii*, being consumed by North American Indians but not brought into cultivation (Kazimierski & Nowacki, 1961). *Lupinus mutabilis* is grown in gardens for home consumption and on farms in larger quantities for market sale (Fig A1.10). The average yield is 300-600 kg/ha but under optimal conditions harvests can yield 3500 kg/ha (Mujica, 1992). The seeds have a high protein content (ca. 40% Chávez & Untied, 1982) and are consumed in soups and salads and ground to make flour. Occasionally the leaves and tender pods are consumed (Gross, 1982). The seeds however contain high alkaloid concentrations (0.72-2.13%) and must be rigorously debittered before use (Blanco Galdes, 1982). Sweet individuals, with alkaloid concentrations of 0.15% or lower are naturally occurring and have also been artificially bred (von Baer & von Baer, 1988).

There is wider agricultural interest in *L. mutabilis* particularly from lupin breeders in Western Australia. One problem which hinders breeding programs is the high outcrossing rate. Although this varies depending on the position of the flower on the plant it can reach

81% in this species (10th International Lupin Conference, 1999). In comparison with Old World domesticated species of lupin, *L. mutabilis* produces lower yields per hectare (Report of an Ad Hoc Panel of the Advisory Committee on Technology Innovation Board on Science and Technology for International Development National Research Council 1989). Most *L. mutabilis* plants are not completely frost hardy and a crop can be destroyed when temperatures drop below -2°C (Tapia & Vargas, 1982). This crop is also susceptible to minor attack from a number of insect species and fungal infections (Frey, 1999).

Lupinus mutabilis is offered as a horticultural plant by major seed companies and is thus infrequently grown as an ornamental in Europe, America, Australia and New Zealand due to it's attractive appearance and rapid growth. Other recorded uses for this species are as a religious food, fertility treatment, medicine (against rheumatism, malaria and internal parasites), stupefacient and insectide (annotation on herbarium specimen Killip EP & Smith AC 22014 1929) (Antúnez de Mayolo, 1982 and references therein; Gross, 1982).

Additional specimens examined. **BOLIVIA, Chuquisaca,** Oropeza 12 km NW of Sucre, rd towards Ravelo 18°59'02" S, 65°20'41" W alt. 3290 m, 21 May 2002, *Hughes CE 2309* (FHO, K); **Cochabamba,** Liriuni, above Cochabamba 17°19' S, 66°20' W alt. 2560 m, Feb 1948, *Cardenas M 4201* (GH, US); Arani, 25 km SE of Arani, rd to Mizque, about 7km beyond the village of Juntatuyo 17°36'44" S, 65°39'58" W alt. 3550 m, 19 May 2002, *Hughes CE 2302* (FHO, K); Ayopaya, 22km ENE of Independencia, rd to la Mina 17°00'50" S, 66°47'26" W alt. 3330 m, 14 May 2002, *Hughes CE 2279* (FHO, K); Ayopaya, 22km ENE of Independencia, rd to la Mina 17°00'50" S, 66°47'26" W alt. 3330 m, 14 May 2002, *Hughes CE 2280* (FHO, K); Quillacollo, 13km N of Quillacollo, rd to Morochata, S side of the Cordillera Tunari 17°17'24" S, 66°18'05" W alt. 3130 m, 18 May 2002, *Hughes CE 2291* (FHO, K); **La Paz,** Region of Miraflores, 8 May 1931, *Buchtien O 8758* (CAS, NY, US); La Paz 16°30' S, 68°09' W, 24 Mar 1929, *Holway EWD 459* (US); Inquisivi, Quime 16°58' S, 67°13' W alt. 3000 m, 16 Mar 1988, *Nee M 36648* (NY); Larecaja, Quiabaya, 55km from Sorata towards Consata 15°39' S, 68°45' W alt. 3170 m, 9 Sep 1989, *Hermann M 287* (NY); Larecaja, 6km beyond the pass on the rd from Achacachi to Sorata, 2 km before the turnoff to Millipaya 15°51'29" S, 68°38'24" W alt. 4060 m, 9 May 2002, *Hughes CE 2253* (FHO, K); Larecaja, Vicinity of Sorata 15°50' S, 68°38'16" W alt. 3200 m, Apr 1858, *Mandon G 681a* (G, GH); Manco Kapac, Outskirts of Copacabana 16°08'39" S, 69°04'36" W, 19 Feb 1983, *Bohs L 2043* (F, GH); Manco Kapac 3 km N of Copacabana, track along shores of Lago Titicaca towards Chani and Challal 16°08'39" S, 69°04'36" W alt. 3845 m, 25 May 2002, *Hughes CE 2323* (FHO, K); Manco Kapac, Challapampa, nr the NE tip of Isla del Sol, Lago Titicaca 16°00' S, 69°10'56" W alt. 3840 m, 26 May 2002, *Hughes CE 2326* (FHO, K); Manco Kapac, 3.5km SW of Janckoamaya 16°12' S, 68°49' W alt. 3900 m, 6 Mar 1984, *Solomon JC 11643* (G, NY); Murillo, 200 m NE of rd along Rio Khallapa, Chinchaya 15°54' S, 68°50' W alt. 3540 m, 9 May 1987, *Vargas E 290* (FHO); **Potosí** Chaqui, nr Potosí 19°59' S, 65°35' W alt. 3500 m, 3 Feb 1939, *Balls EK 6144* (BM, E, K, UC, US); Saavedra, Km 86 rd from Sucre to Potosí, ca. 26 km SSW of Millares 19°35'26" S, 65°18'09" W alt. 3290 m, 22 May 2002, *Hughes CE 2317* (FHO, K);

CHILE, 1 Jul 1873, *Buysman M 60* (E);

COLOMBIA, Antioquia, vicinity of Medellin 6°45' N, 75°40' W, 1 Feb 1928, *Toro RA 963* (NY); **Cundinamarca**, Nr Sibate 5°27' N, 74°15' W alt. 3000 m, 12 Oct 1920, *Popenoe W 1150* (US); Bogota, Ciudad Universitaria, Sabana de Bogota 5°20' N, 74°05' W alt. 2600 m, 8 Sep 1940, *Cuatrecasas J 13638* (F, GH, US); Bogota, Ciudad Universitaria, Bogota 5°20' N, 74°05' W alt. 2620 m, 20 Mar 1946, *Duque-Jaramillo JM 3052* (NY); Bogota, Botanic Garden, Sabana de Bogota 5°20' N, 74°05' W alt. 2600 m, 12 Oct 1943, *Garcia-Barriga H 10902* (US); Bogota, Páramo de Bogota 5°17' N, 73°55' W alt. 3300 m, 1851, *Triana J* 4230 (NY); **Santander,** La Mesita, Pamplona 7°25' N, 72°45' W alt. 2725 m, 23 Feb 1939, *Alston AHG 7231* (BM, F, GH, NY, US); S slope, Páramo de Romeral, 30 Jan 1927, *Killip EP & Smith AC 18503* (GH, NY); S slopes Páramo de Romeral, 30 Jan 1927, *Killip EP & Smith AC 18709* (GH, NY);

ECUADOR, Carchi, Nr Angel 0°36' N, 77°57' W alt. 3250 m, 6 Jul 1935, *Mexia Y 7447*. (UC, US); **Chimborazo,** Pungalá, 6 km N on the rd from Chambo 1°47'96" S, 78°34'40" W alt. 3000 m, 14 Jul 2004,

Eastwood RJ 197 (FHO, QCNE); 6 km S of the turning to Palmira on the rd from Alausi to Cajabamba 2°07'12" S, 78°44'59" W alt. 3300 m, 25 Jul 2004, *Eastwood RJ 223* (FHO, QCNE); **Cotopaxi,** 3km E of Latacunga 0°55' S, 78°34' W alt. 2550 m, 24 Oct 1981, *Sperling CR 5296* (BM, E, F, GH, K, NY, QCNE); **Imbabura,** Rd Otavalo to Laguna Mojanda 0°12' N, 78°17' W alt. 3050 m, 29 Mar 1987, *Balslev H 62458* (K, NY, QCA, QCNE); Ayora, 8 km E on rd to Olmedo 0°07' N, 78°05' W alt. 3300 m, 7 Jul 2004, *Eastwood RJ 166* (FHO, QCNE); **Loja**, Cerro Sozoranga, km 2.6 rd from Colaisaca to Cariamanga 4°18'52" S, 79°40'24" W alt. 2490 m, 24 Apr 1994, *Jorgensen PM 557* (FHO); Loja, Saraguro, Plaza, N of Loja 4°04' S, 79°48' W alt. 2700 m, 17 Jul 1977, *Hart J 851* (GH); Loja, Km 3 rd Saraguro to Loja 4°04' S, 79°48' W alt. 1000 m, 30 Apr 1996, *Klitgaard BB 161* (FHO, K, NY, QCNE); Vicinity of Tablón de Ona, 27 Sep 1918, *Rose JN 23066* (US); **Napo,** Tena, Km 68-72 rd from Salcedo to Tena, banks of Rio Mulatos, nr Rio Sapala, Llanganates National Park 0°59' S, 78°12' W alt. 2230 m, 15 Sep 1998, *Vargas H 2424* (FHO, QCNE); **Pichincha,** Slopes of Volcán Pichincha overlooking Quito, 0°10' S, 78°34' W alt. 3400 m, 11 Jul 1959, *Barclay HG 7840* (US); Vicinity of Tambillo 0°24' S, 78°33' W alt. 2800 m, 26 Jun 1939, *Penland & Summers 961* (F, GH); Quito, Universidad Católica, 0°12' S, 78°30' W alt. 2850 m, 22 Apr 1985, *Arregui T 2* (NY); Quito, 7 km E of Pifo, rd to Papallacta 0°15' S, 78°17' W alt. 2950 m, 2 Jun 1997, *Neill D 10715* (NY, QCNE);

PERU, *Herrera 901*, (B); Uspachaca, 23 Jun 1922, *MacBride JF 1300* (F, US); 4 Jul 1922 (US); **Ancash,** Km 97 rd from Casma to Huaraz, W slopes of the Cordillera Negra, below the village of Pira 9°33'20" S, 77°43'03" W alt. 3120 m, 17 Jun 2001, *Hughes CE 2039* (FHO, K, MOL, NY); Huaraz. *Ames VM s.n.*, 16 Jun 1973 (USM); Huaraz Tumba, Yungay, 5 Feb 1974, *Ames VM s.n.* (USM); Recuay, 2km S of Recuay, rd from Huaraz to Catac, upper reaches of the Rio Santa valley 9°44'09" S, 77°27'08" W alt. 3430 m, 29 Apr 2002, *Hughes CE 2237* (FHO, K, MOL); **Arequipa,** Morro Verde, 18km above Yura by rail 16°07' S, 71°22' W alt. 2920 m, 13 Nov 1963, *Straw RM 2360* (US); **Ayacucho**, Carrapa, between Huanta and Río Apurímac 13°52' S, 74°20' W alt. 2800 m, 5 May 1929, *Killip EP & Smith AC 22281* (NY, US); **Cajamarca,** Las Arenas 4°55' S, 78°16' W alt. 3700 m, 8 Apr 1952, *Cabrera D 1821* (US); Surroundings of Cajamarca 7°10' S, 78°30' W alt. 3200 m, 18 Feb 1986, *Becker B 389* (NY); Km 25-30, rd from Cajamarca to Hualgayoc 7°05'53" S, 78°33'58" W alt. 3350 m, 20 Mar 1998, *Dostert N 98/128* (F, MSB); Cajamarca, Huayapongo on rd from Baños del Inca to Llacanora, about 2 km NW of Llacanora, 7°10'44" S, 78°26'32" W alt. 2610 m, 8 Jun 2001, *Hughes CE 1999*, (CAS, FHO, K, MOL, NY); 8.5 km N of Cajamarca rd to Yanacocha and Chota 7°05'53" S, 78°33'58" W alt. 2920 m, 10 Jun 2001, *Hughes CE 2009* (FHO, K, MOL, NY); Cajamarca, Vicinity of Namora 7°13' S, 78°19' W alt. 2600 m, 16 Aug 1973, *Sagastegui A 7743* (F, NY); In rolling hills 20 km SW of Celendín, rd to Cajamarca 6°56'06" S, 78°11'02" W alt. 3150 m, 23 Apr 2002, *Hughes CE 2220* (FHO, K, MOL); Celendín, Challuayacu (Kumulca-Sendamal) 6°58' S, 78°13' W alt. 3250 m, 17 Aug 1984, *Sagastegui A 12062* (NY); Sendamal, rd to Celendín 6°58' S, 78°13' W alt. 3100 m, 19 Aug 1984, *Sagastegui A 12238* (F, NY); Chota, Above Cochabamba 6°07' S, 78°51' W alt. 2300 m, 13 Nov 1986, *Mostacero J 1588* (NY); Contumaza, Pampa de la Sal (Cascabamba Pozo Kuan) 7°21' S, 78°48' W alt. 3500 m, 13 Jun 1981, *Sagastegui A 10035* (NY); Contumaza, Vicinity of Guzmango 7°12' S, 78°53' W alt. 2300 m, 29 Apr 1982, *Sagastegui A 10346* (NY); 3-4km NE of San Pablo 7°06'57" S, 78°48'37" W alt. 2590 m, 20 Apr 2002, *Hughes CE 2195* (FHO, K, MOL); Santa Cruz, 4.3 km (as the crow flies) NE of Monteseco 6°37' S, 78°57' W alt. 2600 m, 16 May 1987, *Santisteban J 69* (F, NY); **Cuzco,** 21-23 km NE of Paucartambo, or km 98-100 rd to Patria 13°05' S, 71°12' W alt. 3270 m, 9 Dec 1978, *Luteyn JL 6433* (NY); Calca, Centre of town of Amparaes around the bridge and SW from Cuzco and S of the town 13°10' S, 71°50' W alt. 3360 m, 13 Dec 1986, *Nuñez P 6679* (FHO, USM); Calca, Above Pisac 13°27' S, 71°45' W alt. 3700 m, May 1914, *Weberbauer A 6980* (F, G, GH, US, USM); Cuzco, Granja Experimental K'ayqa, 13°34' S, 7°57'10" W, 11 Dec 1972, *Brunel GR 155* (FHO); Ollantaytambo. 13°15' S, 72°15' W alt. 3000 m, 16 May 1913, *Cook OF 767* (US); Urubamba, Cordillera Salcantay 13°20' S, 71°58' W alt. 4000 m, 29 May 1954, *Hirsch R P1437* (NY); Urumbamba, Chincheros. Community of Pukamarka, fields between house of Jose Cusihuaman and Lake Piuray 13°25' S, 72°20' W alt. 3800 m, 4 Feb 1982, *King S 151* (F, GH, USM); **Huánuco,** Carpish Pass 9°45' S, 76°08' W, 22 Aug 1946, *Ferreyra R 1201* (UC, US, USM); Huánuco 9°55' S, 76°42' W alt. 2100 m, *MacBride JF 1365* (US); Dos de Mayo 5 km NNE of Pachas, km 20 rd from NE from La Union towards Huánuco, valley of the Rio Vizcarra 9°41'55" S, 76°45'25" W alt. 3450 m, 30 Apr 2002, *Hughes CE 2244* (FHO, K, MOL); Huánuco 9°53' S, 76°26' W alt. 3000 m, Apr 1868, *Pearce R 116* (K); **Junín,** Concepción Nueva Esperanza. 6 km E on Rd from Apata to Comas. E side of the Mantaro Valley 11°49'52" S, 75°19'02" W alt. 3680 m, 26 Jun 2004, *Eastwood RJ 151* (FHO, K, MOL, USM); Huancayo. 1922, *Chavez LA 10649* (USM); Nr Huancayo 12°04' S, 75°14' W alt. 3300 m, 28 May 1977, *Hart J 567* (GH); San Gerónimo, Mantaro Valley, 11°55' S, 75°16' W alt. 3300 m, 25 Apr 1929, *Killip*

EP & Smith AC 22014 (F, NY, US); Tarma. 1918, *Esposto N 10647* (USM); **La Libertad,** Rd from Otuzco to Usquil 7°50'44" S, 78°30'41" W alt. 3100 m, 6 Jun 2004, *Eastwood RJ 37* (FHO, K, MOL, USM); Santiago de Chuco, Rd Santiago de Chuco to Shorey 8°09'09" S, 78°13'27" W alt. 3610 m, 8 Jun 2004, *Eastwood RJ 56* (FHO, K, MOL); **Lambayeque**, Ferrenare Incahuasi 6°14' S, 79°24' W alt. 3500 m, 22 Jun 1986, *Quiroz SL 1947* (F, NY); **Lima,** Río Blanco 11°43' S, 76°16' W alt. 3250 m, 15 Apr 1929, *Killip EP & Smith AC 21564* (F, NY, US); **Pasco**, Valley of ?Cainus, nr Pasco, Andes of Peru, *Cruckshanks s.n.* (F, K); **Moquegua,** Comas, rd from Arequipa to Puquiña, in valley immediately before Puquiña 16°36'36" S, 71°11'48" W alt. 3100 m, 29 Apr 2000, *Weigend M 2000/581* (NY, USM); **Piura,** Vicinity of Huancabamba 5°23' S, 79°32' W, 22 Feb 1981, *Davis EW 755* (GH); **Puno,** Carabaya Ollachea, slopes of Koropina mountain 13°45' S, 70°30' W, 22 Aug 1980, *Boeke JD 3156* (NY); Chucuito Edge of Lake Titicaca, Yunguyo 15°53' S, 69°53' W alt. 3815 m, 18 Feb 1975, *Plowman T 5127* (F, GH); Lampa, Vicinity of Puno. 23 Jul 1965, *Tovar O 5113* (USM); Yunguyo, San Ramon. 14 May 1992, *APDAL s.n.* (USM);

VENEZUELA, Mérida, Rangel, Páramo de Mucuchies 8°45' N, 70°55' W alt. 4000 m, 29 Sep 1942, *Tamayo F 2413* (UC, US);

FRANCE, Cultivated in 'Jardin du Luxemburg' - originating from Bogota. 10 Jul 1834, *Gay J s.n.* (K);

SWITZERLAND, Geneva, Geneva Botanic Garden. 1851, *Barbey W s.n.* (G);

UK, Oxfordshire, Oxford Botanic Garden. 2 Sep 2003, *Eastwood RJ 11* (FHO); Oxford Botanic Garden. 2 Oct 2003, *Eastwood RJ 14* (FHO); Oxford Botanic Garden. 9 Oct 2003, *Eastwood RJ 18* (FHO).

Notes. *Domestication.* The Andes is one of the 5-9 regions where agriculture arose independently and is a hotspot for early crop plant and livestock domestication (Diamond, 2002). *Lupinus mutabilis* is one of a set of indigenous food plants domesticated in the Andes often referred to as "The Lost Crops of the Incas" (NRC, 1989). Wild lupin seeds are reported in various archaeological sites in South America, the oldest dating to 8000-6000 BC (Fig. 4.1). Domesticated lupin seed appears ca. 200 AD. Following this, putative seeds of *L. mutabilis* have been found in Nazca tombs and the species is thought to be depicted on Tiahuanacan pottery (700-800 AD) (Antúnez de Mayolo, 1982; Yacovleff & Herrera, 1934)(Fig. 4.3). The window of domestication is wide from 1500 BC to 200 AD but it is likely domestication occurred toward the beginning of this period coincident with a rapid expansion of the diversity of crops found in the archaeological record. In common with other Andean domesticates, cultivation of *L. mutabilis* has suffered a decline since the conquest of South America by the Spanish in the 17th century due in part to displacement by introduced crops such as broad bean, *Vicia faba*.

Lupinus mutabilis differs from wild Andean *Lupinus* species in a number of characters including;

- indehiscent pods,
- large seed size (producing large cotyledons and seedlings),
- a soft, water permeable testa,
- reduced pigmentation of the testa.

It is thought that these differences are the result of human selection during domestication. Some or all of these character changes have been documented for other domesticated legumes such as chickpea, cowpea, lentil and bean (Evans, 1993; Ladizinsky, 1979; Smartt, 1985). There are no reports of indehiscent pods from any wild Andean lupin species. A survey of seed size within *Lupinus* has shown that only seeds of Old World species reach comparable sizes to those of *L. mutabilis* (unpublished data). *Lupinus mutabilis* seeds are at least twice the size of the largest seeded Andean species and ca. six times those of the putative progenitor *L. piurensis.*

Relatives and origin. Multiple gene trees place *L. mutabilis* within a New World clade with high support (Chapter 3) (Ainouche & Bayer, 1999; Ainouche *et al.*, 2004; Kass & Wink, 1997). Currently well-supported resolution is largely lacking within this group due to the lack of molecular variation. This low molecular variation in combination with the high morphological diversity seen has led to the hypothesis that the Andean lupins are the result of a rapid radiation (Chapter 3). Thus, the domestication of *L. mutabilis* is nested within this wider radiation of Andean lupins and although there has been considerable morphological evolution due to both radiation and domestication this has not been accompanied by comparable levels of molecular change in the loci examined so far (Bateman, 1999).

Lupinus condensiflorus C.P.Sm. (Tapia & Vargas, 1982), *L. eanophyllus* C.P.Sm. (Smith, 1953), *L. falsomutabilis* C.P. Sm. (Smith, 1953) *L. haughtianus* C.P.Sm. (Smith, 1944), *L. malacotrichus* (Tapia & Vargas, 1982), *L. paniculatus* (Tapia & Vargas, 1982), *L. praestabilis* C.P.Sm. (Smith, 1948b; Tapia & Vargas, 1982), and *L. tayacajensis* C.P.Sm. (Smith, 1953) have all been proposed as close relatives of *L. mutabilis*. Detailed investigation of protologues and type material shows that *L. praestabilis* shows close morphological similarity to *L. mutabilis*. The holotype of *L. praestabilis* C.P.Sm. Vargas 5052 (CAS!) is annotated by Vargas as 'very similar to *L. mutabilis*'. However, *L. praestabilis* is markedly distinct from *L. mutabilis* in a number of characters, most notably in its large stipules. *L. condensiflorus*, *L. paniculatus* and *L. malacotrichus* belong with the widespread blue-flowered shrubby complex. *L. eanophyllus* is similar to but apparently more robust than *L. piurensis* and may be related to *L. falsomutabilis.* Further material is required to assess the status of this species. *Lupinus haughtianus* is conspecific with *L. jamesonianus* a sprawling woody perennial from Ecuador. Smith (1953) commented that *L. tayacajensis* C.P. Sm. is 'quite a wonderful diversion from the ordinary *L. mutabilis*'. However, while the holotype of *L. tayacajensis* (*Ochoa 974*; CAS!) is similar to *L. mutabilis* in its pithy stem and large leaves, it has large papery stipules, small flowers, a greater number of leaflets and a strigose indumentum over all vegetative parts.

A number of collections have been seen which are putative hybrids involving *L. mutabilis* (personal observation and Gross, 1982). Some of these have been recognized as distinct species. Although sharing leaflet shape characters with *L. mutabilis*, the type of *L. xenophytus* C.P.Sm. Fiebrig, 2414 (Holotype:CAS!) from Bolivia also shares many morphological characters (diamond shaped bracteoles, long thin stipules, dense silky abaxial leaf indumentum) with *L. bandelierae* C.P.Sm. Only one collection of this species has been made. Dunn's proposal (annotation on herbarium label, date) that this is a hybrid between *L. mutabilis* and *L. bandelierae* is possible given the morphology but divergent chromosome numbers mitigate against this hypothesis (Conterato & Schifino-Wittmann, 2006; unpublished data Schifino-Wittmann). Hughes 2329 (FHO) and 2293 (FHO) are putative hybrids of *L. mutabilis* with wild species.

There has been speculation about the origin of *L. mutabilis.* Kazimierski and Nowacki (1961) postulated an origin from one of two North American species, *L. douglasii* or *L. ornatus,* supported by sparsely sampled alkaloid data. This would be inconsistent with the patterns of domestication and movement of crops in the New World known so far and the placement of *L. mutabilis* in an exclusively Andean clade in current phylogenetic trees. The major centres for domestication in the New World were Mexico, the Amazon and the Andes. The only major crop thought to have been domesticated in North America

is the sunflower, *Helianthus annuus* L. (Harter et al., 2004). Furthermore, observations of herbarium material of *L. douglasii* and *L. ornatus* have shown these species are not highly similar in their morphology to *L. mutabilis.* A local origin for *L. mutabilis* is more consistent with the presence of several species that are morphologically very similar to *L. mutabilis* in the Andean region, allozyme data and with the close relationship to Andean species consistently proposed by molecular data (Ainouche & Bayer, 1999; Kass & Wink, 1997; Wolko & Weeden, 1990).

Putative wild *L. mutabilis* populations have been recorded from the department of Cuzco, Peru (Blanco Galdes, 1982). These are reported to be morphologically similar to *L. mutabilis* but are highly pigmented, with leaves, flowers, pods and seeds smaller than cultivated plants (Blanco Galdes, 1982). Tapia and Vargas (1982) named the wild species as *L. praestabilis*. Like all wild Andean lupins the pods of *L. praestabilis* are dehiscent and the seeds, which are impermeable to water, show the common marbled pattern (Blanco Galdes, 1982, 1984). One specimen of *L. praestabilis* (*Mandon 681*) is annotated as cultivated. However, this collection is mixed with specimens of *L. mutabilis* and I believe the cultivated status refers to *L. mutabilis* only. The Cuzco populations showed elevated levels of outcrossing when compared to domesticated *L. mutabilis*. Blanco Galdes (1984) crossed plants from these populations with cultivated *L. mutabilis.* Hybrids were readily formed and exhibited the small leaves, small flowers and dehiscent pods seen in the putative wild population yet had large white seeds, a character seen in the cultivar. However, hybridization within the genus occurs both naturally between sympatric species (field observation) and even artificially at a lower rate between distantly related species e.g. *L. mutabilis* and the Old World species *L. albus* (10th International Lupin Conference, 1999). The ability of the Cuzco populations to hybridize with *L. mutabilis* and retain domesticated characteristics is not thus strong evidence to confirm their status as wild *L. mutabilis*. Additionally *L. praestabilis* has very distinctive large stipules which very different from the small stipules of *L. mutabilis.*

Lupinus piurensis var. *piurensis* has been identified as a putative progenitor of *L. mutabilis* based on morphological similarity (Fig. 4.14). *Lupinus piurensis* C.P.Sm., an erect herbaceous annual or short-lived perennial, has small stipules and glabrous, glaucous stems. Populations are distributed from the Province of Loja in southern Ecuador to central Peru (Fig. A2.2). This distribution indicates a domestication event in southern Ecuador or northern/central Peru, regions known to have had sophisticated early civilizations.

Types and Nomenclature. The name *L. cruckshanksii* has been published twice. First by Hooker in 1831 and this name is treated here as a synonym of *L. mutabilis* Sweet. The name published later by A. Gray (1856) is illegitimate. IPNI lists *L. mutabilis* Lindley as a validly published name. Given that when Lindley used this name he acknowledged Sweet's prior publication and noted the source of the plant as from Robert Barclay in the Botanical Register (1832), I do not think this was an attempt to publish this name. Cook (1925) utilised *L. cunninghamii* in a figure legend of a photograph of *L. mutabilis.* It has been suggest that the name *L. cunninghamii* was confused with *L. cruckshanksii* (the determination of *Cook 767* and a synonym for *L. mutabilis* (Towle, 1961)). No further use or reference to this name has been found thus it is believed to be illegitimate. Several authors (e.g. Hanelt, 1986) list *L. tauris* as an additional synonym of *L. mutabilis* but this is clearly misplaced as *L. tauris* is quite a distinct species not readily confused with *L. mutabilis.* It is unclear where this came from and there does not appear to be any publication of *L. tauris* which relates to *L. mutabilis.*

Although a type specimen was not cited in the description of *L. cruckshanksii* Hook. the protologue gives the locality of the Andes of Peru, near Pasco. Two sheets from Hookers Herbarium at Kew are from this locality and are here designated lectotypes. When Sweet later described *L. mutabilis* var. *cruckshanksii* (Hooker) Sweet. he commented that this species "was discovered by Mr Cruckshanks, not far from Pasco, on the Peruvian Andes". The illustration accompanying this protologue was drawn from nursery material.

No types were designated in the protologue of *L. mutabilis* Sweet. However, the original material used in the protologue has been located in the Linnaean herbarium (Fig. A1.4). The LINN specimen was originally in the herbarium of E.D. Smith, the artist who drew the illustration accompanying the protologue of *L. mutabilis*, and is annotated "From Bogata, Mr Barclay's Bury Hill July 1 1826". Although it is dated one year later than the protologue, this specimen matches the source material for the illustration of *L. mutabilis* Sweet. (Fig. A1.3), alluded to in the text, and furthermore, the illustration and the specimen are a close match suggesting that the illustration was originally drawn from this material (compare Figs. A1.3-4) justifying its designation here as the lectotype of *L. mutabilis*. The plant described by Sweet was grown from seed from Bogota (mis-spelled as Bogato) by Robert Barclay at Bury Hill and drawn by E.D. Smith from a specimen presented to him by Barclay. Robert Barclay tended an exotic garden on his estate and was an enthusiastic importer of plants from across the world (Hooker, 1830-31). At this time there was huge influx of plants from abroad and also a large network of growers and gardeners in the United Kingdom and Europe. In fact the earliest dated specimens, after the type, (Gay s.n. 1831 and Gay s.n. 1834) of *L. mutabilis* Sweet were made from material cultivated in the "Jardin du Luxemburg" in Paris from seed also originating in "Bogato". Sweet and Barclay were in communication and Sweet visited the garden at Bury Hill in 1725 (Sweet & Smith, from 1823). Barclay himself did not make a herbarium but he did employ artists whose work contributed to botanical works such as Sweet's British Flower Garden and Botanical Register.

It is interesting to note that none of the early descriptions of *L. mutabilis* refer to or illustrate the conspicuously large indehiscent pods even though the ability to increase stocks of this plant by seeds in the UK was clearly stated (Hooker, 1830-31). However, these early descriptions are in horticultural publications and emphasize that *L. mutabilis* was introduced into Europe as a garden plant with attractive scented flowers during an intense period of plant collection and distribution. Mature pods are not common in herbarium collections and the earliest specimen showing mature pods is Mandon 681 (G, GH not BM or K) in 1860.

Two formal varieties of *L. mutabilis* var. *cruckshanksii* (Hooker) Sweet and var. *roseus* Vargas and six invalidly published varieties: var. *albiflorus* Atabekova, var. *albo-roseus* Atabekova, var. *elegantissimus* Libk, var. *indianus* Libk, var. *montanus* Libk, and var. *violaceus* Atabekova, have been proposed essentially on the basis of flower colour variation (Maisurian & Atabekova, 1974; Smith, 1948b; Sweet, 1835). The latter are judged to be invalidly published since no latin diagnosis or types are presented by Maisurian and Atabekova (1974) and no further references to these elements have been uncovered. However, given the book is published in Russian it is possible that these may have been overlooked. *Lupinus mutabilis* var. *cruckshanksii* (Hooker) Sweet describes the common bright blue-flowered form differing from the original description of *L. mutabilis* Sweet which is based on a pale blue and

white-flowered form. Flower colour variants are not formally recognised here because pale blue, lilac, pink and white-flowered individuals are often seen growing sporadically in otherwise blue-flowered populations and are thought to be no more than flower-colour mutants. Similar pale blue, lilac, pink or white-flowered individuals occur in populations of otherwise purple-blue flowered wild species such as *L. piurensis* and *L. tauris.* This is in line with diverse colour morphs that are also found as sporadic individuals in populations of some North American lupin species. According varietal status to such sporadic mutants seems unjustified. All flowers show colour changes as the flowers fade (Fig. A1.7) except for some pure white individuals which remain white.

Informally, many named cultivars and agricultural varieties of *L. mutabilis* have been recognised (Mujica, 1992). Major seed bank collections of these are held at Universidad Nacional San Antonio Abad del Cusco, Estación Experimental Baños del Inca, Universidad Nacional del Altiplano - Peru, Centro de Investigaciones Fitoecogenéticas de Pairumani -Bolivia, Instituto de Ciencias Naturales – Ecuador and The Institute of Plant Genetics, Poland. Tapia (1982) recognised a set of regional ecotypes of *L. mutabilis* two of which are particularly distinctive, a high valley ecotype (Figs. A1.5a-c) and an Altiplano ecotype (A1.5d) found at altitudes above 3,800m in the department of La Paz, Bolivia, particularly around Lake Titicaca. The basis of the distinction is length of vegetative period with the high valley ecotype exhibiting a growing period of 270-300 days while the Altiplano ecotype has a shorter period of 250-270 days.

Gross informally (1982) distinguished two 'subspecies' *tarhui* north of the 10º 5' S latitude line and *chocho* south of this line. *Tauri* is more foliose, matures more slowly and has narrower leaflets than *chocho* which also has flatter seeds. Gross also stated that *tauri* has greater resistance to anthracnose. The diversity of seed colour and patterning is reported to be greater in *chocho* (Gross, 1982).

Finally, Blanco (1984) recognised three informal infraspecific groups. One corresponds to the Altiplano ecotype in the altitude range and short vegetation cycle but differs in geographical location. These plants are generally small, less than 60 cm, with a single axis and little or no lateral branching. The environment faced by the Altiplano plants is that of low rainfall and low temperatures. The second group is similar to the high valley ecotype described by Tapia (1982), with plants 1.5 m tall with one main axis yet also with lateral branches. Thirdly, Blanco recognised a highly branched group distributed in northern Peru and Ecuador where the climate is milder and there is little risk of frost. This comprises plants to 1.8 m tall, with multiple axes and also those with the longest vegetative period of up to a year. Cowling *et al* (1998) reported morphological and phenotypic differences corresponding to latitude.

All the ecotypes relate essentially to different growth forms (Figs. A1.4 & A1.6). Environmental factors such as amount of rainfall and proximity to other plants can affect growth form particularly with regard to branching (Tapia, 1982). However, there does appear to be a general geographic trend of decreased branching and woodiness, reduced stipule and bract length and shorter vegetative period from North to South. However, a survey of *L. mutabilis* across its range has revealed the variation in habit is not consistent with any other morphological characters. While is may be useful in terms of agricultural utilisation to identify this level of phenotypic variation in terms of ecotypes such variants do not merit formal botanical recognition.

Distinguishing features. The key characters distinguishing *L. mutabilis* species from all other Andean species are the large indehiscent pods and large pale-coloured seeds (Figs. A1.8-9).

Excluded names.

Lupinus cunninghamii O. F. Cook, (1925) J. Hered. 16(2): 33-46, 16(3): 94-110 nom. This name was used by Cook to accompany a photograph of *L. mutabilis*. No protologue, type, or other references to this name have been located, and the name does not appear to have been validly published.

Lupinus mutabilis Lindley (1833) Bot. Reg. t. 1539. This was not an attempt to publish this name since Lindley acknowledged the earlier publication of *L. mutabilis* in the Botanical Register (1832).

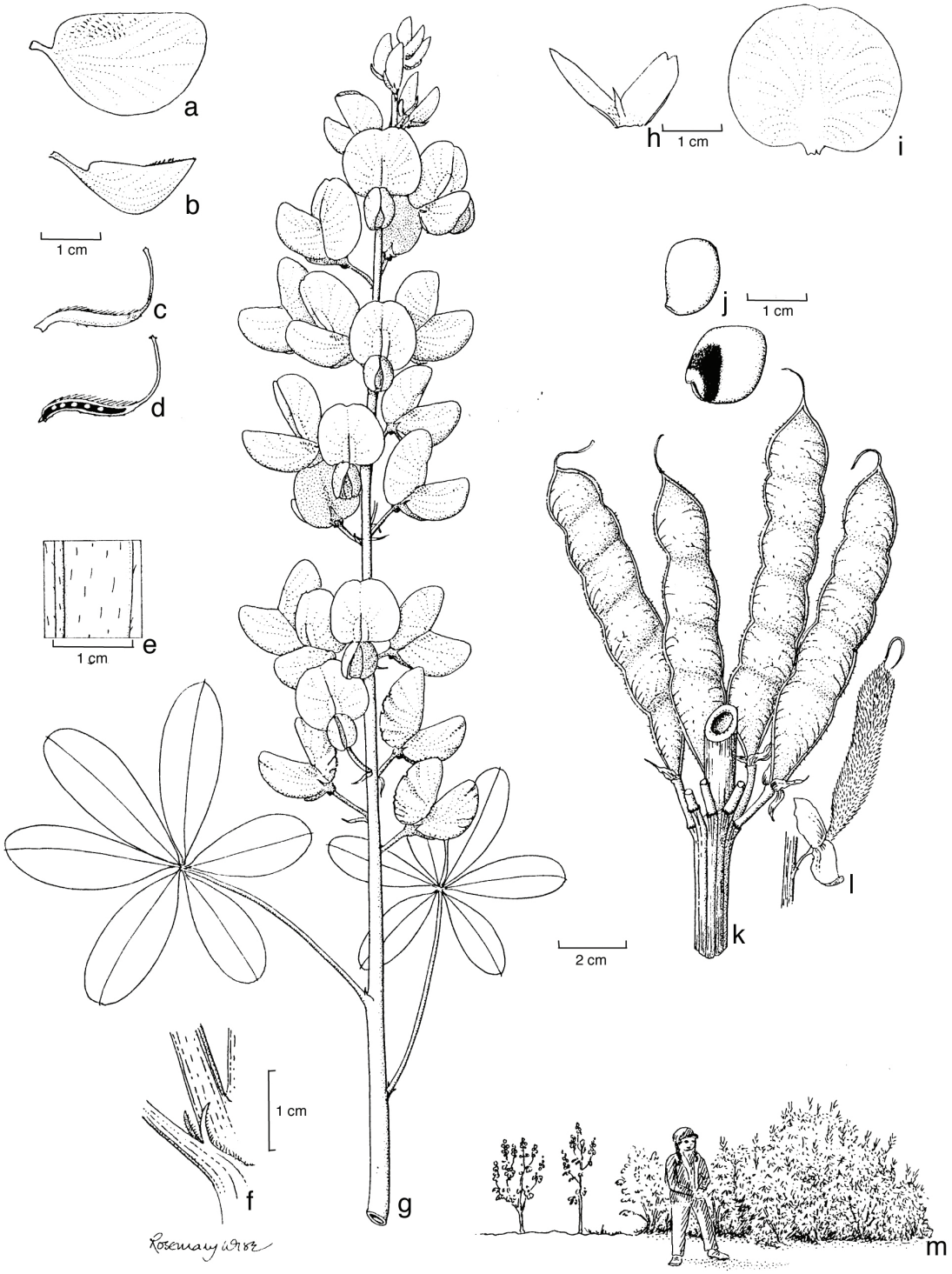


Fig. A1.1 *Lupinus mutabilis* Sweet
a. Wing petal, b. Keel petal, c. Gynoecium, d. Longitudinal section of gynoecium, e. Detail of abaxial leaflet surface, f. Stipule, g. Flowering shoot, h. Calyx, i. Banner petal, j. Seeds, k. Mature infructescence, l. Immature pod, m. Habit. Based on: a-d, f-i & l, *CEH2279*; e, *CEH2220*; j, University of Western Australia collection *P28725* & *P26961*; k, *CEH2302*; m, *RJE37* & *CEHs 2326*. Illustration by Rosemary Wise.

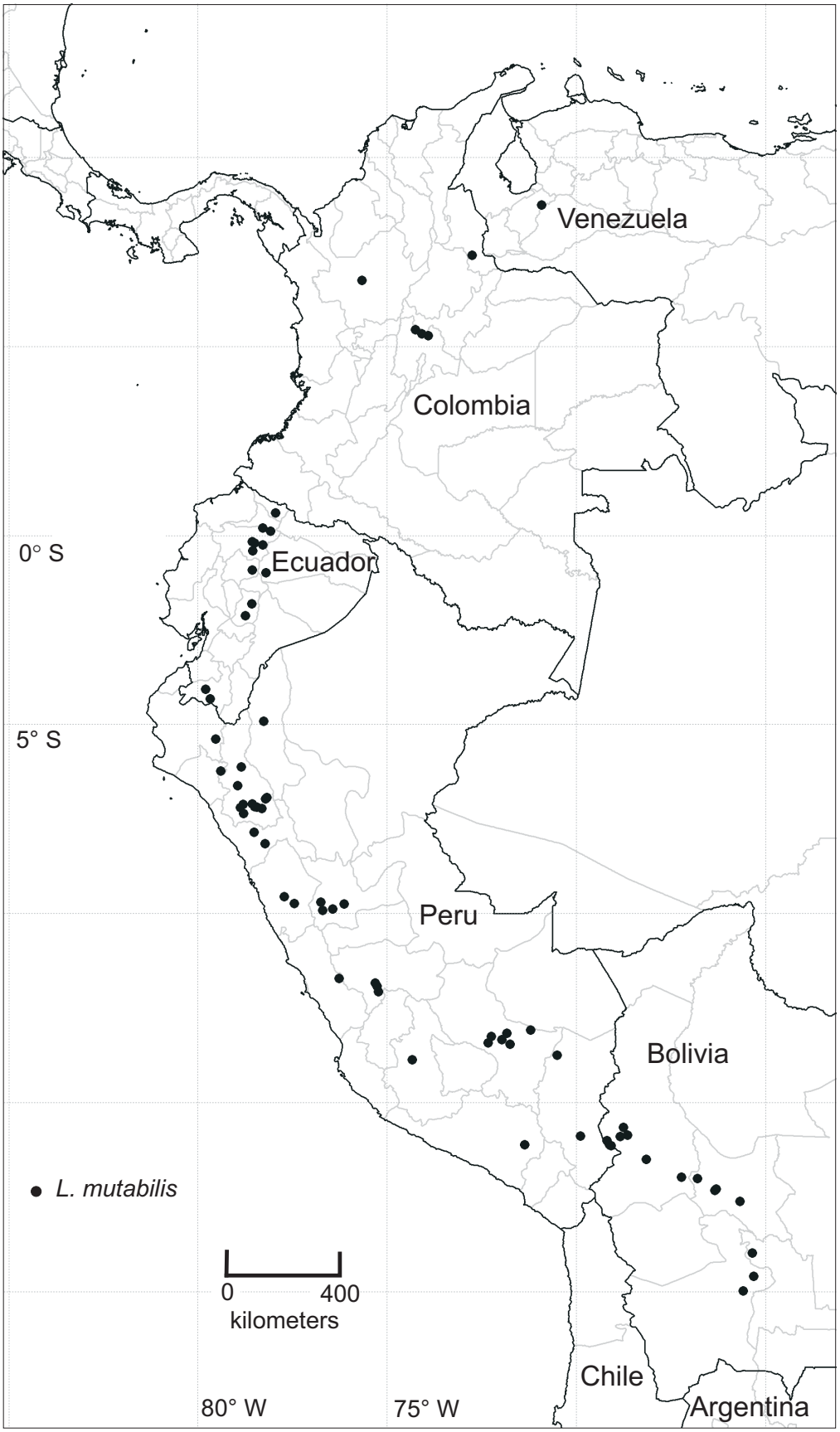
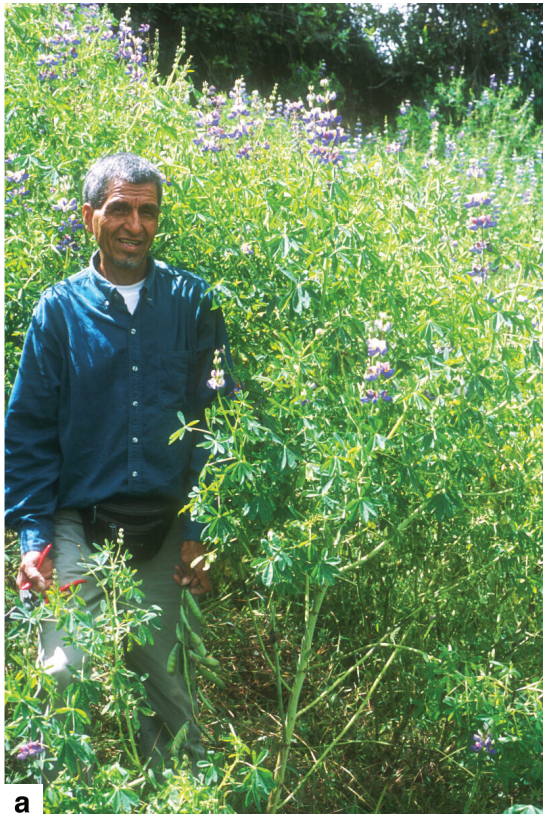


Fig. A1.2 Distribution of *L. mutabilis* Sweet
One specimen was seen from Chile but with no locality details. No specimens have been seen from Argentina but there are reports of *L. mutabilis* cultivation in the Andean region of NW Argentina (Blanco, 1984). This species is found at an altitudinal range of 2000 - 4100 m.

Fig. A1.3 ***Lupinus mutabilis*** **Sweet** Sweet's British Flower Garden 1825 Ser. I. ii: t. 130 Illustration accompanying the protologue of *L. mutabilis* Sweet. Note the similarity (with artistic license) to the lectotype specimen designated here (Fig. A1.4).

Fig. A1.4 Original material of *Lupinus mutabilis* Sweet here designated as the lectotype (LINN) The notes which correspond to those in the protologue and the similarities of the illustration accompanying it indicate that this specimen provided the original material for the description of *L. mutabilis* Sweet.



a

b

c

d

e

Fig. A1.5 Ecotypic variation in habit within *L. mutabilis* – Annual or biennial herb or shrub with erect habit.

a. *CEH2220*, Celedín, Peru, b. *RJE37*, Otuzco, Peru, c. & e. *RJE56*, Santiago de Chuco, Peru, d. *CEH2326*, Manco Kapac, Bolivia. Photos a and d CEH and b, c and e RJE.

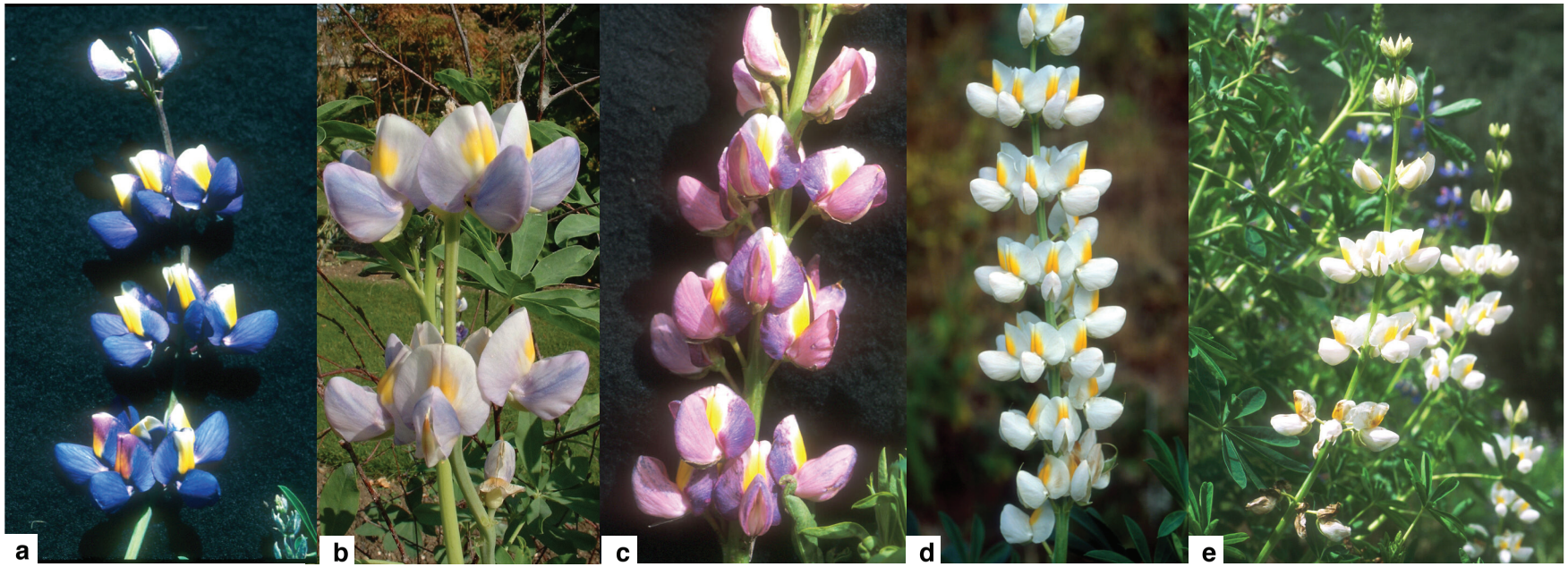


Fig. A1.6 Flower colour variation within *L. mutabilis* a. Bright blue, *CEH2039*, Pira, Peru, b. Lavender, Oxford Botanic Garden, c. Pink, *CEH2253*, Millipaya, Bolivia, d. Pale blue, *CEH2291*, Quillacollo, Bolivia, e. White, *CEH2220*, Celedín, Peru. Photographs a, b-e CEH and b RJE.

Fig. A1.7 Colour changes in flowers of *L. mutabilis RJE56* Photograph RJE.

Fig. A1.8 Pods and seeds of *L. mutabilis* *RJE56* Note that pod is indehiscent and was shelled. Photograph copyright RJE.

Fig. A1.9 Seed colour variation in *L. mutabilis* in cultivation around Lake Titicaca in Bolivia *CEH2326* Photograph RJE.

Fig. A1.10 Field of *L. mutabilis* Photograph RJE. Northern Peru

Fig. A1.11 Tarwi *L. mutabilis* seed for sale in a market in Copacabana, Lake Titicaca, Bolivia Photograph CEH.

2 Lupinus piurensis C. P. Smith, Sp. Lupinorum 768 (1953). – Type: PERU. Piura: Huancabamba, Ascending from Canchaque to Huancabamba, nr Canchaque, 5º21' S, 79 º 37' W alt. 2000 m, 22 Jun 1952, *Ochoa C 1799,* (Holotype CAS!; Isotype MOL!).

Annual or short-lived herbaceous perennial subshrub, with one principal stem or branching from near the base, 0.4-0.6 (-1) m in height. Stems sometimes weakly woody at base, often hollow, glabrous, glaucous or sparsely pubescent. Stipules 2.5-4 (-6) mm long and 0.5-1 mm wide, adnate region linear, adnate 50% - 75% of length, free tip ovate, often reflexed, often caducous, glabrous. Leaves palmately digitate, (20-) 40-130 mm long with 5-7 (-8) leaflets per leaf; petioles (15-) 40-80 (-100) mm in length and 0.5-1 mm in diameter, glabrous; leaflets (10-) 25-40 (-60) mm long and 7-13 mm wide, variable within each leaf, leaflet apex mucronate, blade oblanceolate, narrow-oblanceolate or narrow-elliptic, base attenuate, immature leaflets conduplicate, mature leaflets held horizontally, thin, adaxial surface glabrous, glaucous or with sparse cover of short white hairs. Erect terminal inflorescence (10-) 15 - 30 cm tall held predominantly above foliage, racemes loosely verticillate with 10-35 flowers in 3-6 whorls each of 3-5 flowers; peduncle diameter 1-2 mm, glabrous or with a sparse covering of white appressed hairs; pedicels at anthesis 6-7 mm long, 0.35-0.6 mm in diameter glabrous or with a sparse covering of appressed white hairs; floral bracts 3-5 mm long and 0.5-1 mm wide, lanceolate, apex attenuate, glabrous or sparsely covered with appressed white hairs, present in immature bud but readily shed before anthesis; bracteoles 0.5-0.8 mm long and 0.2-0.4 mm wide, lanceolate, apex attenuate, glabrous or with sparse cover of appressed white hairs, readily shed before anthesis; calyx upper lip 6-8 mm long and 1-2 mm high, bidentate the notch 1-1.5 mm, lower lip 8-10 mm long and 2-3 mm high, concave, green sometimes tinged inky blue, glabrous or with sparse cover of appressed white hairs; banner petal orbicular 10-21 mm high by 11-18 mm wide, bright blue with a central white stripe and bright yellow spot, both stripe and spot darkening to purple as flowers fade, rarely white with yellow spot fading to pinky purple, glabrous or with few white hairs on back of banner only in central region and predominantly towards apex; wing petals 12-20 mm wide and 5-11 mm high, lobe height 2-3 mm, claw length 2-2.5 mm, united at tip, bright blue or rarely white, glabrous with 5-8 rows of corrugated petal sculpturing at base near lobe; keel petals 10-13 mm long, 5-6 mm high, white/cream with dark blue/purple apex, glabrous or with minute hairs on sub-apical margin; floral parts and particularly staminal sheath persistent through to pod maturity. Pods 5-22 on hollow infructescences, (25-) 30-53 mm long, 7-10 mm wide and 3-6 mm thick, laterally compressed, narrow-oblong, apex mucronate occasionally with persistent style, base cuneate asymmetrical towards ventral suture, dorsal suture undulate, ventral suture entire, 0.5 mm thickening surrounding both sutures, lateral constrictions visible, venation often visible near ventral suture, immature pods mid-green, glabrous and glossy or densely covered in white or yellow silky hairs, mature pods mid to dark brown glabrous or with medium-dense cover of silky white hairs, 3-7 seeds per pod, dehiscent. Seeds in outline elliptical or broadly oblong, (4.6-) 5-5.6 mm long, (3-) 3.4-6.8 mm wide and 2.5-3.4 mm thick, 0.01-0.03 g weight, marbled creams, greys and browns with dark brown line separating corner containing the hilar, testa smooth to touch. Cotyledons petiolate, fleshy. First seedling leaves with (4-) 5 (-6) leaflets. Chromosome number 2n=48 (Conterato & Schifino-Wittmann, 2006). (Fig. A2.1, A2.3).

Phenology. Flowering and fruiting material has been seen from March to October but the main season is May and June.

Distribution. This species is found in scattered populations on gravelly or scrubby hillsides, disturbed roadsides and pastures from southern Ecuador (Provinces: LO) to central Peru (Departments: CAJ, LL, PI). Commonly associated species include shrubby Asteraceae, *Alnus*, *Dalea, Eucalyptus*, *Pinus*, *Spartium* and grasses. This species is restricted to the western slopes of the Andes and has been found at an altitudinal range of 1650-3300 m. (Fig. A2.2).

Vernacular names. chocho kon, chocho silvestre, churco, tabre, chocho de campo

Notes. Smith (page 768, 1953) named this species *piurensis* (Greek for thin-leaved), in recognition of it's thin leaflets and commented this species is *"perhaps the thinnest leaved and most nearly glabrous-leaved lupine"* and perhaps the more obvious origin of the name from 'Piura' the department where it was first collected is coincidental. This character is also shared by *L. mutabilis* of which this species is a putative progenitor. In fact both in the field and herbarium *L. piurensis* specimens closely resemble *L. mutabilis* but lack the large pods and large seeds of that species. Although *L. piurensis* forms a smaller and generally less robust plant than *L. mutabilis*, without mature pods confident identification is often difficult.

Herbarium specimens and particularly fragmentary ones can also be difficult to distinguish from *L. jamesonianus* C.P.Sm., a scrambler from Ecuador. The distribution of these two species overlaps in the province of Loja, Ecuador. In the field and greenhouse (*personal observation*) however these two species are clearly distinguished by habit. Additionally, *L. jamesonianus* has pubescent stems and swollen nodes.

Lupinus piurensis var. *cotumazicus* is described as a new variety in recognition of material from the Cotumaza region of Cajamarca, Peru, which shares a number of fixed morphological characters and additionally to account for the extremes of quantitative measurements for stipule, banner and wing size. This group of specimens have been maintained as a variety within *L. piurensis* rather than as a distinct species since one collection (*Hutchinson 5055*) contains plants which are pubescent (F, M) and those which are glabrous (GH, K) i.e. corresponding to both subspecies but which are otherwise identical.

Lupinus piurensis var. *cotumazicus* shares similarities in flower size and indumentum with *L. ellsworthianus* but has a more erect habit and grows on the western slopes of the Andes.

2a L. piurensis var. piurensis

Additional specimens examined. **ECUADOR,** Loja, Km 8 rd from Sozoranga to Cariamanga, 4°22' S, 79°43' W alt. 2000 m, 14 Jun 1996, *Klitgaard BB 214* (K, NY, QCA); Rd from Puente de Lucarque to Celica, 14 km from junction with Catacocha - Macara rd, 4°08'77" S, 79°52'82" W alt. 1150 m, 27 May 1996, *Lewis GP 2330* (FHO, K, QCNE);

PERU, Cajamarca, Chiquidan de San Juan to Cajamarca, 7°13' S, 78°07'60" W, 17 May 1952, *Ochoa C 1473* (F, GH); Cajamarca, Yumagual, District of San Juan, 20 Jun 1993, *Cabanillas J 669* (F); Nr km 154 on rd Cajamarca to San Juan, 7°13' S, 78°30' W alt. 2730 m, 16 Jun 2004, *Eastwood RJ 110* (FHO, K, MOL,

USM); Nr km 151 on rd Cajamarca to San Juan, 7°16' S, 78°30' W alt. 2600 m, 16 Jun 2004, *Eastwood RJ 113* (FHO, K, MOL, USM); Cajamarca, Cerro Chupacotoy. 2 Jun 1992, *Ochoa C 16138* (US); rd from Cajamarca to Pacasmayo, nr Paso El Gavilan, 7°10' S, 78°31' W alt. 2950 m, 10 Apr 1984, *Sanchez-Vega I 3350* (F); Contumaza, La Pampa, Guzmango, 7°25' S, 78°53' W alt. 2500 m, 7 May 1989, *Sagastegui A 14116* (F, NY); San Juan, Caserio Choten, main rd E from Pacasmayo towards Cajamarca, 7°15'28" S, 78°30'41" W alt. 2710 m, 7 Jun 2001, *Hughes CE 1997* (FHO, K, MOL, NY); San Juan, Caserio Choten, main rd E from Pacasmayo towards Cajamarca, 7°15'28" S, 78°30'41" W alt. 2710 m, 7 Jun 2001, *Hughes CE 1998* (FHO, K, MOL, NY); San Juan, Yumagual, between San Juan and El Gavilan. 10 Jun 1975, *Sanchez-Vega I 1536* (F); San Miguel, 6°59'28" S, 78°49'38" W alt. 2700 m, *Diaz A s.n.* (USM); San Pablo, El Molino. 13 Jun 1993, *Cabanillas J 552* (F); Adjacent to Cerro San Pedro and San Pablo, 7°06' S, 78°50' W alt. 2500 m, 13 Jun 1993, *Cabanillas J 595* (F); San Pablo, El Molino, 7°08' S, 78°50' W alt. 2600 m, 29 May 1982, *Sanchez-Vega I 6582* (NY); San Pablo, 1 km above and NE of the town of San Pablo, 7°12' S, 78°50' W alt. 2630 m, 13 May 1964, *Hutchison PC 5055* (GH, K); La Libertad, Otuzco, Rd to Paranday (Sinsicap), 7°50' S, 78°45'80" W alt. 2650 m, 1 May 1954, *Lopez A 1049* (US); **Piura,** Huancabamba, El Tambo, 5°22' S, 79°33' W alt. 2770 m, *Acleto C 267* (USM); Huancabamba, 12 km E Canchaque on rd to Huancabamba, 5°21' S, 79°34' W alt. 1650 m, 17 Jun 2004, *Eastwood RJ 122* (FHO, K, MOL); Huancabamba, Palambla. 10 km E on rd from Huancabamba to Canchaque, 5°21' S, 79°34' W alt. 1660 m, 19 Jun 2004, *Eastwood RJ 134* (FHO, K, MOL, USM); Porculla, Cerro Porculla, 3-11km W of the continental divide down W slopes of Cerro, 5°50' S, 79°30' W, 12 Jun 1966, *Edwin G 3774* (F, GH, NY, US, USM);

UK, Cultivated in Oxford Botanic Garden. 2 Oct 2003, *Eastwood RJ 16* (FHO).

Description of distinguishing characters. The most reliable characters to distinguish var. *piurensis* from var. *contumazicus* are the glabrous, glaucous stems (Fig. A2.3d) which are commonly pithy and barely lignified. Stipules tend to be shorter (2.5-4 mm) and flowers larger (banner (12-) 18-21 mm high, (11-) 15-18 mm wide, wing 8-11 mm high and (12-) 14-20 mm wide) but these are continuously variable quantitative characters.

Distribution. *Lupinus piurensis* var. *piurensis* is distributed across the range of the species from southern Ecuador (Province: LO) to central Peru (Departments: CAJ, LL, PI) (Fig A2.2). This variety is abundant in Piura and Cajamarca and apparently scarce at the limits of it's distribution in Loja, Ecuador and La Libertad, Peru.

2b L. piurensis var. cotumazicus Eastwood & Hughes var. nov**.** – Type: PERU, Cajamarca, Contumaza, Guzmango, 7º12' S, 78º53' W alt 2600 m, 27 Jul 1973, *Sagastegui A 7724,* (Holotype F!; Isotype NY!).

Additional specimens examined. **PERU**, **Cajamarca,** Above San Pablo, 7°06' S, 78°50' W alt. 2450 m, 22 May 1975, *Sagastegui A 7932* (GH, US); Vicinity of Contumaza, 7°22'08" S, 78°48' W alt. 2935 m, 3 May 1999, *Binder M 1999/64* (F, M); Between Cascas and Contumaza, 7°27' S, 78°49' W alt. 2250 m, 19 May 1962, *Lopez A 3674* (US); Contumaza, Guzmango, 7°12' S, 78°53' W alt. 2600 m, 5 Oct 1966, *Lopez A 6244* (ISC); Bosque de Cachil (Cascas-Contumaza). Lindero del bosque, *Lopez A 9117* (FHO); Contumaza, Juque, Trinidad, 7°20' S, 78°46' W alt. 3230 m, 19 Jun 1962, *Sagastegui A 3819* (US); Contumaza, Trinidad to Ventarron at Fila la Carga, 7°22'08" S, 78°48' W alt. 2800 m, 6 Jul 1977, *Sagastegui A 8973* (NY); Snamon, above Contumaza, 7°04' S, 78°49' W alt. 2800 m, 8 Jul 1977, *Sagastegui A 9004* (NY); Between Toledo and Contumaza. 25 May 1981, *Sagastegui A 9906* (F, NY); Contumaza, Totorillas to Marin, Guzmango, 7°22' S, 78°51' W alt. 3300 m, 1 May 1982, *Sagastegui A 10371* (GH, NY); Contumaza, San Benito, between San Lorenzo and Tucac, 7°26' S, 78°56' W alt. 2200 m, 14 Jun 1983, *Sagastegui A 10672* (F, GH, NY); Contumaza, Lleden, 7°04' S, 78°48' W alt. 2200 m, 29 Jun 1983, *Sagastegui A 10871* (F, NY); Contumaza, Turnoff to Bosque de Cachil between Cascas and Contumaza, 30 May 1990, *Sagastegui A 14280* (F); Vicinity of Contumaza, route to Cascas, 7°22' S, 87°48' W alt. 2600 m, 17 Apr 1992, *Sagastegui A 14564* (F); San Pablo, 3-4 km NE of San Pablo, 7°06'57" S, 78°48'37" W alt. 2590 m, 20 Apr 2002, *Hughes CE 2196* (FHO, K, MOL); San Pablo, 1 km above and NE of the town of San Pablo, 7°12' S, 78°50' W alt. 2630 m, 13 May 1964, *Hutchison PC 5055* (F, M).

Description of distinguishing characters**.** *Lupinus piurensis* var. *cotumazicus* has sparsely pubescent stems which are commonly woody. Stipules are often larger than var. *piurensis* ranging from 3-6 mm. Flowers are commonly smaller, the banner petal 10-12 mm high, 11-14 mm wide and the wing petals 5-9 mm high 12-13 mm wide.

Distribution. The distribution of the var. is restricted to the Cotumaza region of Cajamarca, Peru, where it is quite common (Fig A2.2). It grows in sympatry with *L. piurensis* var. *piurensis* in this region.

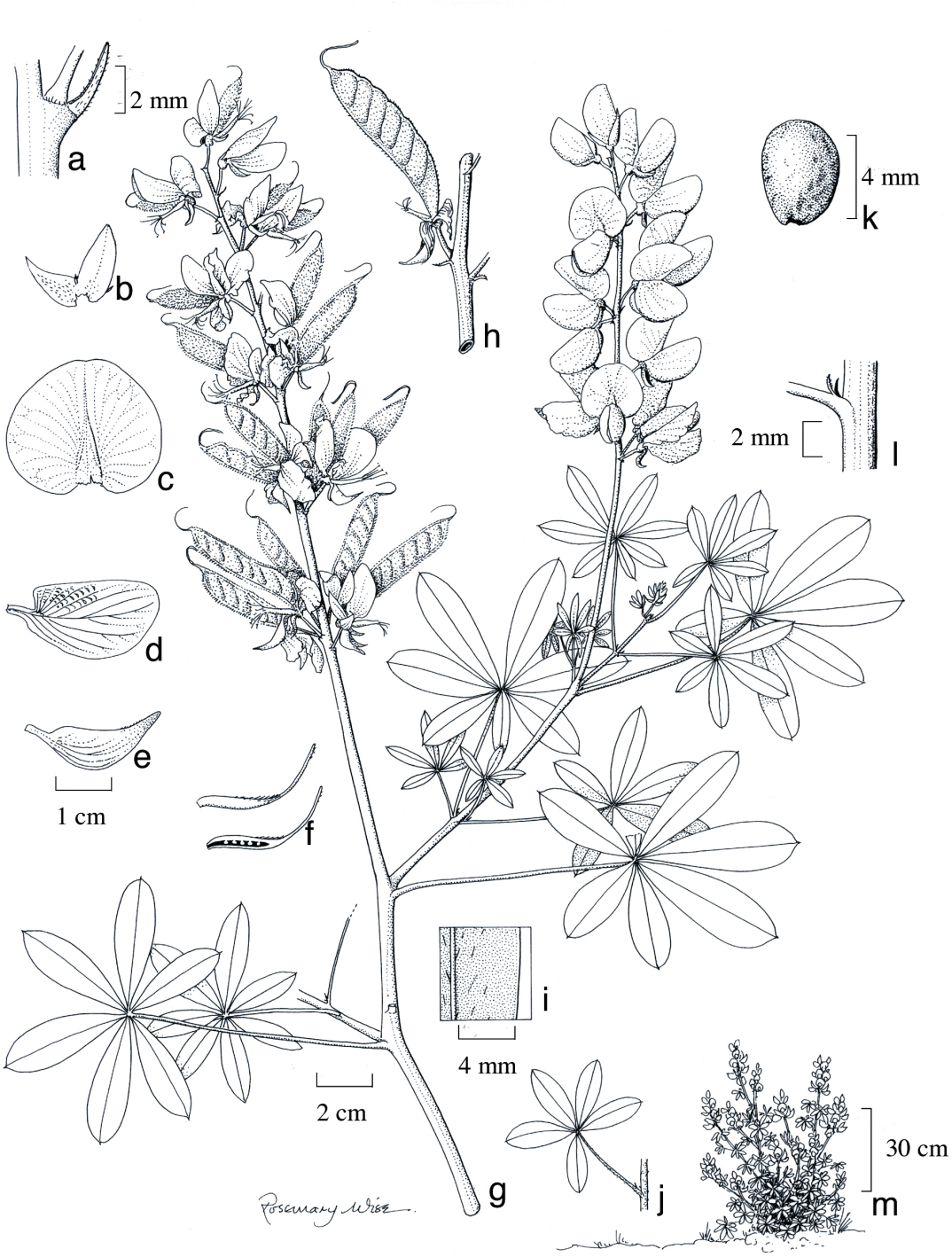


Fig. A2.1 *Lupinus piurensis* C. P. Smith
a. Floral bract b. Calyx, c. Banner, d. Wing petal, e. Keel petal, f. Gynoecium, g. Flowering shoot with immature pods, h. Mature pod, i. Detail of abaxial leaflet surface, j. Leaf, k. Seed, l. Stipule, m. Habit. a.-i., and k.-m. *L. piurensis* var. *piurensis,* j. *L. piurensis* var. *contumazicus*. Based on a-i and l. *CEH1998*, j. *Sagastegui 7724* k. *RJE113*, and m. *CEH1997*. Illustration by Rosemary Wise.

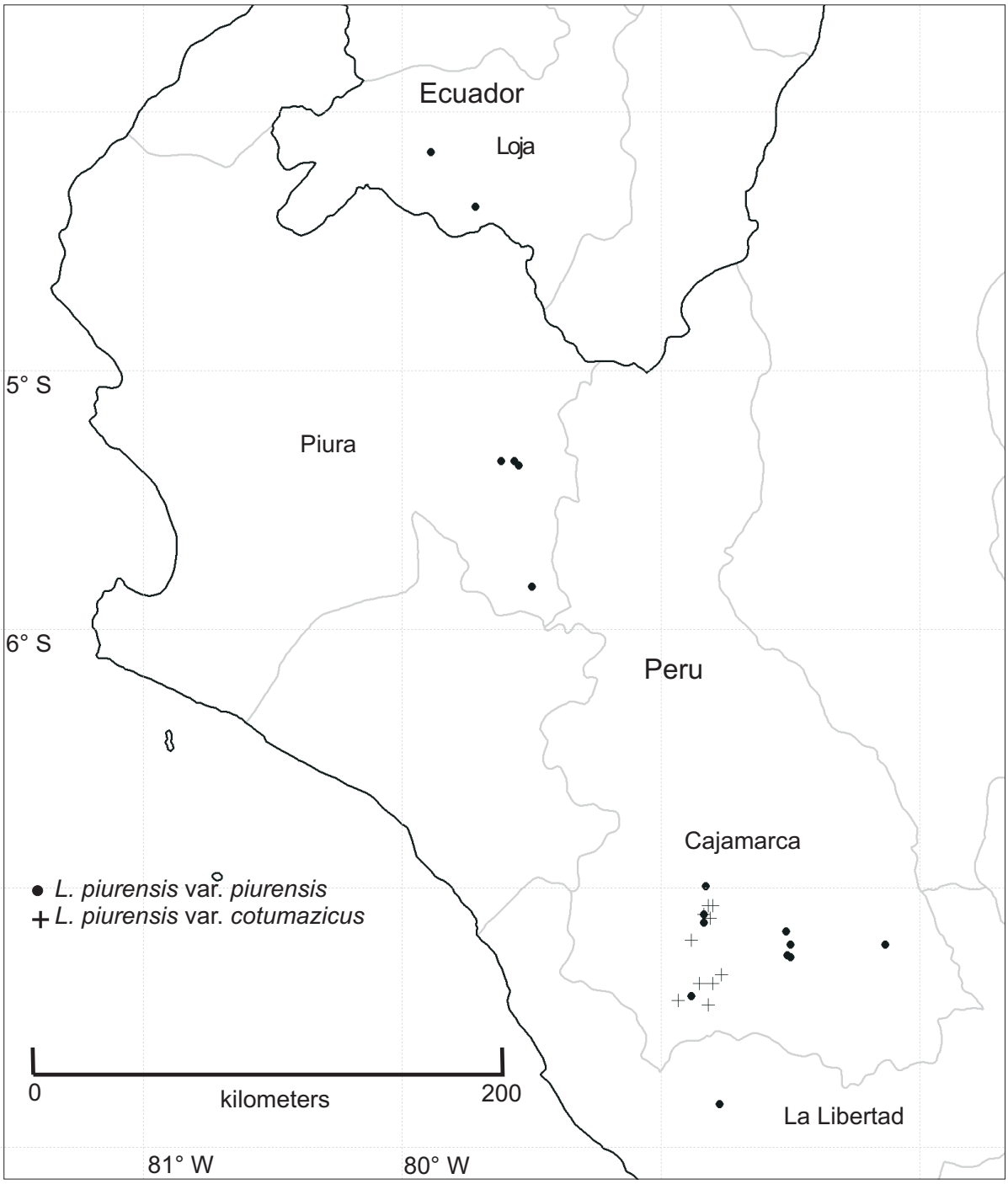


Fig. A2.2 Distribution of *L. piurensis* C.P. Smith
This species is found at an altitudinal range of 1650-3300m.

Fig. A2.3 *L. piurensis* var. *piurensis* a. Habit, *CEH1997*, b. Leaf, Seedling of *CEH1997*, c. Seed *RJE113*, d. glaucous shoot with stipules, *RJE113*, e, Infloresence, *RJE110*, f. Fertile shoot, *CEH1997*. No photographs of *L. piurenesis* var. *cotumazicus* were available. Scale bars a. 5 cm, b., d., e. and f. 10 mm and c, 5 mm. Photographs a, e. and f. CEH and b-d RJE.

3 Lupinus semperflorens Hartw. ex Benth. Pl. Hartw. 127 (1844). – Type: ECUADOR. Loja: Mountains nr Loja, Jun 1843, *Hartweg 724* (Holotype: K!; Isotype: BM!, CGE!, E!, F!, G, NY!, OXF!, P, W).

Perennial shrub or small treelet with one principal stem (0.8-) 1.5-6 (-10) m in height, dbh of principal stem to 10 cm. Outer bark grey to yellow-brown, vertically striate or smooth. Inner bark green. Stems of seedlings glabrous and glaucous. Young stems brown with dense cover of white silky hairs. Stipules (7-) 12-22 (-32) mm long and 1-2 (-3) mm wide, adnate region linear, adnate length 6-12 (-16) mm ~30% - 50% of length, free tips lanceolate often reflexed, persistent, densely covered with silky hairs. Leaves palmately digitate, 30-150 (-230) mm long with 7-9 leaflets per leaf; petioles 15-96 (-145) mm long and 0.5-1.5 (-3) mm in diameter, densely covered in sparse short white appressed or erect hairs; leaflets 18-60 (-100) mm long and 4-14 (-20) mm wide with a range of leaflet sizes in each leaf, leaflet apex mucronate, blade oblanceolate, narrow-oblanceolate, narrow-oblong or narrow-elliptic, base attenuate, immature leaflets conduplicate, mature leaflets often drooping, thin, adaxial surface glabrous, abaxial surface with dense cover of appressed white silky hairs, markedly discolorous when dry. Erect terminal inflorescence 15 - >30.0 cm tall held predominantly above foliage, racemes markedly or loosely verticillate with (15-) 25-45 flowers in 4-12 whorls each with 3-5 flowers; peduncle diameter 1.5-2.5 (-4.5) mm with dense cover of erect or appressed white hairs; pedicels at anthesis 6-11 mm long, 0.75-1 mm in diameter with dense cover of erect or appressed white hairs; floral bracts 10-14 mm long and 1.5-2.5 mm wide, ovate, attenuate, with dense cover of erect or appressed white hairs, present in immature bud but not persisting at anthesis; bracteoles 3.0-3.9 mm long and 0.5-1 mm wide, lanceolate, apex attenuate, densely covered in white silky hairs, often deciduous before anthesis; calyx upper lip 10-11 (-14.5) mm long and 2-3 mm high, bidentate the notch 2-2.5 mm, lower lip 10-13 mm long and 2-3.5 mm high, concave, green sometimes tinged inky blue, densely covered in silky white hairs; banner petal orbicular (13-) 15-19 mm high by 19-25 mm wide, lavender to bright blue with or without a central white stripe, pale to bright yellow spot, both stripe and spot darken to purple as flower fades, white hairs on back of banner only on central region and predominantly towards apex; wing petals (15-) 17-20.5 mm wide and 9-13 mm high, lobe height 2-3 mm, claw length 2-3.5 mm, united at apex, lavender to bright blue, glabrous with 6-8 rows of corrugated petal sculpturing at base near lobe; keel 10-12 (-16) mm long, 6-7 mm high, white/cream with dark blue/purple apical region, glabrous or rarely with few hairs on sub-apical margin; faintly sweetly scented, floral parts particularly staminal sheath persistent through to pod maturity. Pods on a slightly woody or hollow pithy infructescence, 5-30 pods per infructscence, pods 40-80 mm long, 10-12 mm wide and 3.5-5 mm thick, laterally compressed, narrow-oblong, apex mucronate occasionally with a persistent style, base cuneate, asymmetrical, dorsal suture undulate, ventral suture entire, 0.5 mm thickening surrounding both margins, lateral constrictions visible, venation often visible and prominent near ventral suture, immature pods mid-green often tinged inky blue at sutures and particularly along dorsal suture, mature pods mid to dark brown, immature pods densely covered in white/grey pubescence, mature pods sparsely hairy or almost glabrous, (1-) 4-7 seeds per pod, dehiscent. Seeds in outline elliptical or broadly oblong, 4.5-7.5 mm long, 2.5-5.5 mm wide, and 2-3.5 mm thick, 0.02-0.06 g weight, marbled creams, greys and browns with dark brown line separating corner containing the hilar, testa smooth to touch. Cotyledons petiolate, fleshy. First seedling leaves with 4-5 leaflets. Chromosome number 2n=48 (Conterato & Schifino-Wittmann, 2006). (Fig. A3.1).

Phenology. This species has been collected in flower and fruit all year round.

Distribution. *Lupinus semperflorens* is restricted to the central Andes largely between latitudes 3°5' S and 7°2' S but two single outlying collections have been seen from further south near Mito, Junin and Abancay, Apurímac, Peru at 11°53' S and 13°3' S respectively. The distribution spans southern Ecuador (Provinces: LO) to central Peru (Departments: AM, AP, CAJ, JU, PI) (Fig. A3.2). The altitudinal range of this species is 1750-3600 m. *Lupinus semperflorens* is found on dry or moist scrubby hillsides, in disturbed roadside scrub and in grassy pasture. It is locally abundant and can form a major component of secondary vegetation in some areas. Commonly associated species are *Alnus,* shrubby *Asteraceae, Baccharis, Escalonia,* and *Polylepis* and planted *Eucalyptus* and *Pinus.*

Vernacular names. chocho soro, chugur, margaretta, tabre (this is more commonly applied to *L. piurensis* C. P. Sm.)

Ethnobotany. Used as a source of firewood.

Notes. Identification of this species is relatively straightforward and thus the two outlying collections of this species are intriguing. Additional to specimens, photographs of the most southerly collection (Hughes 2386A) were examined. These show clearly the characteristic drooping leaflet habit seen in populations in Loja, Ecuador, Piura and Cajamarca, Peru (*personal observation*). This disjunct distribution can be attributed to the geographic structure, environmental variation and glacial history, all of which have been highly influential in determining Andean plant distributions.

There is considerable size variation of all parts of this species across its range. Particularly small plants are found in the moist montane scrub near Canchaque, northern Peru, whilst the largest individuals come from Cajamarca, central Peru, and Loja, southern Ecuador. All specimens observed except CEH 2386A, the most southerly collection, have glabrous keel petals.

L. semperflorens is identified by a combination of characters: silky stem and abaxial leaf indumentum, leaves drying discolorous, commonly drooping leaflets, and stipules (7-) 12-22 (-32) mm long with a silky indumentum.

Herbarium specimens of two species *L. ellsworthianus* C.P. Sm. and *L. ochoanus* C.P. Sm. can resemble *L. semperflorens* at first glance due to a tendancy for the leaflets to dry discolorous. On further inspection these can be distinguished from *L. semperflorens. L. ellsworthianus* C.P. Sm. is much smaller in all respects and has dainty 'button-like' flowers. The abaxial leaflet indumentum of *L. ochoanus* C.P. Sm is formed of short curled hairs dissimilar to the appressed straight silky indumentum of *L. semperflorens. L. ochoanus* C.P. Sm also has much smaller flowers.

Some herbarium material of *L. semperflorens* Benth. has been mis-identified as *L. mutabilis* to which it shows some similarity particularly in it's large flower size. However, *L. semperflorens* Benth. is a much larger woody perennial and it is immature plants which are most easily confused with *L. mutabilis*. It is distinguished from *L. mutabilis* by the larger stipules, indumentum and the dehiscent pods.

L. semperflorens is split into two varieties to recognise a group of plants distributed in the Amazonas department of Peru which show fixed leaflet shape differences.

3a Lupinus semperflorens var. semperflorens

Lupinus macbrideanus C. P. Sm., Sp. Lupinorum 159 (1940). – Type: PERU. Junín, Mito, Cani, 11°53' S, 75°19' W alt. 2700 m, 21 Mar 1923, *Macbride JF 3464* (Holotype: US!; Isotype: F!,).

Lupinus storkianus C. P. Sm., Sp. Lupinorum 606 (1948). – Type: PERU. Cajamarca, Chota, Above Huambos, 6°28' S, 78°58' W alt. 2500 m, 21 Apr 1923, *Stork HE & Horton OB 10201* (Holotype: UC! - Unicate).

Lupinus semperflorens var. peruensis C. P. Sm., Sp. Lupinorum 768 (1953). – Type: PERU. Piura, Montaña de Cuijos, nr Arabaca (possibly spelling error for Ayabaca), 4°39' S, 79°44' W alt 2000 m, Jun 1952, *Ochoa 1793* (Lectotype: CAS! here designated; Isolectotype: US!). Protologue omits type collector and number citing only the locality.

Additional specimens examined. **ECUADOR,** Jul 1884, *Cross s.n.*, (BM, NY); **Loja,** Las Chinchas. 12 Apr 1944, *Acosta Solis M 7781* (F, M); Loja. 8 km W on the old rd from Catamayo, 4°00'06" S, 79°15'11" W alt. 2550 m, 20 Jul 2004, *Eastwood RJ 219* (FHO, K, QCNE); 2 km S of Loja Botanic Garden on the road from Loja to Villcabamba, 4°03'20" S, 79°11'38" W alt. 2200 m, 21 Jul 2004, *Eastwood RJ 220* (FHO, K, QCNE); Mountains nr Loja, 3°57' S, 79°19' W, Jun 1843, *Hartweg 724* (BM, CGE, E, F, G, K, NY, OXF, P, W); Km 10 new rd Loja to Catamayo, 3°64' S, 79°19' W alt. 2000 m, 9 Apr 1996, *Klitgaard BB 126* (FHO, K, NY, QCNE); 1 km S of Yungana, track to Sierra Toledo, km 3.3, 4°23' S, 79°09' W alt. 2150 m, 14 Jan 1998, *Lewis GP 3775* (FHO, K); Loja, Vicinity of Loja, 4°00' S, 79°13' W, 30 Sep 1918, *Rose JN 23297* (NY);

PERU, Amazonas, Cos. de Calla-Calla, on rd 30 km SW of Leimebamba (Km 391), 6°40' S, 77°50' W, 12 Sep 1970, *Ambrose JD 2381* (F, NY); Km 99 rd from Celendín to Leimebamba, 42 km NE of Balsas on upper slopes E of the gargantuan cleft cut by the Rio Maranon Valley, 6°47'04" S, 77°56'10" W alt. 2830 m, 12 Jun 2001, *Hughes CE 2025* (CAS, FHO, K, MOL, NY); Bongara, 35 km NE of Pedro Ruíz, 2km E of Florída, rd to Rioja, immediately N of Laguna Pomacochas, 5°48'31" S, 77°57'09" W alt. 2230 m, 15 Jun 2001, *Hughes CE 2032* (CAS, F, FHO, K, MEXU, MO, MOL, NY); Bongara, 20 km NE of Pedro Ruíz, rd to Pomacochas, 5°45' S, 77°54' W alt. 2750 m, 18 Feb 1978, *Wasshausen DC 947* (G, K, NY, US, USM); Chachapoyas, Cordillera Calla Calla on the Balsas - Leimebamba rd nr km 386, 6°40' S, 77°50' W alt. 3200 m, 9 Jul 1977, *Duncan T 2612* (UC); Chachapoyas, Between Saullamur and Calla Calla, 6°40' S, 77°50' W alt. 3250 m, 3 Aug 1978, *Ruiz Vigo W 2139* (F); Chachapoyas, 41 km from Puente Chocanto, rd from Celendín to Chachapoyas, Maranon Valley, 6°20' S, 77°55' W alt. 2800 m, 24 May 1984, *Smith DN 7092* (FHO, USM); Chachapoyas, Middle and upper slopes of Puma-urcu, 6°80' S, 77°43' W alt. 2600 m, 25 May 1962, *Wurdack JJ 540* (F, NY, UC, US, USM); Chachapoyas, Upper slopes of Puma-urcu, 6°80' S, 77°43' W alt. 2850 m, 1 Jun 1962, *Wurdack JJ 673* (F, K, NY, UC, US, USM); Luya, Along trail between Balsas and Leimebamba, 6°20' S, 77°46' W alt. 2550 m, 21 Aug 1943, *Evinger EL 479* (US); **Apurímac**, Approx 20 km S of Abancay by rd, ca. km 166-167.5 rd from Cuzco to Abancay, 13°34'28" S, 72°45'51" W alt. 3100 m, 21 Nov 2003, *Hughes CE 2386A* (FHO, K, MOL); **Cajamarca,** Porcón Alto. 7°04' S, 78°04' W alt. 3300 m, 15 Jun 2004, *Eastwood RJ 109* (FHO, K, MOL, USM); Cajamarca, Just below Porcón Bajo, 12.5km N of Cajamarca rd to Yanacocha and Chota, 7°05'42" S, 78°34'28" W alt. 3040 m, 10 Jun 2001, *Hughes CE 2012* (FHO); Cajamarca, Sais, Jose Carlos Mariategui, km 1-19 on San Juan Sunchubamba rd, 4 Jun 1984, *Smith DN 7465* (FHO, USM); Chota, Paccha towards Chadin nr Caserio Marcopata. 22 Jul 1993, *Sanchez-Vega I 851* (F); Contumaza, Between Cascas and Contumaza, 7°22' S, 78°46' W alt. 2640 m, 20 May 1962, *Lopez A 3719* (US); Contumaza, Vicinity of Guzmango, 7°12' S, 78°53' W alt. 2600 m, 24 Jul 1992, *Sagastegui A 14773* (F); Contumaza, Entrance to Bosque Cachil, 6°23' S, 79°17' W alt. 2500 m, 29 Jul 1993, *Sagastegui A 14964* (F); Cutervo, Between Cutervo and Cochabamba, 6°22' S, 78°49' W alt. 2300 m, 1 Aug 1946, *Ferreyra R 841* (UC, USM); Cutervo, Puente El Suro, San Andres, 6°51' S, 78°10' W alt. 2200 m, 16 Jul 1990, *Llatas S 2799* (F); Cutervo, NE of El Suro, rd to Madre Mia, 6°51' S, 78°10' W

alt. 2350 m, 25 Jun 1992, *Sanchez-Vega I 6318* (F); Hualgayoc. Rd from Cajamarca to Bambamarca, 6°46' S, 78°36' W alt. 3600 m, 9 May 1999, *Binder M 1999/97* (F, M); Hualgayoc, nr mining camp on Bambamarca to Cajamarca rd, 6°46' S, 78°36' W alt. 3200 m, 21 Feb 1983, *Smith DN 3629* (ISC, NY, USM); Jaen, Rd between El Espino and La Union, 6°31' S, 79°46' W, 17 Jul 1998, *Diaz C 9862* (FHO, MO); San Miguel, Vicinity of El Tingo, 7°11' S, 79°00' W alt. 2930 m, 9 Feb 2000, *Alvitez E 1068* (F, M); San Miguel, Quishuarpampa Agua Blanca, 7°02' S, 79°04' W alt. 2900 m, 4 Jul 1986, *Mostacero J 1199* (F, NY); San Miguel, Quishuarpampa, Agua Blanca, 7°02' S, 79°04' W alt. 2900 m, 4 Jul 1986, *Mostacero J 1202* (F, NY); San Miguel, Cerro Lanchez, Niepos, 6°55' S, 79°07' W alt. 2300 m, 29 Aug 1992, *Quiroz SL 3186* (F); San Miguel, Between Agua Blanca and El Tingo, 7°11' S, 79°00' W alt. 2900 m, 16 Feb 2000, *Rodriguez E 2305* (F, M); San Miguel, Niepos, rd to Lanche, 6°55' S, 79°07' W alt. 2500 m, 21 Aug 1992, *Yupaita G 34* (USM); Santa Cruz, Monteseco. Ca. 4.3 km (as the crow flies) NE of Montesceo. 16 May 1987, *Sagastegui A 69* (F); Cusco, Nr Riobamba, Lucumayo Valley, 12°58' S 72°39' W. 19 Jun 1915, *Cook OF 1332* (US); **Junín,** Mito, Village 11 km NE of Mito, nr Cani, 11°53' S, 75°19' W alt. 2700 m, 21 Apr 1923, *MacBride JF 3464* (F, G, US); **Lambayeque,** Ferrenafe, Huacapampa, above Tute, Canaris, 6°54' S, 78°80' W alt. 2800 m, 25 Jun 1989, *Quiroz SL 2515* (F); Piura, Ayabaca, 49 km E of Ayabaca, Cerro Aypate, 4°35' S, 79°32' W alt. 2750 m, 22 Sep 1991, *Gentry A 74996* (FHO); Ayabaca, Yacupampa - Cuyas (Ayabaca), 4°39' S, 79°44' W alt. 2500 m, 26 May 1971, *Lopez A 7783-a* (FHO, NY); Ayabaca, Ruins of Aypate, Coumindad Campesina de Tacalpo, Anexo Yanchala, 4°42'94" S, 79°34'25" W alt. 2750 m, 21 May 1996, *Quipuscoa V 516* (F); Huancabamba, 16 km E Canchaque on road to Huancabamba, 5°21' S, 79°32' W alt. 1880 m, 17 Jun 2004, *Eastwood RJ 124* (FHO, K, MOL, USM); Huancabamba, Above Palambla, 5°20' S, 79°34' W alt. 1750 m, 15 Sep 1985, *Ferreyra R 20385* (USM); Huancabamba, Cuello del Indo. Below Cuello del Indio. 4 May 1990, *Sanchez-Vega I 5310* (F).

Description of distinguishing characters. This variety has oblanceolate or narrow-oblanceolate leaflets.

Distribution**.** The distribution spans southern Ecuador (Provinces: LO) to central Peru (Departments: AM, AP, CAJ, JU, PI) (Fig. A3.2). Commonly associated species are *Alnus*, *Astragalus, Baccharis* and other shrubby Asteraceae, *Dalea* and *Escalonia*.

Notes. Observation of the type material of *L. storkianus* C.P.Sm. and *L. macbrideanus* C.P.Sm. suggests that these species should be treated as conspecific with *L. semperflorens* since they share a dense cover of appressed silky white hairs on the abaxial surface of obovate leaflets and have stipules of similar size to *L. semperflorens* and do not show unique characters. The locality for the type collection of *L. macbrideanus* C.P.Sm. is noted as "Peru, Cani, pueblo 7 miles NE of Mito" however in the protologue Smith gives Peru, Junin, Puno as the location. There is a Mito in Junin and a Cani in Puno. Since Mito is the major locality I think this collection is most likely from Junin. This locality is also more harmonious with the other collections of this species. Smith (page 768, 1953) described *L. semperflorens* var. *peruensis* C. P. Smith. in 'Species Lupinorum' whilst also commenting it is *"perhaps to close too* L. semperflorens *of Ecuador".* Following observation of type material and considering the geographic component this name is also treated as a synonym of *L. semperflorens*.

3b Lupinus semperflorens var. amazonus Eastwood & Hughes subsp. nov. – Type: PERU, Amazonas, Balsas, Km 96 rd from Celendín to Leimebamba, 39km E of Balsas, E side of Marañon Valley, 6°48'13" S, 77°56'41" W alt. 2700 m, 12 Jun 2001, *Hughes 2024* (Holotype: MOL!, Isotypes: CAS, FHO, NY)

Additional specimens examined. **PERU, Amazonas,** Leimebamba, 6°40' S, 77°50' W alt. 2300 m, 17 Dec 1962, *Woytkowski F 7764* (GH, UC, US); Bongara, 20 km NE of Pedro Ruíz, rd to Pomacochas, 5°45' S, 77°54' W alt. 2750 m, 18 Feb 1978, *Wasshausen DC 939* (US); Bongesa, Km 322 rd from Olmos to

Rioja, above Campamento Ingenio (km 300) on Rio Utcabamba, rd to Pomacocha. 31 Jan 1964, *Hutchison PC 3896* (F, G, GH, M, NY, UC, US, USM); Chachapoyas, Vicinity of Leimebamba, 6°40' S, 77°50' W alt. 2050 m, 31 May 1963, *Lopez A 4381* (US); Between Chachapoyas and Leimebamba, rd Chachapoyas to Cajamarca, Utcubamba Valley, 6°27' S, 77°50' W alt. 2300 m, 3 Sep 1983, *Smith DN 4970* (ISC, USM); Chachapoyas, Lower slopes of Puma-urcu above Hacienda Santa Isabel, 1.5-2 km SE of Chachapoyas, 6°80' S, 77°43' W alt. 2450 m, 20 May 1962, *Wurdack JJ 421* (F, K, NY, UC, US); Chachapoyas, Km 11 rd from Tingo Maria to Kuelap in mountains W of Rio Utcubamba, 6°24'04" S, 77°56'53" W alt. 2360 m, 14 Jun 2001, *Hughes CE 2030* (FHO); Rocky slopes immediately below the ancient fortified Chachapoyan city of Kuelap, 38 km (by rd) W of Tingo, in mountains W of the valley of the Rio Utcubamba, 6°25'30" S, 77°55'31" W alt. 2950 m, 14 Jun 2001, *Hughes CE 2031* (FHO); Chachapoyas, 6°28' S, 77°43' W *Mathews A 1570* (E, OXF); Mendoza to Chachapoyas, 6°22' S, 77°37' W alt. 2000 m, 15 Nov 2004, *Pennington TD 17693* (FHO, K, MOL); 12 km SSW of Leimebamba, rd to Balsas and Celendín, 6°42'50" S, 77°48'47" W alt. 2660 m, 13 Jun 2001, *Hughes CE 2029* (FHO); Jalca de Calla Calla, 30-37 km from Leimebamba, rd to Cajamarca, just below the pass on the Marañon valley side, 6°50' S, 77°50' W alt. 3550 m, 4 Sep 1983, *Smith DN 5043* (ISC); Luya, Tingo, Kuelap, 4 Sep 1998, *Vasquez R 25460*, (FHO, MO).

Description of distinguishing characters. This subspecies has narrow-oblong or narrow-elliptic leaflets.

Distribution. This subspecies is only found in Amazonas province of Peru (Fig. A3.2) where it can be sympatric with *L. semperflorens* var. *semperflorens.*

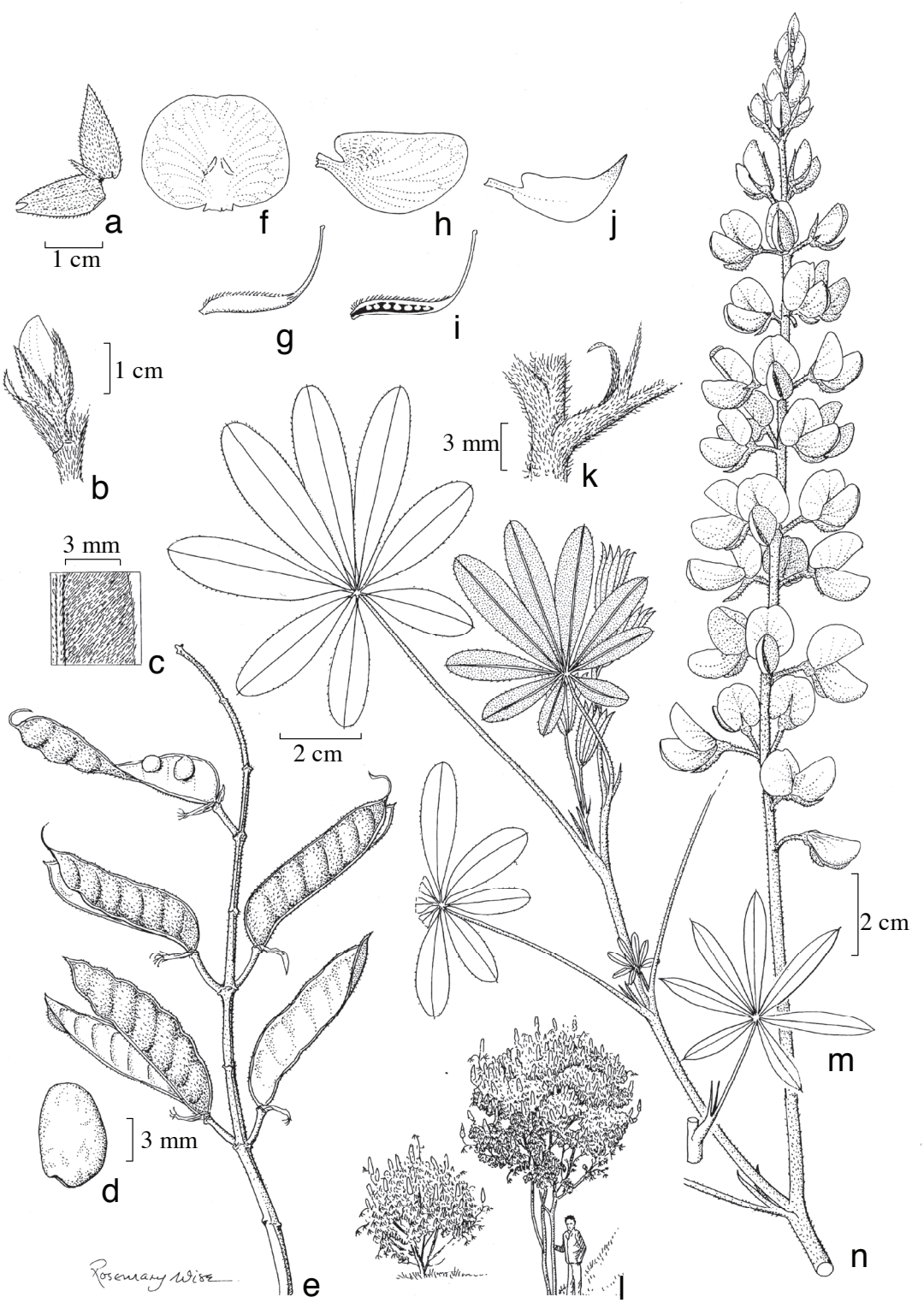


Fig. A3.1 *Lupinus semperflorens* Hartw. ex. Benth.
a. Calyx, b. Bud and floral bract, c.Detail of abaxial leaflet surface, d. Seed, e. Mature infructescence, f. Banner petal, g. Gynoecium, h. Wing petal, i. Longitudinal section of gynoecium, j. Keel petal, k. Stipule, l. Habit, m. Leaf, n. Flowering shoot. a-j, l from *CEH2032*, k. *Lewis 3775*, and m. *CEH2024*. a.-l. & n. *L. semperflorens* var. semperflorens, m. *L. semperflorens* var. *amazonicus*. Illustration by Rosemary Wise.

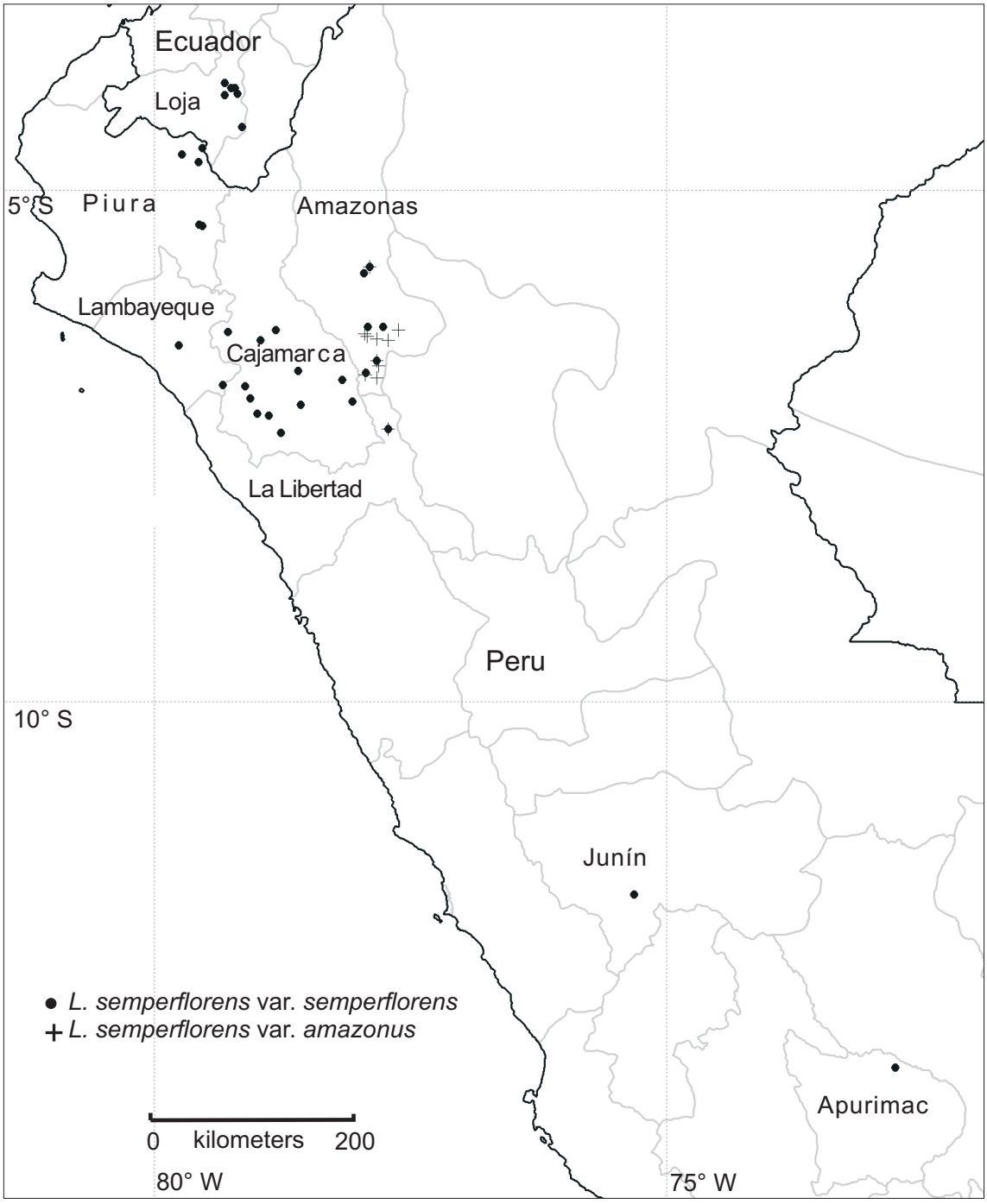


Fig. A3.2 Distribution of *L. semperflorens* Hartw. ex Benth.
This species is found at an altitudinal range of 1750 - 3600 m

Fig. A3.3 ***L. semperflorens*** **var.** ***semperflorens*** a. Habit, CEH2032, b. Seedling of *CEH2030*, c. Seeds, *RJE201*, d. Inflorescence, *RJE219*, e. Mature pods, *CEH2030*. Scale bar b. and d. 1 cm, c. 5 mm and e. 2cm. No photographs of var. *amazonus* were available. Photographs a. CEH and b-e. RJE.

4 Lupinus ellsworthianus C.P. Smith Sp. Lupinorum 159 (1940). – Type: PERU. Junín, Carpapata, above Huacapistana, 11°15' S, 75°33' W alt. 2400 m, 7 Jun 1929, *Killip EP (with Smith AC) 24338* (holotype US!, isotypes CAS!, F!, NY!) .

Lupinus albert-smithianus C. P. Smith, Sp. Lupinorum 279 (1941). – Type: PERU. Junín, Huacapistana, 11°14' S, 75°29' W alt. 2100 m, 7 Jun 1929, *Killip EP (with Smith AC) 24196* (holotype CAS!, type fragment F!, isotypes GH!, K!, NY!, US!).

Lupinus amboensis C. P. Smith, Sp. Lupinorum 764 (1953). – Type: PERU. Huánuco, Ambo, San Rafael, 10°4' S, 76° W alt. 2600m, *Ochoa C 1109,* 20 Mar 1951 (holotype CAS!).

Lupinus carhuamayus C. P. Smith, Sp. Lupinorum 761 (1953). – Type: PERU. Junín, Km 30 Carhuamayo to Paucartambo, alt. 3700 m, 15 Mar 1951, *Ochoa C 1025* (holotype CAS!).

Lupinus carpapaticus C. P. Smith, Sp. Lupinorum 181 (1940). – Type: PERU. Junín, Carpapata, above Huacapistana, 11°15' S, 75°33' W alt. 2400 m, 7 Jun 1929 *Killip EP (with Smith AC) 24360* (holotype CAS!, isotypes F!, type fragment F!, GH!, NY!, US!).

Lupinus cesar-vargasi C. P. Smith, Sp. Lupinorum 538 (1945). – Type: PERU. Cuzco, Paucartambo, Pillahuata, 13°38' S, 71°44' W, 11 Dec 1942, *Vargas C 3017* (holotype CAS!).

Lupinus diasemus C. P. Smith, Sp. Lupinorum 629 (1948). – Type: PERU. Pasco, Between La Quinua and Salccachupan, 10°41' S, 76°13' W alt. 3660 m, May 1947, *Vargas C 6604* (holotype CAS!).

Lupinus disjunctus C. P. Smith, Sp. Lupinorum 284 (1941). – Type: PERU. Junín, Huacapistana, 11°14' S, 75°29' W alt. 1900m, *Killip EP (with Smith AC) 24237,* 7 Jul 1929 (holotype CAS!, type fragment F!, isotypes NY!, US!).

Lupinus juninensis C. P. Smith, Sp. Lupinorum 278 (1941). – Type: PERU. Pasco, Huancaycc, Huariaca, 10°25' S, 76°07' W alt. 2850 m, 3 Apr 1923, *MacBride JF (with Featherstone) 3119* (holotype US!, isotypes F!, G!).

Lupinus maleopinatus C. P. Smith, Sp. Lupinorum 595 (1948). – Type: PERU. Cuzco, Paucartambo, Huailabamba, alt. 3700 m, 6 May 1939, *Balls EK 6735B* (holotype CAS!, isotypes BM!, UC!).

Lupinus pachitensis C. P. Smith, Sp. Lupinorum 765 (1953). – Type: PERU. Huánuco, Pachitea, on rd from Huánuco to Panano, ca. km 46, alt. 3000 m, 19 Mar 1951, *Ochoa C 1106,* (holotype CAS!, isotype MOL!).

Lupinus paucartambensis C. P. Smith, Sp. Lupinorum 542 (1945). – Type: PERU. Cuzco, Paucartambo, Quebrada Llulluchayoc, alt. 3350 m, 4 Jun 1944, *Vargas C 4315* (holotype CAS!).

Lupinus quellomayus C. P. Smith, Sp. Lupinorum 592 (1948). – Type: PERU. Cuzco, Convención, Tanamanche, Quellomayo, alt .3700 m, 25 Jul 1944, *Vargas C 4449* (holotype CAS!).

Lupinus semiprostratus C. P. Smith, Sp. Lupinorum 244 (1941). – Type: PERU. Cuzco, Paucartambo, Huailabamba, alt. 3700 m, 6 May 1939, *Balls EK 6735* (holotype US!, isotypes E!, F!, K!, NY!).

Lupinus tarmaensis C. P. Smith, Sp. Lupinorum 767 (1953). – Type: PERU. Junín, Valley of Chanchamayo between Carpapata and Tarma, 11°15' S, 75°33' W, Mar 1947, *Ochoa C 71* (holotype CAS!, isotype MOL!).

Sprawling or erect annual or short-lived perennial herb or shrublet branching from base (0.2-) 0.6-1.0 (-1.8) m in height. Stems hollow, rarely pithy, commonly slightly woody at base with dense white pubescence or short silky indumentum. Stipules (3-) 5-10 (-12) mm long and 0.25-1 (-1.5) mm wide, adnate region linear, adnate length 2-3 mm ~60% of length, free region lanceolate often reflexed, persistent, densely pubescent. Leaves palmately digitate, (20-) 35-100 (-120) mm long with 5-7 leaflets per leaf; petioles (10-) 20-60 (-80) mm long and 0.25-0.5 (-1) mm in diameter, densely pubescent or silky; leaflets (8-) 15-40 (-60) mm long and (2-) 5-10 (-15) mm wide although variation in leaflet size is seen in each leaf, leaflet apex mucronate, blade narrow-elliptic, or narrow-oblanceolate, base attenuate, adaxial surface glabrous or with white appressed hairs, abaxial surface densely subsericeus, commonly discolorous when dry. Erect terminal inflorescence (10-) 15-30 cm tall held predominantly above foliage, racemes markedly or loosely verticillate with 20-35 (-50) flowers in 5-9 (-12) whorls each with 3-5 flowers; peduncle diameter 1.5-3 mm, pubescent; pedicels at anthesis 4-5 mm long, 0.25-0.5 mm in diameter pubescent; floral bracts 8-9 mm long and 0.5-1 mm wide, lanceolate, pubescent, present in immature bud but not persisting at anthesis; bracts 0.5-1 mm long and 0.5-0.8 mm wide, lanceolate, pubescent, deciduous; calyx upper lip 6-7 mm long, 2.5 mm high, bidentate the notch 2-3 mm, lower lip 8-9 mm long, 1.5 mm high, concave, pubescent; banner petal orbicular (6-) 9-11 (-14) mm high by (8-) 12-16 mm wide, common flower colour bright purple-blue with central white stripe and bright yellow spot, both stripe and spot darken as flower fades, glabrous or with sparse cover of short white silky hairs along centre of banner; wing petals 9-13 (-17) mm wide and 5-8 mm high, lobe height 1.5-2 mm, claw length 1-2 mm, united at apex, bright purple-blue, glabrous with 5-7 rows of corrugated petal sculpturing at base near lobe; keel petals 10.5-11.5 (-16) mm long, 4-5 mm high, cream with dark blue/purple apical region, glabrous; floral parts persistent through to pod maturity. Occasional pink or white-flowered individuals are seen. Pods 2-8 (-30) on a slender hollow pithy infructscence, (20-) 30-45 mm long, 7-9 (-13) mm wide and 3.0-4.0 mm thick, laterally compressed, narrow-oblong, apex mucronate occasionally with persistent style asymmetrical towards ventral suture, base cuneate, dorsal suture undulate, ventral suture entire, 0.25 mm thickening surrounding all margins, lateral constrictions visible, venation often visible and prominent near ventral suture, immature pods mid-green often tinged inky blue along sutures and particularly along dorsal suture, mature pods sandy to dark brown, immature pods densely covered in white/grey pubescence or silky hairs, mature pods pubescent, 3-5 (-7) seeds per pod, dehiscent. Seeds in outline broadly oblong, 4-5 mm long, 3-4 mm wide, and 2-2.5 mm thick, marbled creams and browns with faint eyebrow, testa smooth to touch. Chromosome number unknown. Figs. A4.1 and A4.3.

Phenology. This species has been collected in flower and fruit all year round but the peak flowering season is March-May.

Distribution. This species is found in disturbed moist or very moist evergreen montane forest and secondary vegetation and roadsides on eastern slopes of the Andes in Peru (Departments: HAV, HUC, JU, PA, PU, SM) and northern Bolivia (Department: LP) (Fig. A4.2). It has an altitudinal range of 1700–3750 m. Commonly associated species are bromeliads, Ferns, *Lantana*, *Nicotiana, Pampus*, *Salvia, Verbascum.*

Additional specimens examined. **BOLIVIA,** Cochabamba, Carrasco, 8.7 km E of Epizana on rd from Cochabamba to Santa Cruz, 17°4' S 65°09' W, alt. 3200 m, 4 Feb 1987, *Solomon J C 15924* (G, NY);

Chapare, N slopes of Cerro San Benito, 17°31' S 66°55' W alt. 3400 m, 18 Mar 1929, *Steinbach J 9668* (E, F, GH, NY); Sacaba, Cadelaria, alt. 3300 m, 7 Feb 1980, *Toledo W L-5* (US!); La Paz, Franz Tamayo, Trail between Pelechuco and Pata, 14°46' S 69°01' W alt. 3150 m, 16 Nov 1988, *Lewis M 881716* (G, NY); Franz Tamayo, Pelechuco, 14°48' S, 69°05' W alt. 3150 m, 2 Mar 1980, *Feuerer T 9082b* (NY); Franz Tamayo, Pelechuco, Descending path of Río Abajo at the foot of the Río Pelechuco, 14°46' S 69°20' W alt. 3500 m, 19 Oct 1999, *Beck S G 25223* (FHO); Pelechuco, 14°46' S 69°20' W May 1960, Cardenas M 5814 (K, US);

PERU, Muna, 26 Jun 1923, *MacBride JF 3949* (F, G, US); **Cuzco,** Rd from Puerto Maldonado to Urcos, c. 20km after Marcapata, 13°35'23" S, 70°58'20" W alt. 3000 m, 7 Oct 1987, *Khan F 2287* (NY, USM); Pillahuata, Cerro de Cusilluyoc, 13°08' S, 71°25' W alt. 3150 m, 4 May 1925, *Pennell FW 14113* (GH); Machu Pichu, Ceja de la Montaña, 13°09' S, 72°30'28" W alt. 2000 m, 21 Apr 1954, *Rauh & Hirsch P888* (NY); Above Marcapata, 13°35'23" S, 70°58'20" W alt. 3280 m, 14 Oct 1976, *Wasshausen DC 805* (K, NY, US, USM); Between Calca and Cuzco, 13°20' S, 71°57' W alt. 2000 m, 7 Apr 1983, *Hoogte LD 982* (F); Convención, Amaybamba. Jan 1948, *Marin 864*. (F); La Convención, Vilcabamba, 13°07'27" S, 72°58'55" W alt. 3400 m, 31 May 2002, *Galiano W 4157* (FHO, MO); Paucartambo, Acjanaco, Parque Nacional Manú, 9 May 1990, *Cano A 3614* (USM); Paucartambo, Altura de Telaban, Parque Nacional Manú, 16 Jul 1990, *Cano A 3778* (USM); Paucartambo, Alturas de Pucara, Parque Nacional Manú. 6 Sep 1990, *Cano A 4097* (F, USM); Paucartambo, Qollatamba, Parque Nacional Manú, 11 Sep 1990, *Cano A 4326* (F, USM); Paucartambo, Tres Cruces, Manú National Park, 13°07' S, 71°05' W alt. 3650 m, 6 Mar 1991, *Cano A 4618* (NY, USM); Paucartambo, Callanga, 15 Feb 1953, *Woytkowski F 374* (USM); **Huánuco,** Carpish Pass, km 84 rd from Tingo Maria to Lima, 9°45' S, 76°08' W alt. 2750 m, 19 Nov 1949, *Allard HA 20958* (US); 29-35 km NW of Hacienda Mitoambo, rd from Huánuco to La Union, 9°55' S, 76°43' W alt. 3100 m, 2 Mar 1978, *Luteyn JL 5505* (CAS, F, NY, USM); Vicinity of Piedras Grandes. 30 Nov 1937, *Woytkowski F 138* (F); Acomayo, Rocky hills W of Acomayo, 9°46' S, 76°05' W alt. 2400 m, 22 Apr 1946, *Woytkowski F 34237* (F, G, UC); Huancapallac. Km 28 rd W from Huánuco towards La Union, 9°53'31" S, 76°27'46" W alt. 3220 m, 30 Apr 2002, *Hughes CE 2246* (FHO, K, MOL); **Junín,** Carpapata, above Huacapistana, 11°15' S, 75°33' W alt. 2400 m, 7 Jun 1929, *Killip EP 24360* (CAS, F, GH, NY, US); Huacapistana. 11°14' S, 75°29' W alt. 1828 m, Oct 1943, *Sandeman C 4381* (OXF); Tarma, Palca, 9.3 km NE of Palca, rd from Tarma to San Ramón, above the Rio Huashahuasi, nr bridge nr branch in the road, 11°15' S, 75°36' W alt. 2600 m, 31 Mar 1984, *Croat TB 57660* (FHO); Tarma, Carparpata. Turning for this at km 64 on the rd from Tarma to San Ramón, 11°15'14" S, 75°33' W alt. 2400 m, 28 Jun 2004, *Eastwood RJ 158* (FHO, K, MOL, USM); Tarma, Vicinity of Palca, rd from Tarma to Chanchamayo, 11°15' S, 75°36' W 11 Oct 1863, *Isern J 2448* (F); Tarma, Carpapata, descent from Tarma to San Ramón in the Chanchamayo Valley, 11°15' S, 75°33' W alt. 2400 m, 6 Mar 1948 *Ochoa C 295* (US, USM); Tarma, Carpapata, descending from San Ramón, 11°15' S, 75°33' W alt. 2020 m, 12 Feb 1953, *Ochoa C 2022* (CAS, GH, US, USM); Tarma, Below Carpapata, rd from Tarma to San Ramón, 11°15' S, 75°33' W alt. 1700 m, 16 Mar 1976, *Plowman T 5646* (F, GH); Tarma, Carpapata, 11°15' S, 75°33' W alt. 2230 m, 29 Jan 1983, *Smith DN 3145* (ISC, USM); Tarma, Carpapata, rd from Tarma to La Merced, 11°16' S, 75°33' W alt. 2350 m, 11 Jan 1984, *Smith DN 5720* (ISC, K, USM); Tarma, Rd between Huacaspitana and Palca, 11°27'10" S, 75°40'59" W alt. 1800 m, *Weberbauer A 1965* (G); Pasco, Between Saloachupan and Cerro de Paso, 31 Jan 1950, *Ferreyra R 6624* (MOL, US, USM); La Quinua, 10°36' S, 76°10' W alt. 3700 m, 14 May 1922, *MacBride JF 2017* (F, G, US); Huancaycc, Nr Huariaca, climbing by dirt track from Huariaca towards the ruins of Yarush, 10°25'56" S, 76°07'31" W, 20 Mar 1951, *Ochoa C 1110* (CAS); Vicinity of Yanamchay, between Cerro de Pasco and La Quinua, 10°38' S, 76°11' W alt. 3600 m, 6 Apr 1948, *Ochoa C 474* (CAS, F); **Puno,** Between Sandía and Cuyocuyo, 14°14'43" S, 69°25'52" W alt. 3150 m, 16 May 1966, *Ferreyra R 16783* (USM); Carabaya, Below Ollachea, 13°45' S, 70°30' W alt. 2600 m, 30 Dec 1947, *Vargas C 6932* (CAS); **San Martín,** Mariscal Caceres, Forest patch (C9) above timberline, Rio Abiseo National Park, 12°34' S, 74°57' W alt. 3425 m, 24 Nov 1985, *Young K 2570* (F).

Notes. The key distinguishing feature of this species is the small 'button-like' flower (Figs. A4.1 and A4.3), which is a function of the narrow angle between the wing and the banner petal.

The large majority of specimens of *L. ellsworthianus* are distributed around two geographical locations, Carpapata (Junin) and Paucartambo (Cuzco), where numerous collections have been made due to accessability. No subspecies or varieties are recognised within this species

due to the absence of fixed and discrete characters but some variation can be recognised. The Paucartambo plants tend to be either larger in stature, less woody and with larger leaves or much smaller and slender in stature. Questions remain about the placement of some additional specimens (holotype of *L. sandiensis* Vargas 1295 (CAS), holotype of *L. volubilis* Vargas 6645 (CAS), Meltcalf 30454 (CAS, G, UC, US), Cano 5194 (USM), Cano 3281 (USM), Cano 3322 (USM), Cano 3613 (USM), Macbride 4079 (F)) from Paucartambo. These collections are much smaller in height, herbaceous, pithy, with erect silky hairs and longer narrow stipules but show similarities in flower shape and size and distribution. These collections may represent young material of *L. ellsworthianus* but are currently excluded.

Lupinus piurensis var. *cotumazicus* is similar in morphology to *L. ellsworthianus*. It however has larger flowers sometimes with ciliate keel petals, a more erect habit and a narrow obovate leaf shape.

The extensive synonymy associated this species exemplifies the chaotic state of Andean *Lupinus* taxonomy particularly as perpetrated by the copious naming of new species carried out by C. P. Smith. Fourteen names described by Smith are here treated as conspecific with *L. ellsworthianus*. Smith does acknowledge the similarity between three of these 14 species. Notes in the protologue of *L. carhuamayus* recognise a similarity to the type of *L. diasemus* (Vargas 6604) and *L. juninensis* but Smith (1953) concluded they are distinct but closely related. Confidence can not be had in Smith's names particularly when different sheets of the same collection (Balls 6735) have been designated as two different species (*L. maleopinatus* and *L. semiprostratus*). The name maleopinatus derives from *'badly judged'* as in Smith's opinion misjudgement caused the *"inclusion of such a different plant, under one and the same number"*. However, it is not obvious which duplicates of *Balls 6735* Smith thought was so different. Those sheets with little material clearly only have one shoot whereas the sheets with more material have a larger branching shoot. Observation of the type specimens also shows that many of the type localities are either identical or very close together (e.g. the types of *L. albert-smithianus, L. carpapaticus, L. disjunctus, L. ellsworthianus,* and *L. tarmaensis* are all from near to Carpapata on the main road from Tarma to Chanchamayo, a frequent area for many collectors). Although some sympatry has been observed, such high number of similar species in sympatry adds further doubt about Smith's delimitations.

In effect many of Smith's species, and especially his prolific naming of species based on the collections of Carlos Ochoa and Cesar Vargas simply amount to types. The first name that was applied to this species published by Smith was *L. elsworthianus.* Macbride (1943) commented this was "no doubt by typographical error" since it was named after the collector Ellsworth P. Killip and therefore corrected it to be spelt as ellsworthianus. This is in line with article 60.1 of the ICBN (Greuter & International Association for Plant Taxonomy, 2000).

A further type specimen which shows similarity in flower size and shape to *L. ellsworthianus* is that of *L. praealtus* (Macbride 2144). It has been not been included as a synonym of *L. ellsworthianus* since it is described as a tree to three metres (Macbride, 1943) and the indumentum, present on all vegetative parts, is far denser.

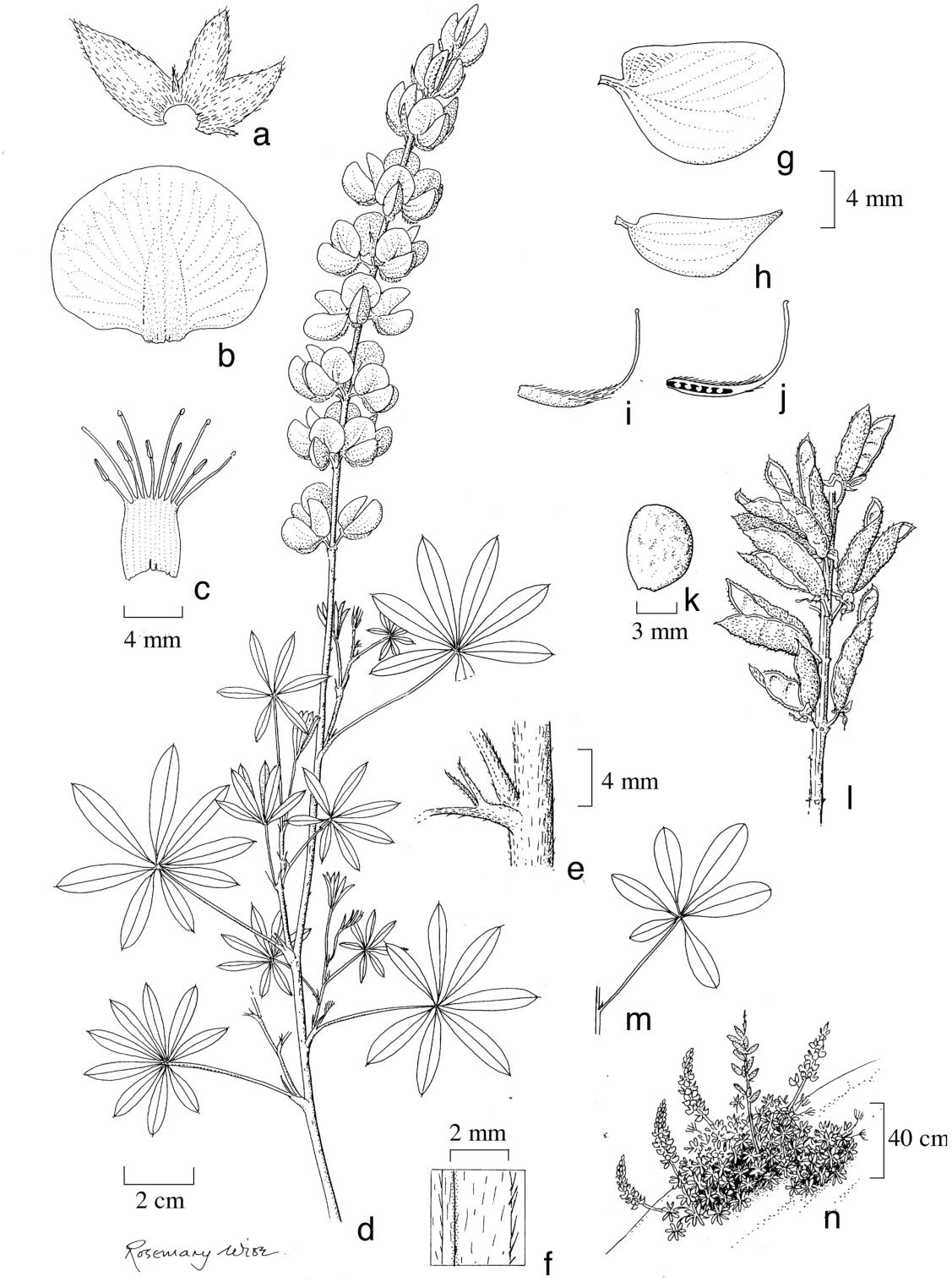


Fig. A4.1 *Lupinus ellsworthianus* C. P. Smith
a. Calyx, b. Banner petal, c. Androecium, d. Flowering shoot, e. Stipule f. Detail of abaxial leaflet surface, g. Wing petal, h. Keel petal, i. Gynoecium, j. Longitudinal section of gynoecium, k. Seed, l. Mature infructescence, m. Leaflet shape variation, n. Habit. A.-l. and n. based on *RJE158,* m. *Cano A 4326* . Illustration by Rosemary Wise.

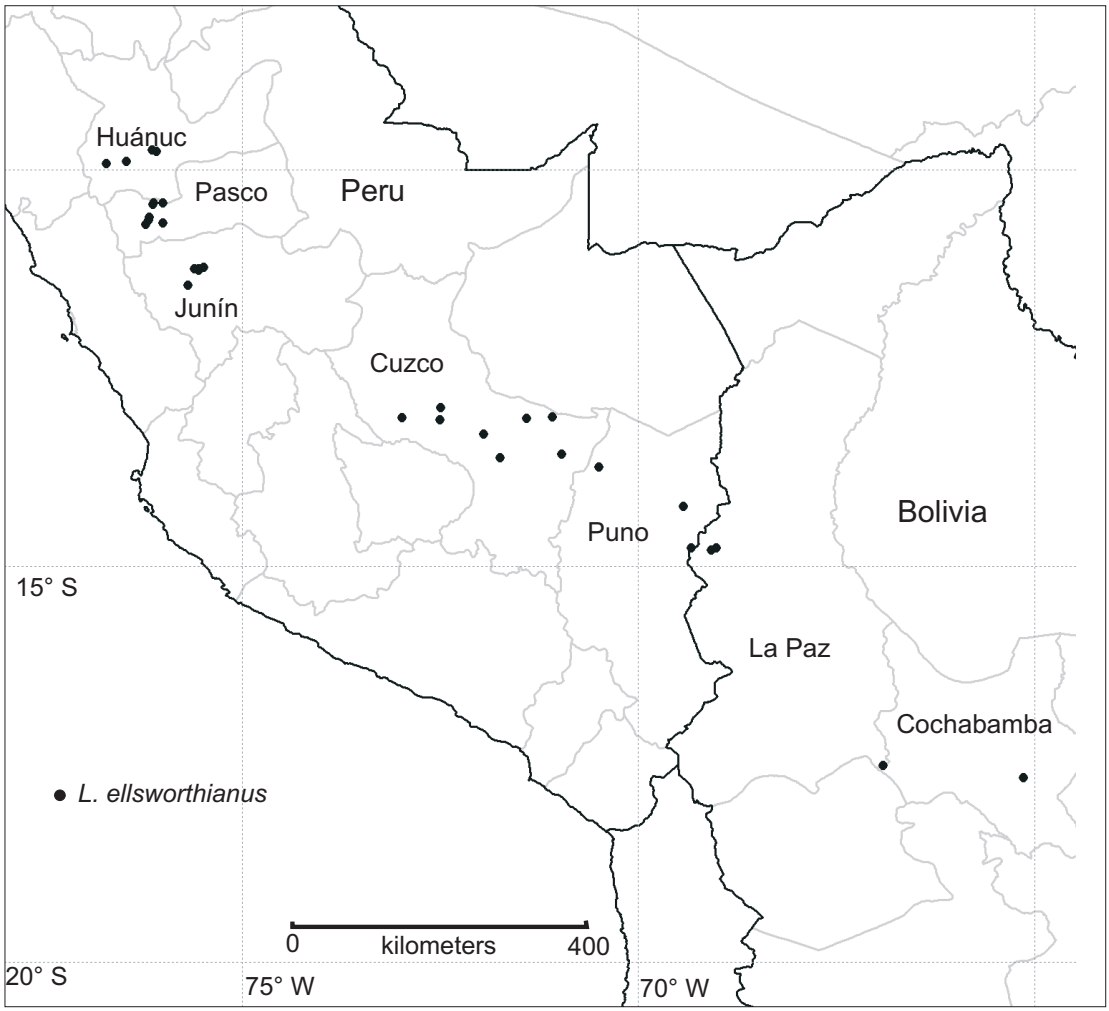


Fig. A4.2 Distribution of *L. ellsworthianus* C.P. Smith
This species is found at an altitudinal range of 1700-3750 m.

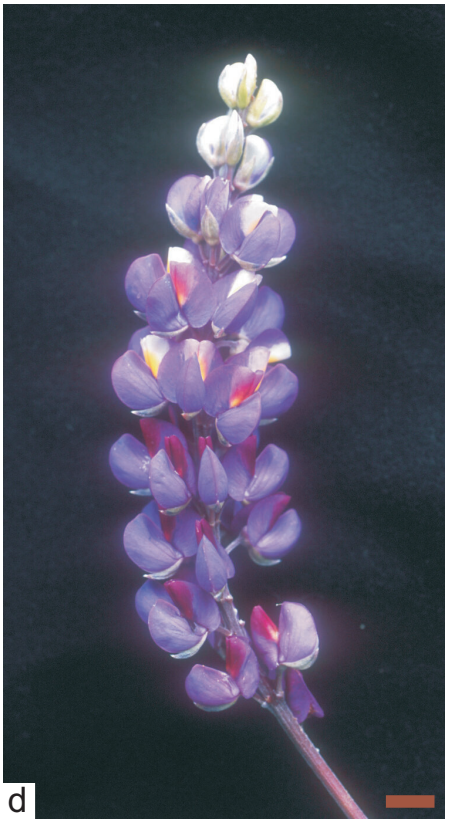


Fig. A4.3 ***L. ellsworthianus***
a. Mature and immature pods, b. Habit, c. Habit, d. Infloresence. All from *RJE158*. Scale bars a. 3 cm, b. 20 cm, c. 10 cm and d. 1 cm. Photographs CEH.

5 Lupinus praestabilis C.P. Sm. Sp. Lupinorum 587 (1948). – Type: PERU. Cuzco, Calca, Pisac, Amaru, 13° 27' S 71° 45' W alt. 3500 m, 19 Mar 1945, *Vargas C 5052* (holotype: CAS!).

Lupinus proculaustrinus C. P. Sm. Sp. Lupinorum 607 (1948). – Type: PERU. Tacna, Tarata, Chuvire, 1km from Challaquaya, 17°17' S 70°01' W alt. 2500 m, Mar 1926, *Metcalf RD 3040* (holotype: CAS!, isotypes: G!, GH!, UC!, US!).

Perennial shrub or treelet with one principal stem 1.5-3 m in height. Shoots hollow, stems woody at base, glabrous, young stems glaucous and often tinged red. Stipules 40-70 mm long and 3-6 mm wide at base, adnate length 30-65 mm (75-) 80-90% of length and sheathing petiole, adnate region linear, free region acute or attenuate, glabrous, becoming papery with age, persistent even after leaflets fall. Leaves palmately digitate, (80-) 110-200 (-240) mm in length with (7-) 8-12 leaflets per leaf; petioles (45-) 60-120 (-170) mm in length and (0.5-) 1-2 (-3) mm in diameter, glabrous; leaflets (25-) 40-70 (-90) mm long and 6-14 mm wide, leaflet apex acute, mucronate or acuminate, blade narrow-elliptic or lorate, base attenuate, immature leaflets conduplicate, adaxial surface glabrous, abaxial surface predominately glabrous but with straight white hairs 1-2 (-3) mm restricted to midrib and margins. Erect terminal inflorescence 30-40 cm tall held above foliage, racemes markedly or loosely verticillate with 40-50 (-60) flowers in 10-11 whorls each with (3-) 4-6 flowers; peduncle diameter 3-4 (-5) mm glabrous; pedicels at anthesis 6-9 (-11) mm long, 0.5-1 mm in diameter with a dense indumentum of white silky hairs; floral bracts (9-) 15-20 mm long, (1-) 3-5 mm wide, lanceolate moderately-densely covered in white silky hairs, deciduous shortly after anthesis; bracteoles (3-) 5-6 (-8) mm long, 1-2 mm wide, lanceolate, glabrous; calyx upper lip 8-11 mm long, 1-3 (-4) mm high, bidentate the notch 1-2 mm, lower lip 9-12 (-14) mm long, 1-3 mm high, concave, green, moderately-densely covered in silky white hairs; banner petal orbicular 14-16 mm high, 15-19 mm wide, the common colour form bright blue with a bright yellow spot which fades to brown, glabrous; wing petals (11-) 15-17 mm wide, 8-10 mm high, lobe height 2.5-3.5 mm, claw length 1.5-2.5 mm, bright blue, glabrous with 5-9 rows of corrugated petal sculpturing at base near lobe; keel petals 12.5-14 mm long, 5-6 mm high, white/cream with blue apical region, margin of apex and basal half glabrous, sub-apical region ciliate, often minutely so; all floral parts frequently persistent through to pod maturity, occasional white-flowered individuals are found. Pods held more or less horizontal to stem, on tough peduncles, (11-) 20-30 pods per infructescence, 40-50 mm long, 7-9 mm wide and 3-4 mm thick, laterally compressed narrow-oblong, apex mucronate, occasionally with a persistent style, base cuneate, dorsal suture undulate, ventral suture entire, 0.5 mm thickening surrounding sutures, lateral constrictions visible, venation sometimes visible near ventral suture, immature pods green tinged inky blue, mature pods light brown tinged inky blue at sutures, pods densely covered in white silky hairs, 4-6 (-7) seeds per pod, dehiscent. Seeds broadly oblong 4-5 mm long, 6-7 mm wide and 2-3 mm thick, ca. 0.52 g weight, marbled brown, cream and grey, testa smooth to touch. Cotyledons petiolar. First eophylls with 3 leaflets. Chromosome number 2n=48 (unpublished data, Maria Schifino-Wittmann). (Fig. A5.1 and A5.3).

Phenology. This species can be found in flower and fruit all year but the peak flowering season is June-July.

Distribution. *Lupinus praestabilis* is distributed in southern Peru (Departments: AR, CU, LI, MO, TA) and north Bolivia (Department: LP) in a restricted altitudinal zone between 2500-3500m (Fig. A5.2). It is found in the transitional dry thorn scrub between low desert scrub and high altitude scrub and dry *Polylepis* forest and also in moister seasonal stream beds in this zone. Associated vegetation comprises of *Baccharis, Barnadesia* and other shrubby Asteraceae species, *Adesmia, Calceolaria, Polylepis* and occasionally members of the Cactaceae.

Additional specimens examined. **BOLIVIA,** La Paz, Larecaja, Vicinity of Sorata 15°50' S, 68°38'16" W alt. 3300 m, Apr 1860, *Mandon G 681* (BM, G, K, NY);

PERU, Arequipa, 25 km N of Arequipa, old rd to Chivay above Charcani Chico about 5 km above turnoff to hydroelectric plant on the SE flanks of Nevado Chachani 16°16'07" S, 71°30'26" W alt. 3450 m, 14 Nov 2003, *Hughes CE 2353* (FHO, K, MOL); **Lima,** Matucana, San Damian, 12°01' S 76°23' W alt. 2400-3000 m, -Feb 1913, *Hrdlicka A* Jan (F – photo only); Yauyos, Atjsacki, hill E of Tupe 12°45'40" S, 75°49'30" W alt. 2930 m, 4 Jan 1952, *Cerrate E 1050* (MOL, USM); Yauyos, 2 km SE of Tupe 12°45'40" S, 75°49'30" W alt. 2780 m, 23 Jun 2004, *Eastwood RJ 142* (FHO, K, MOL, USM); Yauyos, S-facing slope, 1 km NW of Tupe 12°44'29" S, 75°47'50" W alt. 2950 m, 23 Jun 2004, *Eastwood RJ 143* (FHO, K, MOL, USM); **Moquegua,** Comas, Rd from Arequipa to Puquina, in valley immediately before Puqina 16°36'36" S, 71°11'48" W alt. 3100 m, 29 Apr 2000, *Weigend M 2000/581* (M); **Tacna,** Tarata, Ticalaco 17°27' S, 70°03' W, 31 Mar 1998, *Cano A 8319* (USM); Tarata, In Chuvire, 1 km from Challaquaya 17°17' S, 70°01' W alt. 2500 m, Apr 1926, *Metcalf RD 30409* (CAS, G, GH, K, UC, US); Tarata, Between Ticalaco and Tarata 17°17' S, 70°01' W alt. 3150 m, 22 Mar 1953, *Ochoa C 2038* (MOL).

Notes. *Lupinus praestabilis* is the only large-stipuled Andean species. When first collected, material (*Hrdlicka s.n.* 1913) was identified as *L. montanus* var. *austrovolcanicus* from central America (Macbride, 1943; Smith, 1938, 1941). Later, two separate large-stipuled Andean species were described by C. P. Smith (1948a; 1948b). *Lupinus praestabilis* has noticeably larger flowers and longer floral bracts than *L. montanus* and does not exhibit an inflated infrutescence stem. The disjunct distribution of large-stipuled species between Mexico and central America and the Andes was hypothesised to be a result of long distance dispersal (Dunn & Harmon, 1977). However, molecular phylogeny reconstruction sampling both *L. praestabilis* and the large-stipuled Mexican species *L. montanus* show that these two species are not closely related and that large stipules have apparently evolved independently in Mexico and the Andes (Chapter 3).

Given the conspicuously large stipules in this species it is surprising that there are so few collections. Particularly there is little mature pod material. Since it is so distinctive it is also strange that both *Weigend 2000/58* and *Mandon 681* are apparently mixed collections. The Munich specimen of *Weigend 2000/58* is *L. praestabilis* whilst the NY and USM specimens are *L. mutabilis.* It appears that Mandon grouped material he thought were the same thing under one collecting number and this may explain this mixed collection.

Lupinus praestabilis has been discussed as a possible progenitor for *L. mutabilis* (Blanco Galdes, 1982, 1984; 1986; Tapia, 1982). The mixed *L. praestabilis/ L. mutabilis* collection of *Mandon 681* is annotated as cultivated. I think it is most likely that this refers to *L. mutabilis* not *L. praestabilis*. Given the paucity of collections of this species overall the lack of further Bolivian collections, even in Sorata which is well-collected, does not suggest to me that *L. praestabilis* was in cultivation. This is discussed further but dismissed in the account of *L. mutabilis*. If the Bolivian material was cultivated this species would be endemic to Peru and largely restricted to the western slopes yet with an outlier in Cuzco.

Types and nomenclature. Examination of the protologue and type of *L. praestabilis* C.P.Sm. shows that it differs from *L. proculaustrinus* in flower colour and some quantitative characters. Smith stated that *L. praestabilis* is "certainly well removed from *L. montanus*" stating a number of size differences. However, he did not discuss why he recognised *L. proculaustrinus* as a separate species except explaining the name *proculaustrinus* translates into *"far off, southern"* recognising the disjunction from Mexican large-stipuled species. I do not find the *L. praestabilis* type to be sufficiently distinct from the *L. proculaustrinus* type given the morphology of the collections sharing the large stipule character and the sporadic occurrence of white and pink individuals in many species.

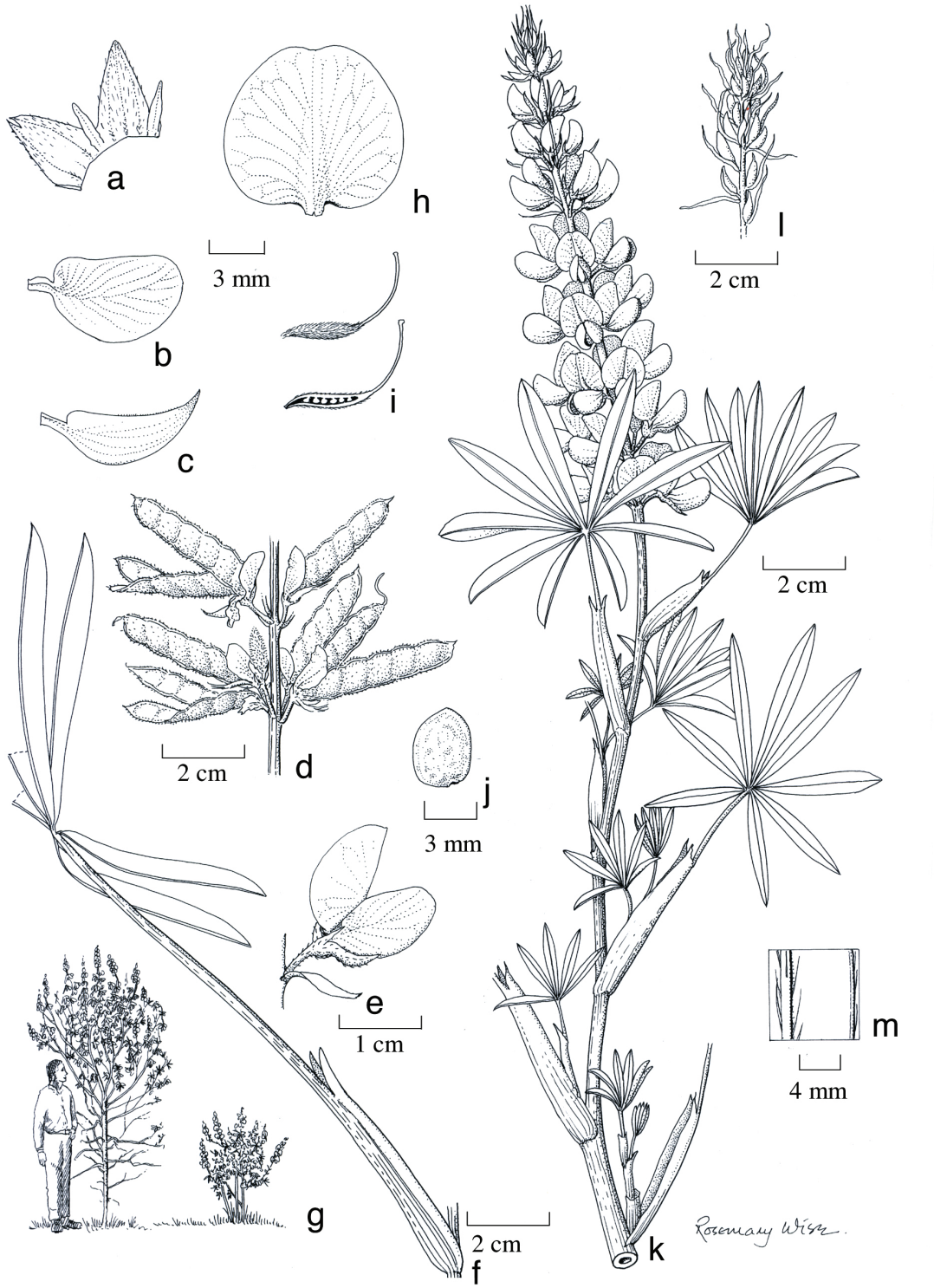


Fig. A5.1 *Lupinus praestabilis* C. P. Smith
a. Calyx, b. Wing petal, c. Keel petal, d. Infructescence, e. Flower with floral bract, f. Leaf, g. Habit, h. Banner petal, i. Gynoecium, j. Seed, k. Flowering shoot, l. Immature inflorescence showing floral bracts, m. Detail of abaxial leaflet surface a-c, e, g, i-l. & m. from *CEH2353*, d. *RJE134,* f. *Metcalf 30409,* j. *RJE142* and l. *Weigend 2000/581*. Illustration by Rosemary Wise.

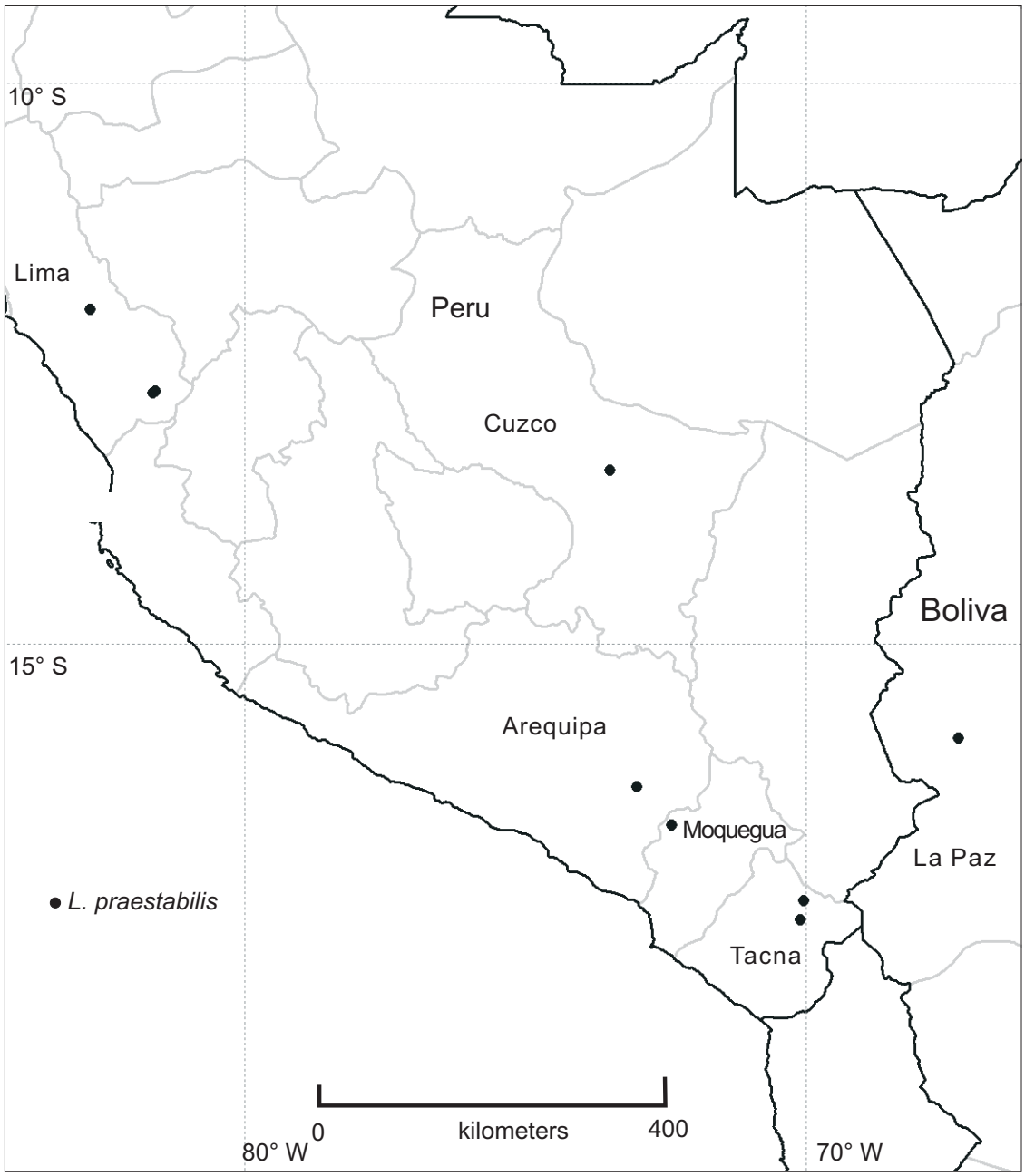


Fig. A5.2 Distribution of *L. praestabilis* C. P. Smith
This species is found at an altitudinal range of 2500 - 3500 m. There has been some suggestion that the Bolivian material is cultivated (*Mandon 681*). However, since this collection is mixed with cultivated *L. mutabilis* it is more likely this annotation refers to the latter species.

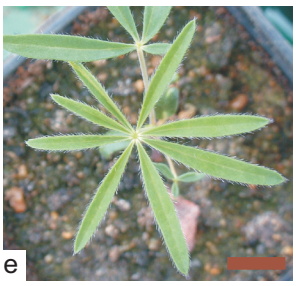


Fig. A5.3 ***Lupinus praestabilis***
a. & b. Habit - treelet and shrub, *CEH2353*, c. Fertile shoot, *RJE142*, d. Glaucous stem, seedling of *RJE142*, e. Leaf, seedling of *RJE142*, f. Stipule, *RJE143*. Scale bar b. 20 cm, c. 2 cm, d-f. 5 mm. Photographs a-b, f. CEH and c-e. RJE.

References for taxonomic account

(1999) Lupin An ancient crop for a new millienia. In *10th International Lupin Conference*, Vol. 1, Germany.

Ainouche, A. & Bayer, R. (1999) Phylogenetic relationships in *Lupinus* (Fabaceae: Papiliondeae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. *American Journal of Botany*, **86**, 590-607.

Ainouche, A., Bayer, R.J., & Misset, M.-T. (2004) Molecular phylogeny, diversification and character evolution in *Lupinus* (Fabaceae) with special attention to Mediterranean and African lupines. *Plant Systematics and Evolution*, **246**, 211-222.

Antúnez de Mayolo, S. (1982) Tarwi in Ancient Peru. *Proceedings of the First International Lupine Conference*, 1-12.

Bateman, R.M. (1999). Integrating molecular and morphological evidence of evolutionary radiations. In *Molecular Systematics and Plant Evolution* (eds P.M. Hollingsworth, R.M. Bateman & R.J. Gornall), Vol. 1, pp. 432-471. Taylor and Francis, London.

Blanco Galdes, O. (1982) Genetic variability of Tarwi (*Lupinus mutabilis* Sweet). *Proceedings of the First International Lupine Conference*, 33-50.

Blanco Galdes, O. (1984) *Lupinus mutabilis* germplasm resources in South America. In 3rs, pp. 159-164.

Blanco Galdes, O. (1986) Wild Stock of *Lupinus mutabilis* Sweet. In *4th International Lupin Conference*, Vol. 1, pp. 284, Geraldton, Western Australia.

Chávez, E. & Untied, P. (1982) Program for Lupine, Tarwi or Chocho production in the Peruvian highlands: Outlook and limitations. *Proceedings of the First International Lupine Conference*, 263-280.

Conterato, I.F. & Schifino-Wittmann, M.T. (2006) New chromosome numbers, meiotic behaviour and pollen ferility in American taxa of *Lupinus* (Leguminosae): contributions to taxonomic and evolutionary studies. *Botanical Journal of the Linnean Society*, **150**, 229-240.

Cowling, W.A., Buirchell, B., & Tapia, M. (1998) *Lupin Lupinus L.* IPGRI.

Darlington, C.D. & Wylie, A.P. (1955) *Chromosome atlas of flowering plants*, 2nd edn. Allen & Unwin, London.

Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, **418**, 700-707.

Dunn, D., B. & Harmon, W.E. (1977) The *Lupinus montanus* complex of Mexico and Central America. *Annals of Missouri Botanic Garden*, **64**, 340-365.

Evans, L.T. (1993) *Crop evolution, adaptation and yield* Cambridge University Press, Cambridge.

Frey, F. (1999) Preliminary observation of diseases and pests of lupinus (*L. mutabilis* Sweet) in Peru and Bolivia. *Lupin, an Ancient Crop for the New Millenium. Proceedings of the 10th International Lupin Conference*, 173-186.

Greuter, W. & International Association for Plant Taxonomy (2000) *International code of botanical nomenclature (Saint Louis Code) adopted by the sixteenth International Botanical Congress, St Louis, Missouri, July-August 1999* Koeltz, Kèonigstein.

Gross, R. (1982) *El cultivo y la utilización del Lupinus mutabilis Sweet* FAO, Roma.

Hanelt, P. (1986). Lupins. In *Verzeichnis Landwirtschaftlicher und Gārtnerischer Kulturpflanzen (Ohne Zierpflanzen)* (ed J. Schultze-Motel), pp. 696-966.

Adademie-Verlag, Berlin.
Harter, A.V., Gardner, K.A., Falush, D., Lentz, D.L., Bye, R.A., & Rieseberg, L.H. (2004) Origin of extant domesticated sunflowers in eastern North America. *Nature*, **430**, 201-205.
Hooker, W.J. (1830-31) The late Mr Barclay. *Botanical Miscellany*, **2**, 122-125.
Hughes, C.E. & Eastwood, R.J. (2006) Island radiation on a continental scale: exceptional rates of species diversification following uplift of the Andes. *Proceedings of the National Academy of Sciences, USA*, **103**, 10334-10339.
Kass, E. & Wink, M. (1997) Molecular phylogeny and phylogeography of *Lupinus* (Leguminosae) inferred from nucleotide sequences of the rbcL gene and ITS 1 + 2 regions of rDNA. *Plant Systematics and Evolution*, **208**, 139-167.
Kazimierski, T. & Nowacki, E. (1961) Indigenous species of lupins reguarded as the initial forms of the cultivated species *Lupinus albus* and *Lupinus mutabilis*. *Flora*, **151**, 202-209.
Ladizinsky, G. (1979) Seed dispersal in relation to the domestication of Middle East legumes. *Economic Botany*, **33**, 284-289.
Macbride, J.F. (1943) *Flora of Peru : family Leguminosae* Field Museum of Natural History, Chicago.
Maisurian, N.A. & Atabekova, A.I. (1974) *Lupin*, Moscow.
Mujica, A. (1992). La agricultura Andina. In *Cultivos marginados otra perspectiva de 1492* (eds J.E. Hernández Bermejo & J. León), Vol. 26. FAO, Rome.
National Research Council (1989). Tarwi. In *Lost crops of the Incas, Little-known plants of the Andes with promise for worldwide cultivation* (ed NRC), pp. 181-189.
Nixon, K.C. & Wheeler, Q.D. (1990) An amplification of the phylogenetic species concept. *Cladistics*, **6**, 211-223.
Schifino-Wittmann, M.T. (2005) *Lupinus* chromosome counts. *Unpublished*.
Smartt, J. (1985) Evolution of Grain Legumes. III. Pulses in the genus Vigna. *Experimental Agriculture*, **21**, 87-100.
Smith, C., Piper (1938). Paper five, The lupines of Central America, south of Mexico. In *Species Lupinorum* (ed C.P. Smith), pp. 89-93. Privately published.
Smith, C., Piper (1941). Signature 18, Paper twenty-five. In *Species Lupinorum* (ed C.P. Smith), pp. 273-288. Privately published.
Smith, C., Piper (1944). Signature twenty-five, paper thirty-three concluded (*Lupinus* in Ecuador). In *Species Lupinorum* (ed C. Smith, Piper), pp. 385-390. privately published.
Smith, C.P. (1948a). Paper sixty-five, Peru ten, Tacna one. In *Species Lupinorum* (ed C.P. Smith), pp. 607-608. Smith, Charles. Piper.
Smith, C.P. (1948b). Paper sixty-two, Peru Seven, Cuzco Three. In *Species Lupinorum* (ed C.P. Smith), pp. 584-593. Smith, Charles. Piper.
Smith, C.P. (1953). Signature fourty-four, Paper one hundred and five. In *Species Lupinorum*, pp. 753-768.
Sweet, R. (1835). 203 Lupinus mutabilis var. Cruckshanksii. In *Sweet's British Flower Garden*, Vol. 3.
Sweet, R. & Smith, E.D. (from 1823) *The British flower garden. The drawings by E.D. Smith*, Lond.
Tapia, M. (1982) Agricultural production and research on Tarwi (*Lupinus mutabilis* Sweet) in Bolivia. *Proceedings of the First International Lupine Conference*, 819-834.
Tapia, M. & Vargas, C. (1982) Wild lupines of the Andes of southern Peru. *Proceedings of*

the First International Lupine Conference, 23-32.
Towle, M.A. (1961) *The Ethnobotany of pre-columbian Peru* Aldine Publishing Company, Chicago.
von Baer, E. & von Baer, D. (1988) Lupinus mutabilis: cultivation and breeding. In 5th International Lupin Conference, pp. 237-247.
Wolko, B. & Weeden, N.F. (1990) Isozyme number as an indication of phylogeny in *Lupinus*. *Genetica Polonica*, **31**, 179-187.
Yacovleff, E. & Herrera, F.L. (1934) El mundo vegetal de los antiguos Peruanos. *Revista del Museo Nacional (Lima)*, **3**, 241-322.

Appendix 2

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2006) **361**, 211–225
doi:10.1098/rstb.2005.1735
Published online 22 November 2005

Review

From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction

Colin E. Hughes[1,*,†], Ruth J. Eastwood[1] and C. Donovan Bailey[2,†]

[1]*Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK*
[2]*Department of Biology, New Mexico State University, PO Box 3001 Department 3AF, Las Cruces, NM 88003-8001, USA*

Phylogenetic analyses of DNA sequences have prompted spectacular progress in assembling the Tree of Life. However, progress in constructing phylogenies among closely related species, at least for plants, has been less encouraging. We show that for plants, the rapid accumulation of DNA characters at higher taxonomic levels has not been matched by conventional sequence loci at the species level, leaving a lack of well-resolved gene trees that is hindering investigations of many fundamental questions in plant evolutionary biology. The most popular approach to address this problem has been to use low-copy nuclear genes as a source of DNA sequence data. However, this has had limited success because levels of variation among nuclear intron sequences across groups of closely related species are extremely variable and generally lower than conventionally used loci, and because no universally useful low-copy nuclear DNA sequence loci have been developed. This suggests that solutions will, for the most part, be lineage-specific, prompting a move away from 'universal' gene thinking for species-level phylogenetics. The benefits and limitations of alternative approaches to locate more variable nuclear loci are discussed and the potential of anonymous non-genic nuclear loci is highlighted. Given the virtually unlimited number of loci that can be generated using these new approaches, it is clear that effective screening will be critical for efficient selection of the most informative loci. Strategies for screening are outlined.

Keywords: Tree of Life; plant phylogeny; nuclear DNA sequence loci; low-copy nuclear gene; comparative anchor tagged sequence; sequence characterized amplified region

1. INTRODUCTION

Reconstructing the Tree of Life has been a central objective of evolutionary biology since phylogenetic trees were first proposed as a way of representing evolutionary relationships. Progress towards that goal has accelerated in the last two decades and assembling the Tree of Life is now the focus of more research than at any time over the past 150 years (e.g. Soltis & Soltis 2001; Donoghue & Cracraft 2004; Palmer *et al.* 2004). This has been prompted by theoretical and methodological advances, vastly increased computational power, and technical improvements in DNA sequencing that enable the generation of large volumes of character data relatively quickly and cheaply. It is now possible to build robust hypotheses of relationships for large numbers of taxa and to contemplate the vision of a complete Tree of Known Life (Soltis & Soltis 2001; Watanabe 2002; Donoghue & Cracraft 2004). Current interest in phylogeny reconstruction has also been driven by ever wider application of phylogenetic trees within as well as beyond the confines of systematics and evolutionary biology (e.g. Soltis *et al.* 1999; Savolainen & Chase 2003; Futuyma 2004; Yates *et al.* 2004). Phylogenetic trees provide not only the basis for classification, but also studies of character evolution (Schultheis & Baldwin 1999), hybridization (e.g. Hughes *et al.* 2002; Linder & Rieseberg 2004), polyploidy (e.g. Doyle *et al.* 2003*b*, 2004), biogeography (e.g. Lavin *et al.* 2004; Pennington *et al.* 2004), origins of domestication (e.g. Wang *et al.* 1999; Nesbitt & Tanksley 2002) and speciation and species diversification (e.g. Barraclough & Vogler 2000; Barraclough & Nee 2001). Phylogenies also provide information to develop comprehensive comparative systems in developmental biology (e.g. Doust & Kellogg 2002; Thießen *et al.* 2002) and comparative genomics (e.g. Bennetzen & Kellogg 1997; Eisen & Fraser 2003; Hong *et al.* 2003). In the last two decades, phylogenies have come of age as a universal component of comparative biology (Hillis 2004).

Questions remain about the best ways to measure progress towards estimating the Tree of Life (Donoghue 2004). Numbers of published phylogenetic trees and taxa represented therein alongside measures of levels of resolution among taxa, statistical confidence in published trees (Sanderson 1995), and discovery of paraphyly, polyphyly, and circumscription of monophyletic groups, all suggest that recent progress has

* Author for correspondence (colin.hughes@plants.ox.ac.uk).
† These authors contributed equally to this work.

Received 1 March 2005
Accepted 12 August 2005

© 2005 The Royal Society

indeed been impressive (reviewed in Cracraft & Donoghue 2004). However, measures of success are greatly influenced by the intended use(s) of the resultant phylogenetic trees. For the purposes of ridding classifications of non-monophyletic groups, completely resolved trees with uniformly high statistical support for all nodes are not necessarily required. Analyses of exemplar taxa that resolve a subset of well-supported nodes will often suffice for the purposes of providing robust phylogenetic classifications. There is little doubt that in this sense there has been massive progress towards assembling the Tree of Life. The angiosperms provide a good example. Despite some persistent areas of poor resolution and/or weak support (e.g. Wortley *et al.* 2005), the branching order of most clades of angiosperms is now relatively clear and well supported (e.g. Soltis *et al.* 1999; Savolainen & Chase 2003; Soltis & Soltis 2004), prompting a new classification (APGII 2003). Similar progress has been made towards delimiting genera and understanding relationships within many angiosperm families (e.g. Lavin *et al.* 2001). In this realm there are clear theoretical reasons to believe that further progress can be made by simply increasing the number of characters via sequencing additional readily accessible loci that can be combined in simultaneous analyses to increase resolution, accuracy (Hillis 1996, 1998) and support (Bremer *et al.* 1999) for critical nodes (Wortley *et al.* 2005). This is borne out by a number of empirical studies that have employed large scale DNA sequence datasets (Herniou *et al.* 2001; Bapteste *et al.* 2002; Matsuoka *et al.* 2002; Rokas *et al.* 2003), albeit with the provisos that increased character sampling is not compromised by sparser taxon sampling (Soltis *et al.* 2004) and that the methodological challenges posed by increasingly large data matrices can be adequately addressed (Sanderson & Driskell 2003).

However, when we come to examine progress in constructing accurate phylogenies among closely related species the situation, at least for plants, is less encouraging. Partially resolved gene and species trees are of limited use for studies of hybridization and polyploidy, character evolution, species diversification, speciation and domestication, or as comprehensive comparative systems underpinning biology more generally. Species-level phylogenies for detailed evolutionary studies often demand not just complete or near-complete taxon sampling, but ideally multiple accessions within species, as well as complete or near-complete resolution and high statistical support. Accurate statements about hybrid and polyploid origins require that the terminal branches leading to putative hybrids and their parents on gene trees are resolved and well supported. For example, lack of resolution in gene trees derived from conventional cpDNA and nrDNA loci in the legume genus *Leucaena* (figure 1) meant that the parentage of hybrids and polyploid species in the genus could not be hypothesized beyond major clades using these loci (Hughes *et al.* 2002). The same applies to studies of domestication where gene trees are used to infer potential progenitors of crop plants (e.g. Emshwiller & Doyle 1999, 2002). Using phylogenies to understand geographical patterns of species diversification or to plot lineage diversification through time and understand speciation processes requires similarly high-quality phylogenetic trees (e.g. Barraclough & Nee 2001).

Conflict attributable to lineage sorting, hybridization, polyploidy and introgression can cause gene genealogies to differ from the branching history of the organisms from which the genes were sampled (Pamilo & Nei 1988; Doyle 1992; Hillis 1995; Wendel & Doyle 1998). The impacts of these processes are thought to be more serious at species level than at higher levels, suggesting that fully resolved divergent species trees should not necessarily be expected for many plant genera (e.g. Linder & Rieseberg 2004). This means that the general desirability of employing multiple independent loci to infer accurate phylogenies can be much more critical in species-level phylogenies especially for plant groups where reticulation is common (Ferguson & Sang 2001; Raymond *et al.* 2002; Linder & Rieseberg 2004; Hegarty & Hiscock 2005). Employing additional DNA sequence loci to obtain greater species-tree resolution works on the premise that individual matrices will either be combined for simultaneous analysis to maximize congruence among independent sources of data (Nixon & Carpenter 1996), or, in the case of reticulation due to hybridization–introgression, that sets of congruent genes can be used to distinguish divergent from reticulate relationships. Differentiating between these alternatives requires gene trees with sufficient resolution to discriminate conflict from congruence.

Here we attempt to encapsulate how far plant species-level phylogenetics has progressed. We show that the rapid accumulation of DNA characters at higher taxonomic levels has not typically been matched in species-level analyses. It is also apparent that success in locating DNA sequence loci that can resolve relationships among species has been extremely patchy, lagging behind studies of animal taxa. Examples of densely sampled, well-resolved and supported species-level phylogenies for plants based on multiple independent loci are scarce and in many cases extremely challenging to reconstruct (see below). This means that many interesting questions in plant evolutionary biology and biogeography are currently frustrated by lack of resolution towards the tips of the tree. There has been little mention of these issues in the recent discussions about completing the Tree of Life (Cracraft & Donoghue 2004), although there have been recurrent calls for improved species-level phylogenies from those involved in evolutionary studies (e.g. Baldwin & Sanderson 1998; Barraclough & Nee 2001; Doyle *et al.* 2003*b*; Bailey *et al.* 2004; Futuyma 2004; Linder & Rieseberg 2004). In the last few years, new approaches to identify nuclear DNA sequence loci have emerged that have the potential to transform the current famine into a veritable feast of hypervariable species-level loci. We provide an explanatory overview that describes and compares the utility, potential benefits and pitfalls of the different approaches. Finally, we discuss the need to implement pilot studies that involve the screening of exemplar accessions to rank loci by potential utility prior to embarking on full scale sequence analysis of any one locus.

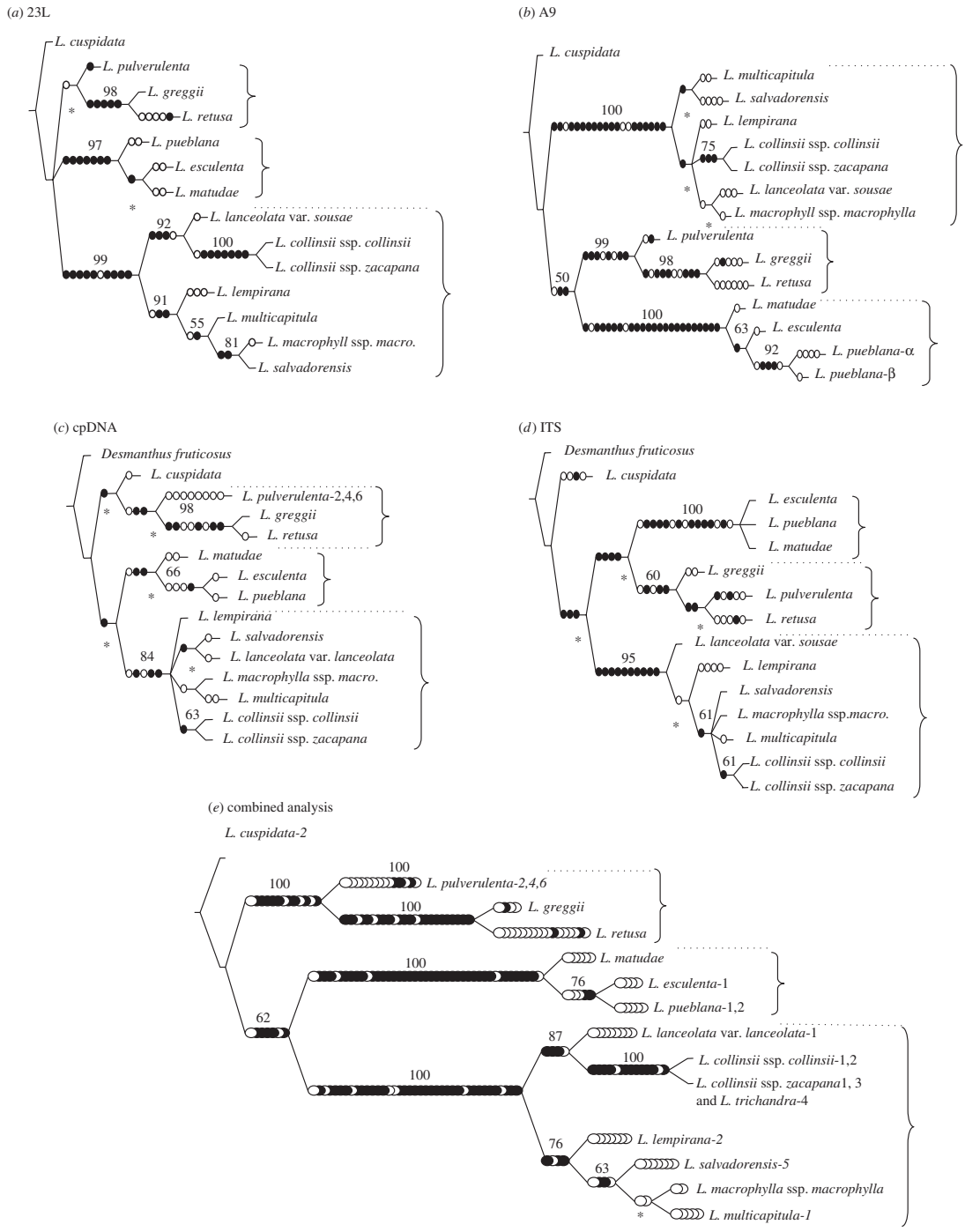


Figure 1. Example of increased resolution and support obtained among closely related diploid species of the Mimosoid legume genus *Leucaena* using a SCAR-based approach (modified and extended from Bailey *et al.* 2004). A fully resolved and well-supported diploid species tree for *Leucaena* is required to investigate the origins of a set of putative hybrid and five polyploid species (Hughes *et al.* 2002). Solid and open circles along branches represent unique and homoplastic unambiguous character state transformations, respectively. Values above branches are strict consensus bootstrap support values; nodes that are not present in the strict consensus trees are marked with an *. (*a*) Anonymous SCAR-based nuclear locus 23L—one of four EMPTs; $L=71$; CFI=0.75; CI=0.80; RI=0.88. (*b*) Anonymous SCAR-based nuclear locus A9—one of six EMPTs; $L=132$; CFI=0.69; CI=0.75; RI=0.88. (*c*) Chloroplast RFLP—one of 24 EMPTs; $L=63$; CFI=0.46; CI=0.58; RI=0.69. (*d*) nrDNA ITS—one of six EMPTs; $L=100$; CFI=0.46; CI=0.70; RI=0.81. (*e*) Combined analysis of 23L, A9, cpDNA and ITS datasets—one of three EMPTs; $L=339$; CFI=0.92; CI=0.73; RI=0.84. EMPT=equally most parsimonious trees; L=length (number of steps); CFI=consensus fork index (calculated from strict consensus trees); CI=consistency index; RI=retention index.

2. SURVEY OF SPECIES-LEVEL MATRICES

One of the strengths of DNA sequence data for resolving relationships is the scope to select among conserved and more rapidly evolving sequence regions, the 'tortoise and hare' of Small *et al.* (1998), to address questions at different hierarchical levels (e.g. Small *et al.* 1998, 2004; Soltis & Soltis 1998; Yang 1998). Molecular systematists have thus been able to tailor the selection of sequence regions to the question at hand—using more slowly evolving loci to analyse higher level relationships and more rapidly evolving regions for studies of closely related species. In this way, sequences from different genomes, genes and coding/non-coding regions can potentially generate data appropriate to resolve all branches of the Tree of Life from species level (or even within species) upwards.

There are a number of different ways to measure the potential informativeness of different DNA sequence loci (Wortley & Scotland in press). Two measures, the number of parsimony informative (PI) characters and the percentage of PI characters (total number of PI characters divided by the aligned length), are widely used. Wortley & Scotland (in press) point to the minimum number of PI character state changes as a more accurate measure of potential utility. However, this measure is rarely reported and for comparisons among molecular datasets is strongly correlated with number of PI characters. Number and percentage PI characters are not independent. From a tree building perspective, the total number of PI characters is more important than the per cent variability, in that a highly variable but very short region may not provide sufficient characters. However, from a practical perspective, very long regions, while potentially contributing more characters, may be inefficient in terms of sequencing effort. In spite of falling costs of sequencing, obtaining the maximum number of characters per sequencing reaction will remain an important factor for species-level phylogeny reconstruction given the scale of taxon sampling with multiple accessions of species that is needed. Thus, both total number of PI characters and per cent PI characters are useful measures for comparing loci. These can be reported relative to the number of taxa as character–taxon ratios.

We have undertaken a two-tier survey of recent plant species-level phylogenetic analyses. First, we surveyed a sample of the most recently published studies (2003 and 2004) from *American Journal of Botany*, *Molecular Phylogenetics and Evolution*, *Systematic Botany* and *Taxon* (136 studies encompassing 345 individual data matrices). These were assessed in terms of their objectives, taxon sampling, character sampling (numbers and types of DNA sequence loci used) and the degree of resolution obtained. In order to assess resolution we used the consensus fork index (CFI; Colless 1980; the number of resolved nodes on a strict consensus tree divided by the number of possible resolved nodes, $n-2$, where n is the number of terminals). Second, we looked in more detail at potential informativeness of different types of sequence loci using a subset of 226 recent plant species-level data matrices from studies that include nuclear (nDNA) and non-coding chloroplast (cpDNA) and/or nuclear ribosomal (nrDNA) loci. We have restricted both surveys and discussions to sets of congeneric species, and data matrix statistics to ingroup species only. This could be viewed as somewhat arbitrary not least because genera are not uniformly applied, but also because some studies focus on subclades within genera. Numbers of PI characters can also be influenced by levels of taxon sampling within genera. Nevertheless, a broad sample of species-level studies allows for the direct comparison of variation provided by different classes of loci to evaluate the general utility of each class.

3. CHLOROPLAST AND ITS DATA

Results from the broad survey of species-level matrices are summarized in figure 2. So far, the vast majority of species-level phylogenetic analyses of plants have relied on a limited set of non-coding cpDNA loci and the nrDNA internal transcribed spacer (ITS) region (ITS 1, 5.8S and ITS 2; figure 2*a*; Alvarez & Wendel 2003; Shaw *et al.* 2005*b*). These loci account for 87% of 345 matrices surveyed (figure 2*a*) and have been favoured because they are variable at low taxonomic levels and easy to amplify using universal primers (Taberlet *et al.* 1991; Baldwin *et al.* 1995; Shaw *et al.* 2005*b*). Routine application of these loci has produced an explosion of new phylogenetic data providing insights into species relationships across a wide range of plant genera. The average resolution across studies as measured by the CFI is 0.64 (figure 2*c*). Even at this level of resolution, many of these analyses serve the primary purpose—i.e. form a basis for generic delimitation and infrageneric classification (figure 2*b*; Alvarez & Wendel 2003). However, the limitations of relying on this limited set of loci are also apparent, especially for studies with objectives beyond classification (figure 2*b*). Only 11% (15 trees) of our sample of recently published species-level phylogenies are fully resolved, more than half are poorly resolved (CFI <0.6), and 14% are very poorly resolved (CFI $<$ 0.4; figure 2*c*). Lack of resolution is, thus, a widespread problem even for morphologically diverse species representing large genera (Wojciechowski *et al.* 1999; Richardson *et al.* 2001; Malcomber 2002; Mitchell & Heenan 2002; Syring *et al.* in press). Usually, this lack of resolution is directly attributable to insufficient variation (Small *et al.* 1998; Bailey *et al.* 2004; Shaw *et al.* 2005*b*), rather than incongruence, although poorly resolved trees can in themselves preclude observation of incongruence. It is also worth noting that taxon sampling is extremely variable across species-level studies, with only 37% fully sampled (figure 2*e*).

Results from the more detailed comparisons of the potential utility of different DNA sequence loci are presented in figure 3. Percentage PI character values for non-coding cpDNA range from 0.2 to 8.5 (-13.1 in Shaw *et al.* 2005*b*) and for ITS from 0.3 to 14 (-24) (figure 3). These values are an order of magnitude lower than those available in many higher level analyses (typically 20–60%). They are also lower than commonly used mitochondrial (mt) DNA loci in animals, where highly variable (typically 15–30% PI characters) DNA sequences have facilitated reconstruction of highly resolved species-level phylogenies (Avise *et al.* 1987; Moritz *et al.* 1987; Moore 1995) and speciation events (e.g. Irwin *et al.* 2001). The low percentages of

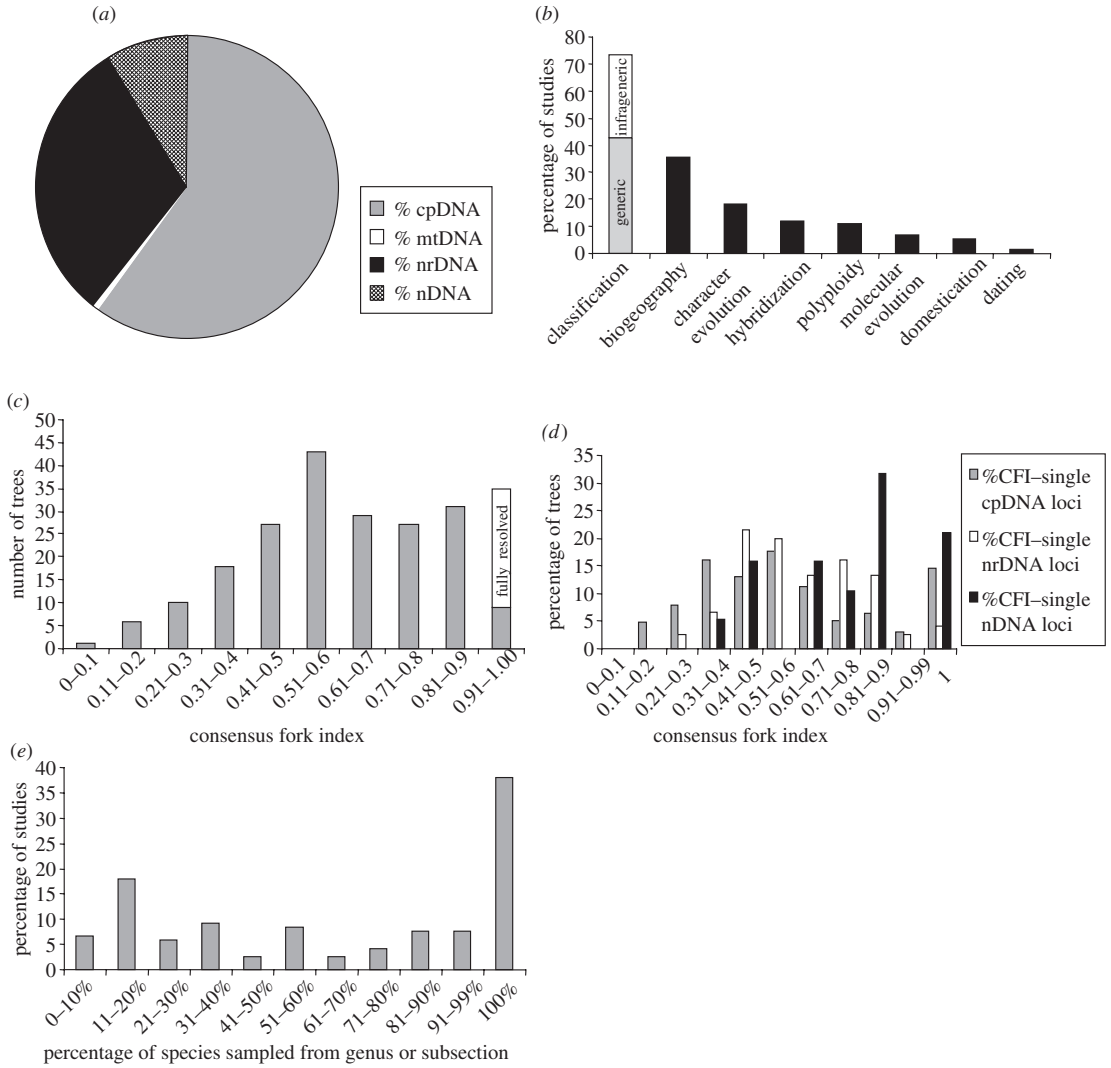


Figure 2. Attributes of current plant species-level phylogenetic studies based on a survey of 345 data matrices from 136 studies (see text for details). (*a*) Loci currently used in species-level analyses: proportions of cpDNA, nrDNA (ITS), mtDNA and nDNA loci. (*b*) Objectives and applications of species-level phylogenetic analyses. (*c*) Resolution obtained in recent species-level analyses across all loci: frequency distribution of consensus fork indices, including both combined and separate analyses. (*d*) Resolution obtained using different loci: frequency distribution of consensus fork indices for analyses of individual loci. (*e*) Variation in taxon sampling in current species-level analyses.

PI characters are reflected in lower CFI values for cpDNA and ITS loci when analysed alone compared to some nuclear loci (figure 2*d*).

In the face of insufficient variation to resolve relationships among sequences from conventional non-coding cpDNA and/or ITS loci, a number of approaches have been adopted (see below). Increasing the amount of cpDNA sequence data, potentially guided by selecting more variable non-coding cpDNA loci (Shaw *et al.* 2005*b*), has been successfully used to obtain greater resolution and support. Inevitably, this approach involves large volumes of DNA sequencing (e.g. 4–7 kb of plastid sequence used by Cronn *et al.* 2002; Clarkson *et al.* 2004; Shaw & Small 2004) and still does not guarantee well-resolved trees (e.g. Shaw & Small 2004). Other approaches have been to sequence the structurally complex external transcribed spacer (ETS; e.g. Linder *et al.* 2000), or resort to dominant PCR-based fragment length characters derived from amplified fragment length polymorphisms (AFLPs), inter-simple sequence repeats (ISSRs) or randomly amplified polymorphic DNAs (RAPDs) (reviewed by Wolfe & Liston 1998; Harris 1999). Contrary to the situation in the animal kingdom, mtDNA sequences have only played a minor role in phylogenetic studies of plants. This is because of structural instability, gene transfer to the nucleus and alleged sequence conservation (Wolfe *et al.* 1987; Palmer 1992; Soltis & Soltis 1998). However, some recent studies have found mtDNA sequence loci sufficient to build species-level phylogenies in plants (e.g. Sanjur *et al.* 2002).

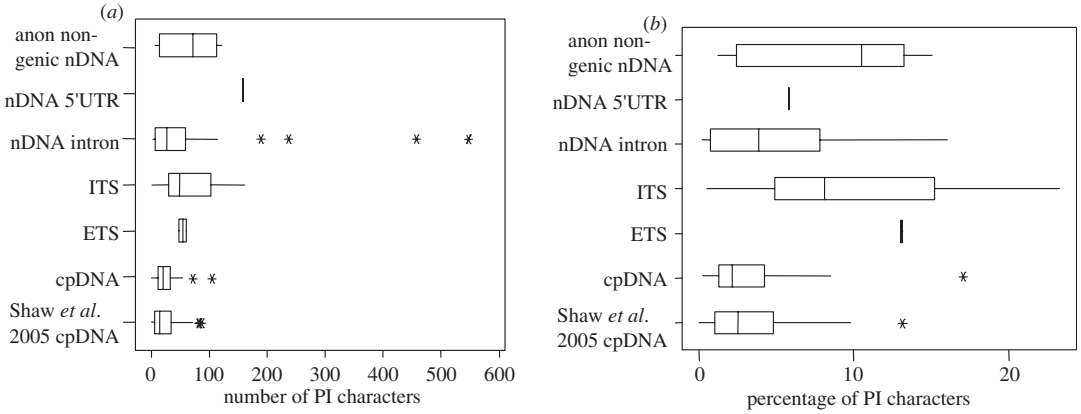


Figure 3. Potential informativeness of cpDNA, nrDNA ITS, nrDNA ETS, nDNA intron, nDNA 5′ UTR and anonymous, SCAR-based DNA sequence loci at species level. (*a*) Number of PI characters per locus. (*b*) Percentage of PI characters for individual loci. Boxes encompass the 75% interquartiles, showing the median while outliers are indicated with an *. Data are derived from species-level phylogenetic analyses of 226 matrices from studies of plant genera. Character statistics are calculated for congeneric terminals only, excluding outgroup taxa. Tallies of characters included both PI substitutions and PI indels coded according to the simple gap coding method of Simmons & Ochoterena (2000). Plant groups included are: *Sphagnum* (Shaw *et al.* 2003), *Leucaena* (Bailey *et al.* 2004 and unpublished data), *Gossypium* (Cronn *et al.* 2002), *Gaetnera* (Malcomber 2002), *Spaerocardamum* (Bailey & Doyle 1999), *Passiflora* (Yockteng & Nadot 2004), *Glycine*—alignments assembled for nine species based on data from Doyle *et al.* (1996, 2003*a*,*b*) and Sakai *et al.* (2003), *Paeonia*—alignments assembled for eight species based on data from Sang *et al.* (1995, 1997), Ferguson & Sang (2001) and Tank & Sang (2001), *Lycopersicum*—alignments assembled for 13 accessions of seven species based on data from Nesbitt & Tanksley (2002), *Viburnum* (Winkworth & Donoghue 2004), *Clarkia* (Ford & Gottlieb 2003), *Hibiscus* (Small 2004), *Neillia*/*Stephandra* (Oh & Potter 2003) and *Bursera* (Weeks & Simpson 2004). The broader survey of non-coding cpDNA loci (Shaw *et al.* 2005*b*) is included as a second cpDNA bar.

Aside from the lack of sufficient variation in many groups, cpDNA and nrDNA loci individually are of reduced utility for reconstructing relationships among potentially reticulating species (Linder & Rieseberg 2004). Uniparental inheritance of plastids makes cpDNA essential for teasing apart maternal versus paternal contributions to putative hybrid species when compared to other data. Similarly, nrDNA multi-copy sequences that have undergone complete concerted evolution do not retain evidence of biparental inheritance and therefore reticulation (Baldwin *et al.* 1995; Wendel *et al.* 1995). The apparently unpredictable extent and direction of concerted evolution among nrDNA repeats, which can even vary among individuals of the same species (Doyle *et al.* 2004), can create situations where nrDNA data provide partial and potentially confusing information about reticulation. Conversely, lack of concerted evolution can create serious paralogy-related problems in some ITS datasets (Alvarez & Wendel 2003; Bailey *et al.* 2003; Doyle *et al.* 2004). These limitations mean that multiple independent bi-parentally inherited loci are essential for species-level phylogenetic studies where reticulation is a possibility (Ferguson & Sang 2001; Raymond *et al.* 2002; Linder & Rieseberg 2004; Hegarty & Hiscock 2005), and these can only come from the nuclear genome.

4. STRATEGIES FOR SELECTING NUCLEAR SEQUENCE LOCI

(a) *Low-copy nuclear gene approaches*

The most widely used approach to resolve relationships among closely related plant species where cpDNA and ITS loci fail to provide resolution has been to use sequence data from low-copy nuclear genes (LCNG), or specific members of multi-gene families and especially their introns (Doyle & Doyle 1999; Cronn *et al.* 2002; Sang 2002; Small *et al.* 2004). Initial studies using nuclear sequences were restricted by the availability of primers to a handful of LCNGs (e.g. floral development genes) under investigation in related taxa in other fields (Strand *et al.* 1997; Small *et al.* 1998; Cronn *et al.* 2002). This approach has succeeded in locating sporadic highly variable loci for specific plant groups (e.g. Histone H3-D in *Glycine*, Doyle *et al.* 1996; GBSSI/*waxy* in grasses, Mason-Gamer *et al.* 1998 and Rosaceae, Evans *et al.* 2000; GPAT in *Paeonia*, Tank & Sang 2001; *pgiC1* and *pgiC2* in *Clarkia*, Ford & Gottlieb 2003; *cycloidea* in *Lupinus*, Ree *et al.* 2004). However, LCNG loci still comprise less than 10% of published species-level matrices (figure 2*a*), and for the vast majority of plant genera choice of LCNG loci remains limited and access to alternatives a significant hurdle.

Also, it is increasingly clear that nuclear-encoded loci vary widely in nucleotide substitution rates among closely related species. For example, Cronn *et al.* (2002) found a fivefold range in substitution rates among 12 nuclear encoded loci developed for *Gossypium*. In combined analysis of these nuclear loci plus ITS, 53% of the PI characters were recovered from just two loci (ITS and Fad2-1). Furthermore, the ITS region provided 33% of PI sites from just 10% of the total sequence data. An even wider sevenfold difference in substitution rates was documented in a survey of 36 nuclear genes for the same species (Senchina *et al.* 2003). There are clear indications of similar variability in rates of evolution among intron sequences of closely

related species from other genera (e.g. Doyle *et al.* 2003*a*, 2004 for *Glycine*; Syring *et al.* in press for *Pinus*). A wider survey of nuclear intron datasets confirms this variability which ranges from 0.21 to 16% PI characters across groups of congeneric species (figure 3). Of the 24 nuclear intron datasets surveyed here which have ITS data for the same taxa, only seven provided more PI characters than ITS (figure 3). While this variability has been recognized for some time (Small *et al.* 2004), what has not been widely acknowledged is that the majority of nuclear introns sequenced across groups of species, so far, are less variable (and often much less variable) and provide fewer PI characters than ITS. Only a small subset exhibits similar or higher levels of variation (figure 3). This is perhaps not surprising considering that some introns, and particularly large introns, have been shown to influence patterns of gene expression, acting as *cis*-regulatory elements (e.g. Sieburth & Meyerowitz 1997). A similarly wide span of variation has been found for LCNG loci used in phylogenetic analyses of animal species where most LCNG sequences are less variable than mtDNA loci (e.g. Helbig *et al.* 2005; Peters *et al.* 2005).

It is also notable that no universally useful LCNG sequence loci have been developed during the last 10 years (Sang 2002). Low-copy nuclear loci have had to be developed specifically for the taxonomic group of interest. In fact, building a phylogeny for a particular plant group has been the primary goal of many of the studies that have also discussed the potential broader utility of specific genes (e.g. Doyle *et al.* 1996; Galloway *et al.* 1998; Mason-Gamer *et al.* 1998; Small *et al.* 1998; Bailey & Doyle 1999; Emshwiller & Doyle 1999; Ree *et al.* 2004), a potential that has not so far been realized. Difficulties of extrapolating loci that have provided informative variation in one group can be due to failure of primers to amplify, variation in copy number, differences in intron presence or length, or differences in nucleotide substitution rates in different plant groups. Thus, up to now it seems that development of LCNG sequence loci has been somewhat of a lottery spawning the occasional lucky jackpot against a backdrop of generally disappointing results and considerable investment in developing and optimizing loci that often fail to produce much data (figure 3). This backdrop is rarely reported in the literature but is reflected in the experiences of several researchers who have acknowledged an element of good fortune in developing highly informative LCNG sequence loci, when they come to try to find similarly variable additional loci for the same group of plants (Doyle *et al.* 2003*a*; Small *et al.* 2004).

A further issue of concern here is the potential imbalance caused by one or a few high variability loci dominating and potentially distorting analyses. In comparisons of closely related species, the range of variation between loci is often skewed towards having few characters per locus. Having one highly variable LCNG locus alongside a set of much less variable loci may create a false sense of security when reconstructing species trees because poorly resolved trees are less likely to reveal incongruence. Furthermore, the combination of such loci in simultaneous analysis runs an increased risk of having one or a few highly variable loci that are incongruent with the underlying species tree negatively impacting the signal provided by less variable loci (Bull *et al.* 1993; Miyamoto & Fitch 1995; Page & Charleston 1997; Slowinski & Page 1999; Page 2000). Methods that screen multiple loci for comparable levels of informative variation prior to generating fully sampled trees (see below) will ensure that individual loci provide more equitable contributions towards combined analyses. While it is worth bearing in mind the idea that different loci may provide resolution on different parts of the tree (e.g. Pennington 1996; Ree *et al.* 2004; Helbig *et al.* 2005), this has not been convincingly demonstrated and balanced contributions of character data from a set of highly variable individual loci is likely to be more informative.

(b) *Comparative anchor tagged sequence-based approaches*

In the absence of universally accessible and informative LCNG loci, alternative approaches to locate more variable nuclear genes have been tried. The first of these involves comparisons of EST and/or complete genome sequences between model organisms to identify evolutionarily conserved regions, termed comparative anchor tagged sequences (CATS) (Lyons *et al.* 1997; Chandappa *et al.* 2005; Syring *et al.* in press) or conserved orthologue set (COS) markers (Fulton *et al.* 2002). While the primary goal of many of these studies has been comparative genome mapping, CATS markers can be used to develop sets of primers circumscribing potentially amplifiable sequence regions for phylogenetic analyses (figure 4*a*). For example, comparison of the tomato EST database with the *Arabidopsis* genome sequence identified a set of more than 1000 COS markers (Fulton *et al.* 2002). Similarly, screening of the *Medicago trunculata* EST database against the *Arabidopsis* genome sequence, and other available legume sequence, identified 274 loci that show strong sequence similarity. A subset of these are being screened as potential DNA sequence loci for species-level phylogenetic reconstruction of the large legume genus *Astragalus* (Scherson *et al.* in press). Similarly, Chandappa *et al.* (2005) used genome and EST sequence alignments to design sets of conserved markers for monocts based on rice–onion and rice–banana alignments. This approach has also been used directly with the *Arabidopsis* and rice genome sequences to identify conserved DNA oligomers that can be employed as primers to amplify orthologous DNA sequence loci for species-level phylogenetics (Padolina *et al.* 2004; Xu *et al.* 2004). One of the most notable features of these methods is the large number of potential primer pairs that can be generated. The few studies undertaken so far identify 100s or 1000s of possible DNA sequence loci and primer pairs. At the extreme, 13 418 candidate primer pair combinations were identified in the studies of Padolina *et al.* (2004) and Xu *et al.* (2004).

Development of CATS-based approaches is at an early stage and, as far as we are aware, they have yielded very few, if any hypervariable DNA sequence loci for use in species-level phylogenetics. CATS-based approaches, as currently implemented, generally

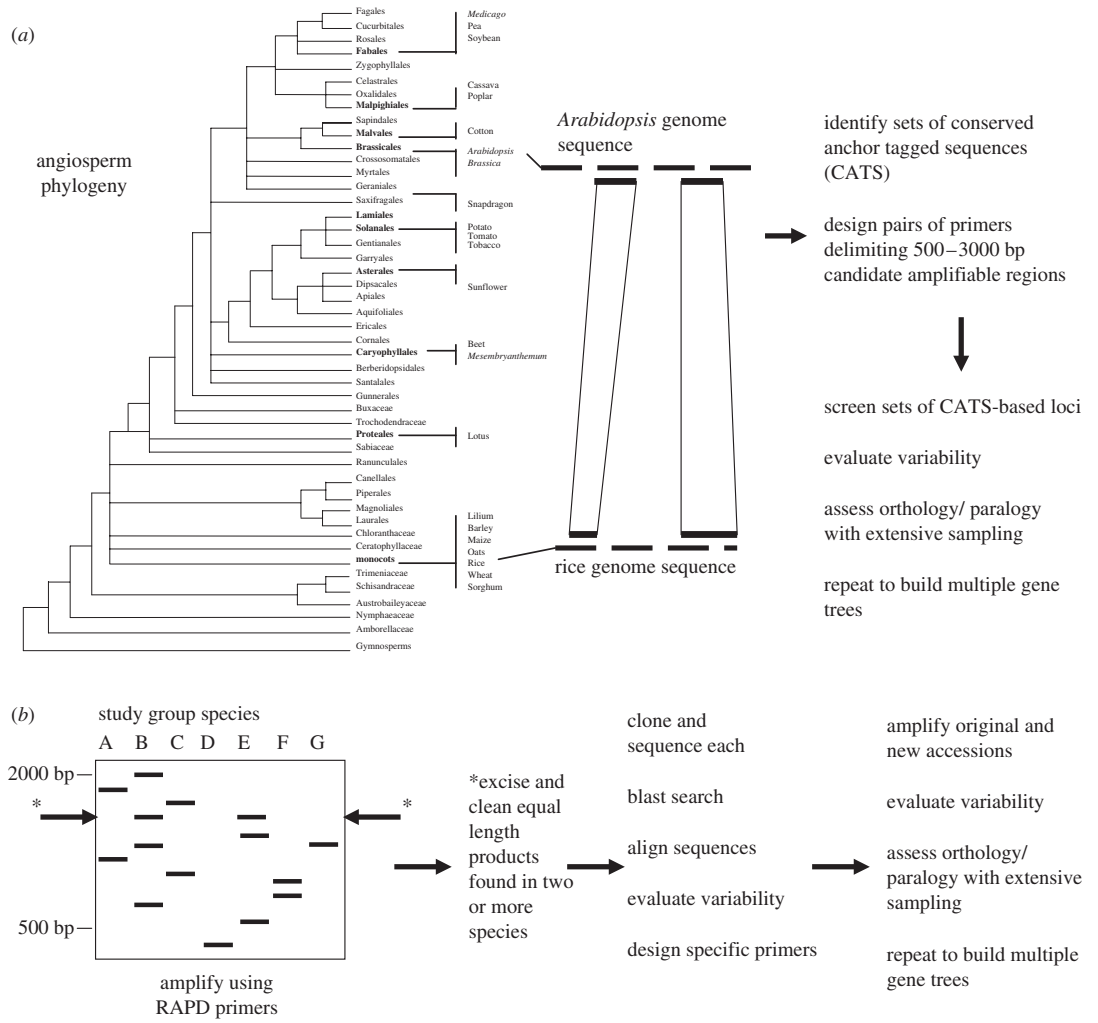


Figure 4. Overview of CATS-based and SCAR-based approaches used to develop nuclear DNA sequence loci. (*a*) The CATS-based approach, where genomic or EST sequences available for (usually) divergent taxa (phylogeny showing distribution of model taxa for which genomic or EST sequence data are available, modified from Soltis & Soltis 2004), are compared to identify sets of conserved anchor tagged sequences and design primer pairs delimiting 400–3000 bp potentially amplifiable regions that are subsequently screened for variability (Syring *et al.* 2004; Xu *et al.* 2004). (*b*) The SCAR-based approach where screening for variability in RAPD-generated DNA fragments among closely related species precedes design of primers (modified from Bailey *et al.* 2004).

compare EST/genomic sequences between widely divergent taxa (e.g. rice and *Arabidopsis*; figure 4*a*). While this may be advantageous for locating widely conserved markers (Fulton *et al.* 2002) thereby facilitating design of potentially 'universal' primers, it also means that primers typically span relatively more conserved coding regions. The simple fact that alignments can be constructed between such divergent taxa suggests that such regions are less likely to provide highly variable loci between closely related taxa. However, this is presumably a temporary limitation of our current data bases that include few taxa with many loci while many taxa have very few (or no) loci (e.g. Driskell *et al.* 2004). As the taxon×loci matrix becomes even a bit more densely populated, future CATS-based comparisons among more closely related taxa will become increasingly powerful for locating DNA sequence loci for specific groups of interest. For example, mapped conifer anchor loci developed from two species of pine (Brown *et al.* 2001) are being used to screen and select potential DNA sequence loci to resolve incongruent and incompletely resolved cpDNA and nrDNA phylogenetic trees for *Pinus* (Syring *et al.* in press). Other examples of comparisons across narrower taxonomic spans illustrate this potential (e.g. in Brassicaceae, Kuittenen *et al.* 2002; Lukens *et al.* 2003 and in monocots, Chandappa *et al.* 2005).

(c) *Sequence characterized amplified region-based approaches*

Another approach is to use RAPD or AFLP primers to generate and identify sequence characterized amplified

regions (SCARs; Melotto *et al.* 1996) that can be screened as potentially useful sequence loci prior to development of specific primers (Shaw *et al.* 2003; Bailey *et al.* 2004). This strategy is a modified version of the AFLP-based method of McLenachan *et al.* (2000) for characterizing population level gel-based markers. Under this method, PCR products of equal length amplified across a subset of species using commercially available random primers are excised, cloned, sequenced and aligned to evaluate variability (figure 4*b*; Bailey *et al.* 2004). By using a range of primer combinations many candidate loci can be generated. Specific SCAR primers are then designed for potentially useful regions that show high levels of variability compared to previously used conventional loci such as ITS. As for the CATS-based approaches, the SCAR strategy can be used to screen and generate numerous candidate loci (see below). The SCAR-based approach has been used successfully to increase resolution and support for groups of closely related species of legumes (Bailey *et al.* 2004; see figure 1) and mosses (Shaw *et al.* 2003, 2005*a*).

There are several potential benefits of using random genomic regions rather than previously characterized genic regions. First, no prior sequence information is required to develop many loci for a group of interest. Second, random amplification of regions across the entire genome eliminates the restriction of working within known genic exon/intron regions (and potentially adjacent 3′ UTRs and 5′ promoter regions), which are likely to represent the less variable half of the nuclear genome. Given that 53–57% of the *Arabidopsis* genome has been classified as non-coding DNA (Initiative 2000), it presents significant untapped potential as a source of DNA characters (figure 3; Shaw *et al.* 2003, 2005*a*; Bailey *et al.* 2004). Third, unlike the CATS-based approaches, the initial screening process facilitates selection of loci with levels of variation needed to address species-level problems prior to development of specific primers. Fourth, by homing in on sets of loci with similar high levels of sequence divergence this reduces the chances of locus imbalance where one highly variable locus overrides signal from less variable loci. Finally, by initially selecting for regions with limited length variation this reduces the chances of encountering highly length variable regions that may be difficult to align (e.g. Syring *et al.* in press). This last feature may of course eliminate many potentially useful but length variable DNA sequence loci from consideration (Britten *et al.* 2003; Fondon & Gamer 2004).

A number of potential drawbacks associated with SCAR-based loci are immediately apparent. First, there is no indication that selected loci will be biparentally inherited which is critical for studies of reticulation. BLAST searches can be used to check for significant similarity to complete angiosperm cpDNA or mtDNA sequences available in GenBank, eliminating known uniparentally inherited loci from those sources. Screening via artificial or other known hybrids provides a more certain route to ascertain this. A second potential limitation is the possibility of amplifying regions of junk repeat DNA, with consequent paralogy problems. While this might suggest greater probability of encountering paralogy problems with SCAR-based loci, in practice all loci are subject to most, if not all of the same considerations. Dealing with paralogy needs to be an integral and central step in screening all DNA sequence loci for phylogenetic analysis (see below). Finally, there are clearly tradeoffs to be weighed up when looking for SCAR-based loci in terms of what span of taxa are to be included in the initial selection of amplified fragments. Inclusion of bands that amplify in just two or three closely related species runs the risk that more divergent taxa will not be amplified by primers developed from the initial sequence data, thereby potentially forfeiting inclusion of outgroup taxa (Bailey *et al.* 2004). Conversely, selecting fragments that amplify across a wider range of taxa than the specific clade of interest may mitigate against finding the hypervariable regions that are needed to resolve relationships within that clade. On the plus side, this is under the control of the investigator; different options can be tested.

5. PILOT STUDIES AND SCREENING

During the short history of plant molecular systematics, the choice of biparentally inherited nuclear loci has been severely limited by the availability of useful primers. The accessibility of new approaches (discussed above) is rapidly expanding the plant systematist's 'toolbox' to a situation with essentially unlimited options (Shaw *et al.* 2003; Bailey *et al.* 2004; Padolina *et al.* 2004; Xu *et al.* 2004; Syring *et al.* in press). For those embarking on the development of new datasets or studies of new groups, these breakthroughs provide tremendous opportunities. To derive maximum benefits from the available loci, investigators need to employ pilot studies incorporating extensive screening to identify the most promising loci prior to investing heavily in any one locus (e.g. Bailey *et al.* 2004; Syring *et al.* in press). The exact order of steps in a screening process will depend on whether or not one is employing an approach that begins with the purchase of specific primers (e.g. CATS as a modified LCNG approach) or the use of random primers (e.g. SCAR-based approaches), but the critical factors that need to be considered remain essentially the same. Any screening approach should consider: (i) how many loci and what DNA samples/taxa to screen; (ii) ease and reliability of PCR amplification and sequencing; (iii) the potential orthology of sequences generated; (iv) alignment difficulties; and (v) last, but certainly not least, relative resolving power of each locus (e.g. Strand *et al.* 1997; Cronn *et al.* 2002; Bailey *et al.* 2004; Syring *et al.* in press).

The number of screened loci and DNA samples used in the selection process are largely dictated by the study group and available resources. Without doubt, selection strategies that screen larger numbers of samples and loci relative to other approaches will provide better starting points. However, these benefits need to be balanced against costs (time and money). The decreasing cost of oligonucleotide synthesis, PCR, and sequencing means that a pre-screening strategy that generates comparative data from 25 to 100 starting loci for 10 or more individuals is generally feasible and

realistic. Such numbers may initially sound excessive, but considering that the goal is to discard problematic loci identified at each level of screening, the actual cost will be much lower than if all loci were carried through every step. Sampling should, whenever possible, focus on non-hybrid diploid individuals spanning the taxonomic breadth of the ultimate study group. Furthermore, the inclusion of some intraspecific samples and samples for which data already exists (e.g. ITS or cpDNA data) will sharpen subsequent screening (e.g. for orthology and variability).

Loci that are easy to amplify and sequence are far preferable to those that are not. Problems with either amplification or sequencing in many accessions suggest that significant unnecessary effort and expense may be involved in primer development, cloning and extra sequencing. Furthermore, templates that are difficult to sequence may be indicative of underlying paralogy problems (see below). Regions which are reasonably straightforward to sequence using PCR primers (plus any additional sequencing primers needed for longer fragments) will facilitate efficient data generation. Of course, there will still be some individual samples that require cloning (e.g. due to hybridization, heterozygosity, or poor primer match), but selection should aim to minimize amplification and sequencing problems.

The assessment of sequence orthology and paralogy can and should be made at several stages. First, simple observation of the number of bands amplified using the 'locus specific' primers can reveal obvious problems. Primer pairs that amplify multiple bands, particularly of similar sizes, should be given lower priority or discarded. Second, difficult to sequence PCR products can be indicative of problems caused by a mixture of paralogous sequencing templates (e.g. Rauscher *et al.* 2002). Furthermore, Scherson *et al.* (in press) have recently noted that repeat patterns of subset polymorphisms in otherwise clean sequencing reads are more likely to be caused by underlying paralogy than by heterozygosity. Consideration of the patterns discussed by Scherson *et al.* (in press) is worthy of inclusion in the screening process. With the large numbers of potential regions now available, there is little reason to intentionally embark on studies that will require the dissection of paralogous gene copies for species-tree reconstruction. Lastly, assessments of orthology/paralogy can be made using comparisons of gene trees developed from each locus in relation to previously generated data (e.g. ITS and cpDNA). In these cases, the inclusion of intraspecific samples can also help identify potential paralogy problems.

Data generated during initial screening to demonstrate reliable single band amplifications and lack of serious sequencing difficulties or obvious paralogy problems can also be used to compare sequence variability, to identify those loci that are easy to use and informative. Assessments of inter-locus variability can be made among the pool of potential new loci and between these loci and previously developed data, such as ITS and cpDNA markers. Comparisons of the former are critical for the selection of the most variable loci, while evaluation of the latter should provide some idea how much additional resolution might (or might not) be gained from a fully sampled matrix using the 'best' of the newly screened loci. Measures of sequence variability may be based on percentage of PI characters, pairwise divergence, or more comprehensive analysis of gene-tree comparisons. Gene-tree approaches best characterize the overall distribution of variability and therefore the relative resolving power of each locus based on the same samples (Wenzel & Siddall 1999). Additional measures of utility for each gene may also be drawn from the individual gene trees (e.g. consistency index, number of resolved nodes, average branch support, etc.).

In plant species-level phylogenetics, rarely has the issue of too much variation been a problem. In general, researchers want to select the most variable loci that fit all the other criteria discussed above. However, one additional measure of variability and potential utility that may be considered in screening is the presence of indels that can provide potentially useful characters, but also potential alignment difficulties. Clean indels contribute useful phylogenetic markers while complicated indel patterns tied to high substitution rates can limit the utility of the data generated. Ideal loci will be relatively straightforward to align while maintaining relatively high levels of information (Bailey *et al.* 2004).

6. SUMMARY COMMENTS

One of the frequently cited advantages of DNA sequence data for reconstructing phylogenies is the virtually unlimited number of characters that can be generated. However, for closely related plant groups it may be extremely difficult to locate sufficiently variable DNA sequence loci to generate enough characters to provide well-resolved and robustly supported hypotheses of species relationships without generating very large volumes of sequence data. In contrast to those who study animals, for whom rapidly evolving mtDNA loci provide a rich source of species-level data, the plant molecular systematist's toolbox lacks readily accessible hypervariable sequence regions. Well-supported resolution at the gene-tree level remains elusive, suggesting limited confidence in how systematists currently infer relationships among closely related species.

Several routes to locating hypervariable nuclear DNA sequence loci for species-level phylogenetic analysis have emerged. The relative efficiency of these alternatives should increase rapidly because locating informative loci depends on how closely related the study group is to species for which significant (genome or EST) DNA sequence is available. SCAR-based approaches seem to present the most efficient strategy for many plant groups at present. However, as the gene$\times$taxon GenBank matrix become more densely populated, CATS-type approaches based on significantly narrower taxonomic spans will become increasingly frequent and productive.

A number of recent studies have demonstrated the high utility of anonymous, presumed non-coding, non-genic nuclear regions for resolving relationships (figures 1 and 3; Shaw *et al.* 2003, 2005*a*; Bailey *et al.* 2004). At the same time, it is increasingly clear that although low-copy nuclear introns range greatly in variability (figure 3; Small *et al.* 2004), the majority are

likely to be less variable (at least on a percentage basis) than ITS. Tapping into the large non-genic, or less-conserved genic fractions of plant genomes may thus be particularly productive. At present, the majority of researchers appear to be following CATS-based approaches to develop species-level nuclear DNA sequence loci (e.g. Syring *et al.* in press; Xu *et al.* 2004; Scherson *et al.* in press). In part this may be due to the large number of primer pairs that CATS approaches can provide, or to the attraction of locating universal or at least widely applicable markers. However, we believe that there is a need for caution to avoid the pitfalls associated with the LCNG approach that has been shown to be sub-optimal.

Whether the focus is on LCNG sequences or loci derived from SCAR-based approaches, the solutions will, for the most part, be lineage-specific. Once we stray beyond well-known nrDNA and cpDNA loci, opportunities to develop universal markers or primers appear to be extremely limited, if not non-existent (Sang 2002), suggesting a need to get away from universal gene thinking for species-level phylogenetics. A requirement for marker development, or at least substantive optimization, for specific plant lineages, and potentially for each individual study, shifts the emphasis towards approaches that involve screening prior to primer development. The need to screen candidate loci is further reinforced by the growing ease of identifying large numbers of candidate primer pairs. Thus, whatever sources of sequence data—non-coding cpDNA, LCNG, CATS-based or SCAR-based—are used, strategies for screening candidate sequence loci are key. Given the proliferation of potential sequence loci, screening can afford, and probably needs, to be more ruthless than anything envisaged before.

Finally, none of these developments denies the ongoing utility of, and need for, non-coding cpDNA and ITS data. Not only will these continue to provide useful data in their own right, but cpDNA and ITS data also form essential foundations for subsequent development of other DNA sequence loci. This will take the form of a framework to compare levels of variation, provide initial hypotheses of relationships to direct taxon sampling and assess orthology/paralogy concerns during screening and data generation for candidate nuclear DNA sequence loci.

The authors thank two anonymous reviewers for comments on an earlier version of the manuscript and John Syring, Rosita Scherson and Alex Wortley for pre-print copies of their papers. C.E.H. is supported by the Royal Society and R.J.E. by BBSRC.

REFERENCES

Alvarez, I. & Wendel, J. F. 2003 Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylog. Evol.* **29**, 417–434. (doi:10.1016/S1055-7903(03)00208-2)

APGII 2003 An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APGII. *Bot. J. Linn. Soc.* **141**, 399–436. (doi:10.1046/j.1095-8339.2003.t01-1-00158.x)

Avise, J. C., Arnold, J., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. 1987 Intraspecific phylogeography: mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* **18**, 489–522.

Bailey, C. D. & Doyle, J. J. 1999 Potential phylogenetic utility of the low-copy nuclear gene pistillata in Dicotylednous plants: comparison to nrDNA ITS and trnL intron in *Sphaerocardamum* and other Brassicaceae. *Mol. Phylog. Evol.* **13**, 20–30. (doi:10.1006/mpev.1999.0627)

Bailey, C. D., Carr, T. G., Harris, S. A. & Hughes, C. E. 2003 Characterization of angiosperm nrDNA polymorphism, paralogy and pseudogenes. *Mol. Phylog. Evol.* **29**, 435–455. (doi:10.1016/j.ympev.2003.08.021)

Bailey, C. D., Hughes, C. E. & Harris, S. A. 2004 Using RAPDs to identify DNA sequence loci for species level phylogeny reconstruction: an example from *Leucaena* (Fabaceae). *Syst. Bot.* **29**, 4–14. (doi:10.1600/036364404772973483)

Baldwin, B. G. & Sanderson, M. J. 1998 Age and rate of diversification of the Hawaiian silversword alliance. *Proc. Natl Acad. Sci.* **95**, 9402–9406. (doi:10.1073/pnas.95.16.9402)

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S. & Donoghue, M. J. 1995 The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann. Missouri Bot. Gard.* **82**, 247–277.

Bapteste, E. *et al.* 2002 The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl Acad. Sci.* **99**, 1414–1419. (doi:10.1073/pnas.032662799)

Barraclough, T. G. & Nee, S. 2001 Phylogenetics and speciation. *Trends Ecol. Evol.* **16**, 391–399. (doi:10.1016/S0169-5347(01)02161-9)

Barraclough, T. G. & Vogler, A. P. 2000 Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.* **155**, 419–434. (doi:10.1086/303332)

Bennetzen, J. L. & Kellogg, E. A. 1997 Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**, 1509–1514. (doi:10.1105/tpc.9.9.1509)

Bremer, B., Jansen, R. K., Oxelman, B., Backland, M., Lantz, H. & Kim, K.-J. 1999 More characters or more taxa for robust phylogeny—case study from the coffee family (Rubiaceae). *Syst. Biol.* **48**, 413–435. (doi:10.1080/106351599260085)

Britten, R. J., Rowen, L., Williams, J. & Cameron, R. A. 2003 Majority of divergence between closely related DNA samples is due to indels. *Proc. Natl Acad. Sci.* **100**, 4661–4665. (doi:10.1073/pnas.0330964100)

Brown, G. R., Kadd, E. E., Bassoni, D. L., Kiehne, K. L., Temesgen, B., van Buijtenen, J. P., Sewell, M. H., Marshall, K. A. & Neale, D. B. 2001 Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. *Genetics* **159**, 799–809.

Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L. & Waddell, P. J. 1993 Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* **42**, 384–397.

Chandappa, L. H., Feltus, F. A., Singh, H. P. & Paterson, A. H. 2005 Conserved PCR primers from rice–banana and rice–onion alignments. XIII Plant–Animal Genome Conf. Abstract, San Diego, USA.

Clarkson, J. J., Knapp, S., Garcia, V. F., Olmstead, R. G., Leitch, A. R. & Chase, M. W. 2004 Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. *Mol. Phylog. Evol.* **33**, 75–90. (doi:10.1016/j.ympev.2004.05.002)

Colless, D. H. 1980 Congruence between morphometric and allozyme data for *Menidia* species: a reappraisal. *Syst. Zool.* **29**, 288–299.

Cracraft, J. & Donoghue, M. J. (eds) 2004 *Assembling the Tree of Life*, p. 576. New York: Oxford University Press.

Cronn, R. C., Small, R. L., Haselkorn, T. & Wendel, J. F. 2002 Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Am. J. Bot.* **89**, 707–725.

Donoghue, M. J. 2004 Immeasurable progress on the Tree of Life. In *Assembling the Tree of Life* (ed. J. Cracraft & M. J. Donoghue), pp. 548–552. New York: Oxford University Press.

Donoghue, M. J. & Cracraft, J. 2004 Introduction. Charting the Tree of Life. In *Assembling the Tree of Life* (ed. J. Cracraft & M. J. Donoghue), pp. 1–4. New York: Oxford University Press.

Doust, A. N. & Kellogg, E. A. 2002 Inflorescence diversification in the panicoid 'bristle grass' clade (Paniceae, Poaceae): evidence from molecular phylogenies and developmental morphology I. *Am. J. Bot.* **89**, 1203–1222.

Doyle, J. J. 1992 Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot.* **17**, 144–163.

Doyle, J. J. & Doyle, J. L. 1999 Nuclear protein-coding genes in phylogeny reconstruction and homology assessment: some examples from Leguminosae. In *Molecular systematics and plant evolution* (ed. P.M Hollingsworth, R. M. Bateman & R. J. Gornall), pp. 229–254. London: Taylor & Francis.

Doyle, J. J., Kanazin, V. & Shoemaker, R. C. 1996 Phylogenetic utility of histone H3 intron sequences in the perennial relatives of soybean (*Glycine*: Leguminosae). *Mol. Phylog. Evol.* **6**, 438–447. (doi:10.1006/mpev.1996.0092)

Doyle, J. J., Doyle, J. L. & Harbison, C. 2003*a* Chloroplast expressed glutamine synthetase in *Glycine* and related Leguminosae: phylogeny, gene duplication and ancient polyploidy. *Syst. Bot.* **28**, 567–577.

Doyle, J. J., Doyle, J. L., Rauscher, J. T. & Brown, A. H. D. 2003*b* Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus Glycine). *New Phytol.* **161**, 121–132. (doi:10.1046/j.1469-8137.2003.00949.x)

Doyle, J. J., Doyle, J. L., Rauscher, J. T. & Brown, A. H. D. 2004 Evolution of the perennial soybean polyploidy complex (*Glycine* subgenus Glycine): a study of contrasts. *Biol. J. Linn. Soc.* **82**, 583–597. (doi:10.1111/j.1095-8312.2004.00343.x)

Driskell, A. C., Ane, C., Burleigh, J. G., McMahon, M. M., O'Meara, B. C. & Sanderson, M. J. 2004 Prospects for building the Tree of Life from large sequence databases. *Science* **306**, 1172–1174. (doi:10.1126/science.1102036)

Eisen, J. A. & Fraser, C. M. 2003 Phylogenomics: intersection of evolution and genomics. *Science* **300**, 1706–1707. (doi:10.1126/science.1086292)

Emshwiller, E. & Doyle, J. J. 1999 Chloroplast-expressed glutamine synthetase (ncpGS): potential utility for phylogenetic analysis with an example from *Oxalis* (Oxalidaceae). *Mol. Phylog. Evol.* **12**, 310–319. (doi:10.1006/mpev.1999.0613)

Emshwiller, E. & Doyle, J. J. 2002 Origins of domestication and polyploidy in Oca (*Oxalis tuberosa*: Oxalidaceae). 2. Chloroplast-expressed glutamine synthetase data. *Am. J. Bot.* **89**, 1042–1056.

Evans, R. C., Alice, L. A., Campbell, C. S., Kellogg, E. A. & Dickinson, T. A. 2000 The granule-bound starch synthase (GBSSI) gene in Rosaceae: multiple loci and phylogenetic utility. *Mol. Phylog. Evol.* **17**, 388–400. (doi:10.1006/mpev.2000.0828)

Ferguson, D. & Sang, T. 2001 Speciation through homoploid hybridization between allotetraploids in peonies (*Paeonia*). *Proc. Natl Acad. Sci.* **98**, 3915–3919. (doi:10.1073/pnas.061288698)

Fondon, J. W. & Gamer, H. R. 2004 Molecular origins of rapid and continuous morphological evolution. *Proc. Natl Acad. Sci.* **101**, 18 058–18 063. (doi:10.1073/pnas.0408118101)

Ford, V. S. & Gottlieb, L. D. 2003 Reassessment of phylogenetic relationships in *Clarkia* sect. Sympherica. *Am. J. Bot.* **90**, 284–292.

Fulton, T. M., Van der Hoeven, R., Eannetta, N. T. & Tanksley, S. D. 2002 Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* **14**, 1457–1467. (doi:10.1105/tpc.010479)

Futuyma, D. J. 2004 The fruit of the Tree of Life. Insights into evolution and ecology. In *Assembling the Tree of Life* (ed. J. Cracraft & M. J. Donoghue), pp. 25–39. New York: Oxford University Press.

Galloway, G. L., Malmberg, R. L. & Price, R. A. 1998 Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. *Mol. Biol. Evol.* **15**, 1312–1320.

Harris, S. A. 1999 RAPDs in systematics—a useful methodology? In *Molecular systematics and plant evolution* (ed. P. M. Hollingsworth, R. M. Bateman & R. J. Gornall), pp. 211–228. London: Taylor & Francis.

Hegarty, M. J. & Hiscock, S. J. 2005 Hybrid speciation in plants: new insights from molecular studies. *New Phytol.* **165**, 411–423. (doi:10.1111/j.1469-8137.2004.01253.x)

Helbig, A. J., Kocum, A., Seibold, I. & Braun, M. J. 2005 A multi-gene phylogeny of aquiline eagles (Aves: Accipitriformes) reveals extensive paraphyly at genus level. *Mol. Phylog. Evol.* **35**, 147–164. (doi:10.1016/j.ympev.2004.10.003)

Herniou, E. A., Luque, T., Chen, X., Vlak, J. M., Winstanley, D., Cory, J. S. & O'Reilly, D. R. 2001 Use of whole genome sequence data to infer baculovirus phylogeny. *J. Virol.* **75**, 8117–8126. (doi:10.1128/JVI.75.17.8117-8126.2001)

Hillis, D. M. 1995 Approaches for assessing phylogenetic accuracy. *Syst. Biol.* **44**, 3–16.

Hillis, D. M. 1996 Inferring complex phylogenies. *Nature* **383**, 130–131. (doi:10.1038/383130a0)

Hillis, D. M. 1998 Taxonomic sampling, phylogenetic accuracy and investigator bias. *Syst. Biol.* **47**, 3–8. (doi:10.1080/106351598260987)

Hillis, D. M. 2004 The Tree of Life and the grand synthesis in biology. In *Assembling the Tree of Life* (ed. J. Cracraft & M. J. Donoghue), pp. 545–547. New York: Oxford University Press.

Hong, R. L., Hamaguchia, M., Maximilian, A. B. & Weigel, D. 2003 Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. *Plant Cell* **15**, 1296–1309. (doi:10.1105/tpc.009548)

Hughes, C. E., Bailey, C. D. & Harris, S. A. 2002 Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: insights into polyploid origins and nrDNA polymorphism. *Am. J. Bot.* **89**, 1057–1073.

Initiative, T. A. G. 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815. (doi:10.1038/35048692)

Irwin, D. E., Bensch, S. & Price, T. D. 2001 Speciation in a ring. *Nature* **409**, 333–337. (doi:10.1038/35053059)

Kuittenen, H. *et al.* 2002 Primers for 22 candidate genes for ecological adaptations in Brassicaceae. *Mol. Ecol. Notes* **2**, 258–262. (doi:10.1046/j.1471-8286.2002.00210.x)

Lavin, M., Pennington, R. T., Klitgaard, B. B., Sprent, J. I., Cavalcante de Lima, H. & Gasson, P. E. 2001 The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *Am. J. Bot.* **88**, 503–533.

Lavin, M., Schrire, B. P., Lewis, G., Pennington, R. T., Delgado-Salinas, A., Thulin, M., Hughes, C. E., Matos, A. B. & Wojciechowski, M. F. 2004 Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. *Phil. Trans. R. Soc. B* **359**, 1509–1522. (doi:10.1098/rstb.2004.1536)

Linder, C. R. & Rieseberg, L. H. 2004 Reconstructing patterns of reticulate evolution in plants. *Am. J. Bot.* **91**, 1700–1708.

Linder, C. R., Goertzen, L. R., Heuval, B. V., Francisco-Ortega, J. & Jansen, R. K. 2000 The complete external transcribed spacer of 18S-26S rDNA: amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families. *Mol. Phylog. Evol.* **14**, 285–303. (doi:10.1006/mpev.1999.0706)

Lukens, L., Zou, F., Lydiate, D., Parkin, I. & Osborn, T. 2003 Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. *Genetics* **146**, 359–372.

Lyons, L. A., Laughlin, T. F., Copeland, N. G., Jenkins, N. A., Womack, J. E. & O'Brien, S. J. 1997 Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat. Genet.* **15**, 47–56. (doi:10.1038/ng0197-47)

Malcomber, S. T. 2002 Phylogeny of *Gaetnera* Lam. (Rubiaceae) based on multiple DNA markers: evidence of a rapid radiation in a widespread, morphologically diverse genus. *Evolution* **56**, 42–57.

Mason-Gamer, R. J., Weil, C. F. & Kellogg, E. A. 1998 Granule-bound starch synthase: structure, function and phylogenetic utility. *Mol. Biol. Evol.* **15**, 1658–1673.

Matsuoka, Y., Yamazaki, Y., Ogihara, Y. & Tsunewaki, K. 2002 Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. *Mol. Biol. Evol.* **19**, 2084–2091.

McLenachan, P. A., Stockler, K., Winkworth, R. C., McBreen, K. & Zauner, S. 2000 Markers derived from amplified fragment length polymorphism gels for plant ecology and evolution studies. *Mol. Ecol.* **9**, 1899–1903. (doi:10.1046/j.1365-294x.2000.01075.x)

Melotto, M., Afanador, L. & Kelly, J. D. 1996 Development of a SCAR marker linked to the I gene in common bean. *Genome* **39**, 1216–1219.

Mitchell, A. D. & Heenan, P. B. 2002 *Sophora* sect Edwardsia (Fabaceae): further evidence from nrDNA sequence data of a recent and rapid radiation around the Southern Oceans. *Bot. J. Linn. Soc.* **140**, 435–441. (doi:10.1046/j.1095-8339.2002.00101.x)

Miyamoto, M. M. & Fitch, W. M. 1995 Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* **44**, 64–76.

Moore, W. S. 1995 Inferring phylogenies from mtDNA variation: mitochondrial-gene versus nuclear-gene trees. *Evolution* **49**, 718–726.

Moritz, C., Dowling, T. E. & Brown, W. M. 1987 Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* **18**, 269–292. (doi:10.1146/annurev.es.18.110187.001413)

Nesbitt, T. C. & Tanksley, S. D. 2002 Comparative sequencing in the genus *Lycopersicon*: implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics* **162**, 365–379.

Nixon, K. C. & Carpenter, J. M. 1996 On simultaneous analysis. *Cladistics* **12**, 221–241. (doi:10.1111/j.1096-0031.1996.tb00010.x)

Oh, S.-H. & Potter, D. 2003 Phylogenetic utility of the second intron of *LEAFY* in *Neillia* and *Stephanandra* (Rosaceae) and implications for the origin of *Stephanandra*. *Mol. Phylog. Evol.* **29**, 203–215. (doi:10.1016/S1055-7903(03)00093-9)

Padolina, J., Timme, R., Linder, R., Briggs, W., Xu, W., Liu, W. & Miranker, D. 2004 Identification of broadly applicable nuclear DNA regions for phylogenetic reconstruction in angiosperms. *Bot. Soc. Am.* **2004**, 528. (Botany 2004 Meeting Abstract 528.)

Page, R. D. M. 2000 Extracting species trees from complex gene trees: reconciled trees and vertebrate phylogeny. *Mol. Phylog. Evol.* **14**, 89–106. (doi:10.1006/mpev.1999.0676)

Page, R. D. M. & Charleston, M. A. 1997 From gene to organismal phylogeny: reconciled trees and the gene-tree species-tree problem. *Mol. Phylog. Evol.* **7**, 231–240. (doi:10.1006/mpev.1996.0390)

Palmer, J. D. 1992 Mitochondrial DNA in plant systematics: applications and limitations. In *Molecular systematics of plants* (ed. P. S. Soltis, D. E. Soltis & J. J. Doyle), pp. 36–49. New York: Chapman & Hall.

Palmer, J. D., Soltis, D. E. & Chase, M. W. 2004 The plant Tree of Life: an overview and some points of view. *Am. J. Bot.* **91**, 1437–1445.

Pamilo, P. & Nei, M. 1988 Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5**, 568–583.

Pennington, R. T. 1996 Molecular and morphological data provide resolution at different levels in *Andira*. *Syst. Biol.* **45**, 496–515.

Pennington, R. T., Cronk, Q. C. B. & Richardson, J. A. 2004 Introduction and synthesis: plant phylogeny and the origin of major biomes. *Phil. Trans. R. Soc. B* **359**, 1455–1464. (doi:10.1098/rstb.2004.1539)

Peters, J. L., McCraken, K. G., Zhuravlev, Y. N., Wilson, R. E., Johnson, K. P. & Omland, K. E. 2005 Phylogenetics of wigeons and allies (Anatidae: Aves): the importance of sampling multiple loci and multiple individuals. *Mol. Phylog. Evol.* **35**, 209–224. (doi:10.1016/j.ympev.2004.12.017)

Rauscher, J. T., Doyle, J. J. & Brown, A. D. H. 2002 Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploidy complex. *Mol. Ecol.* **11**, 2691–2702. (doi:10.1046/j.1365-294X.2002.01640.x)

Raymond, O., Piola, F. & Sanlaville-Boisson, C. 2002 Inferences of reticulation in outcrossing allopolyploid taxa: caveats, likelihood and perspectives. *Trends Ecol. Evol.* **17**, 3–6. (doi:10.1016/S0169-5347(01)02378-3)

Ree, R. H., Citerne, H. L., Lavin, M. & Cronk, Q. C. B. 2004 Heterogeneous selection on *LEGCYC* paralogs in relation to flower morphology and the phylogeny of *Lupinus* (Leguminosae). *Mol. Biol. Evol.* **21**, 321–331. (doi:10.1093/molbev/msh022)

Richardson, J. E., Pennington, R. T., Pennington, T. D. & Hollingsworth, P. M. 2001 Rapid diversification of a species-rich genus of Neotropical rainforest trees. *Science* **293**, 2242–2245. (doi:10.1126/science.1061421)

Rokas, A., Williams, B. L., King, N. & Carroll, S. B. 2003 Genome scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804. (doi:10.1038/nature02053)

Sakai, M., Kanazawa, A., Fujii, A., Thseng, F. S., Abe, J. & Shimamoto, Y. 2003 Phylogenetic relationships of the chloroplast genomes in the genus *Glycine* inferred from four intergenic spacer regions. *Plant Syst. Evol.* **239**, 29–54. (doi:10.1007/s00606-002-0226-9)

Sanderson, M. J. 1995 Objections to bootstrapping phylogenies: a critique. *Syst. Biol.* **44**, 299–320.

Sanderson, M. J. & Driskell, A. C. 2003 The challenge of constructing large phylogenetic trees. *Trends Plant Sci.* **8**, 374–379. (doi:10.1016/S1360-1385(03)00165-1)

Sang, T. 2002 Utility of low-copy nuclear gene sequences in plant phylogenetics. *Crit. Rev. Biochem. Mol. Biol.* **37**, 121–147. (doi:10.1080/10409230290771474)

Sang, T., Crawford, D. J. & Stuessy, T. F. 1995 Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proc. Natl Acad. Sci.* **92**, 6813–6817.

Sang, T., Donoghue, M. J. & Zhang, D. 1997 Evolution of alcohol dehydrogenase genes in peonies (*Paeonia*): phylogenetic relationships of putative nonhybrid species. *Mol. Biol. Evol.* **14**, 994–1007.

Sanjur, O. T., Piperno, D. R., Andres, T. C. & Wessel-Beaver, L. 2002 Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: implications for crop evolution and areas of origin. *Proc. Natl Acad. Sci.* **99**, 535–540. (doi:10.1073/pnas.012577299)

Savolainen, V. & Chase, M. W. 2003 A decade of progress in plant molecular phylogenies. *Trends Genet.* **19**, 717–724. (doi:10.1016/j.tig.2003.10.003)

Scherson, R., Choi, H.-K., Cook, D. & Sanderson, M. J. In press. Phylogenetics of New World *Astragalus*: the utility of genomics technology in reconstructing phylogenies at low taxonomic levels. *Brittonia*.

Schultheis, L. M. & Baldwin, B. G. 1999 Molecular phylogenetics of Fouquieriaceae: evidence from nuclear rDNA ITS studies. *Am. J. Bot.* **86**, 578–589.

Senchina, D. S. *et al.* 2003 Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol. Biol. Evol.* **20**, 633–643. (doi:10.1093/molbev/msg065)

Shaw, A. J., Cox, C. J. & Boles, S. B. 2003 Polarity of peatmoss (*Sphagnum*) evolution: who says bryophytes have no roots. *Am. J. Bot.* **90**, 1777–1787.

Shaw, A. J., Cox, C. J. & Boles, S. B. 2005*a* Phylogeny, species delimitation and recombination in *Spahgnum* Section Acutifolia. *Syst. Bot.* **30**, 16–33. (doi:10.1600/0363644053661823)

Shaw, J. & Small, R. L. 2004 Addressing the 'hardest puzzle in American pomology': phylogeny of *Prunus* Sect. Prunocerasus (Rosaceae) based on seven non-coding chloroplast DNA regions. *Am. J. Bot.* **91**, 985–996.

Shaw, J. *et al.* 2005*b* The tortoise and the hare II: relative utility of 21 non-coding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* **92**, 142–166.

Sieburth, L. E. & Meyerowitz, E. M. 1997 Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. *Plant Cell* **9**, 355–365. (doi:10.1105/tpc.9.3.355)

Simmons, M. P. & Ochoterena, H. 2000 Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **49**, 369–381. (doi:10.1080/10635159950173889)

Slowinski, J. B. & Page, R. D. M. 1999 How should species phylogenies be inferred from sequence data. *Syst. Biol.* **48**, 814–825. (doi:10.1080/106351599260030)

Small, R. L. 2004 Phylogeny of *Hibiscus* sect. Muenchhusia (Malvaceae) based on chloroplast rpl16 and ndhF and nuclear ITS and GBSSI sequences. *Syst. Bot.* **29**, 385–392. (doi:10.1600/036364404774195575)

Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T. & Wendel, J. F. 1998 The tortoise and the hare: choosing between non-coding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group. *Am. J. Bot.* **85**, 1301–1315.

Small, R. L., Cronn, R. C. & Wendel, J. F. 2004 Use of nuclear genes for phylogeny reconstruction in plants. *Aust. Syst. Bot.* **17**, 145–170. (doi:10.1071/SB03015)

Soltis, D. E. & Soltis, P. S. 1998 Choosing an approach and an appropriate gene for phylogenetic analysis. In *Molecular systematics of plants II. DNA sequencing* (ed. D. E. Soltis, P. S. Soltis & J. J. Doyle), pp. 1–42. Boston, MA: Kluwer Academic.

Soltis, P. S. & Soltis, D. E. 2001 Molecular systematics: assembling and using the Tree of Life. *Taxon* **50**, 663–677.

Soltis, P. S. & Soltis, D. E. 2004 The origin and diversification of angiosperms. *Am. J. Bot.* **91**, 1614–1626.

Soltis, P. S., Soltis, D. E. & Chase, M. W. 1999 Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**, 402–404. (doi:10.1038/46528)

Soltis, D. E. *et al.* 2004 Genome-scale data, angiosperm relationships, and 'ending incongruence': a cautionary tale in phylogenetics. *Trends Plant Sci.* **9**, 477–483. (doi:10.1016/j.tplants.2004.08.008)

Strand, A. E., Leebens-Mack, J. & Milligan, B. G. 1997 Nuclear DNA-based markers for plant evolutionary biology. *Mol. Ecol.* **6**, 113–118. (doi:10.1046/j.1365-294X.1997.00153.x)

Syring, J. V., Willyard, A., Cronn, R. C. & Liston, A. In press. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *Am. J. Bot.*

Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. 1991 Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **17**, 1105–1110. (doi:10.1007/BF00037152)

Tank, D. C. & Sang, T. 2001 Phylogenetic utility of the glycerol-3-phosphate acyltransferase gene: evolution and implications in *Paeonia* (Paeoniaceae). *Mol. Phylog. Evol.* **19**, 421–429. (doi:10.1006/mpev.2001.0931)

Thießen, G., Becker, A., Winter, K.-U., Münster, T., Kirchner, C. & Saedler, H. 2002 How the land plants learned their floral ABCs: the role of MADS-box genes in the evolutionary origin of flowers. In *Developmental genetics and plant evolution* (ed. Q. C. B. Cronk, R. M. Bateman & J. A. Hawkins), pp. 173–205. London: Taylor & Francis.

Wang, R. L., Stec, A., Hey, J., Lukens, L. & Doebley, J. 1999 The limits of selection during maize domestication. *Nature* **398**, 236–239. (doi:10.1038/18435)

Watanabe, M. 2002 Describing the 'Tree of Life': attainable goal or stuff of dreams? *BioScience* **52**, 875–880.

Weeks, A. & Simpson, B. B. 2004 Molecular genetic evidence for interspecific hybridization among endemic Hispaniolan *Bursera* (Burseraceae). *Am. J. Bot.* **91**, 976–984.

Wendel, J. F. & Doyle, J. J. 1998 Phylogenetic incongruence: window into genome history and molecular evolution. In *Molecular systematics of plants II. DNA sequencing* (ed. D. E. Soltis, P. S. Soltis & J. J. Doyle), pp. 265–296. Boston, MA: Kluwer Academic.

Wendel, J. F., Schnabel, A. & Seelanan, T. 1995 Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl Acad. Sci.* **92**, 280–284.

Wenzel, J. W. & Siddall, M. E. 1999 Noise. *Cladistics* **15**, 51–64. (doi:10.1111/j.1096-0031.1999.tb00394.x)

Winkworth, R. C. & Donoghue, M. J. 2004 *Viburnum* phylogeny: evidence from the duplicated nuclear gene GBSSI. *Mol. Phylog. Evol.* **33**, 109–126. (doi:10.1016/j.ympev.2004.05.006)

Wojciechowski, M. F., Sanderson, M. J. & Hu, J. M. 1999 Evidence on the monophyly of *Astragalus* and its major subgroups based on nrDNA ITS and cpDNA trnL intron data. *Syst. Bot.* **24**, 408–437.

Wolfe, A. D. & Liston, A. 1998 Contributions of PCR-based methods to plant systematics and evolutionary biology. In *Molecular systematics of plants II. DNA sequencing* (ed. D. E. Soltis, P. S. Soltis & J. J. Doyle), pp. 43–86. Boston, MA: Kluwer Academic.

Wolfe, K. H., Li, W.-H. & Sharp, P. M. 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proc. Natl Acad. Sci.* **84**, 9054–9058.

Wortley, A. H. & Scotland, R.W. 2005 Determining the potential utility of datasets for phylogeny reconstruction. *Taxon*. In press.

Wortley, A. H., Rudall, P. J., Harris, D. J. & Scotland, R. W. 2005 How much data is needed to resolve a difficult phylogeny? Case study in Lamiales. *Syst. Biol.* **54**, 697–709.

Xu, W., Briggs, W. J., Padolina, J., Timme, R. E., Liu, W., Linder, C. R. & Miranker, D. P. 2004 Using MoBIoS scalable genome join to find conserved primer pair candidates between two genomes. *Bioinformatics* **20**(Suppl. 1), i355–i362. (doi:10.1093/bioinformatics/bth929)

Yang, Z. 1998 On the best evolutionary rate for phylogenetic analysis. *Syst. Biol.* **47**, 125–133. (doi:10.1080/106351598261067)

Yates, T. L., Salazar-Bravo, J. & Dragoo, J. W. 2004 The importance of the Tree of Life to society. In *Assembling the Tree of Life* (ed. J. Cracraft & M. J. Donoghue), pp. 7–17. New York: Oxford University Press.

Yockteng, R. & Nadot, S. 2004 Infrageneric phylogenies: a comparison of chloroplast-expressed glutamine synthetase, cytosol-expressed glutamine synthetase and cpDNA maturase K in *Passiflora*. *Mol. Phylog. Evol.* **31**, 397–402. (doi:10.1016/S1055-7903(03)00276-8)

GLOSSARY

AFLP: amplified fragment length polymorphism
CATS: comparative anchor tagged sequences
CFI: consensus fork index
COS: conserved orthologue set
EMPT: equally most parsimonious tree
EST: expressed sequence tag
ITS: internal transcribed spacer
LCNG: low-copy nuclear gene
PI: parsimony informative
RAPD: randomly amplified polymorphic DNA
SCAR: sequence characterized amplified region
UTR: untranscribed region

Appendix 3

Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes

Colin Hughes* and Ruth Eastwood

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, United Kingdom

Edited by Peter R. Crane, Royal Botanic Gardens, Kew, Surrey, United Kingdom, and approved May 19, 2006 (received for review March 9, 2006)

Species radiations provide unique insights into evolutionary processes underlying species diversification and patterns of biodiversity. To compare plant diversification over a similar time period to the recent cichlid fish radiations, which are an order of magnitude faster than documented bird, arthropod, and plant radiations, we focus on the high-altitude flora of the Andes, which is the most species-rich of any tropical mountains. Because of the recent uplift of the northern Andes, the upland environments where much of this rich endemic flora is found have been available for colonization only since the late Pliocene or Pleistocene, 2–4 million years (Myr) ago. Using DNA sequence data we identify a monophyletic group within the genus *Lupinus* representing 81 species endemic to the Andes. The age of this clade is estimated to be 1.18–1.76 Myr, implying a diversification rate of 2.49–3.72 species per Myr. This exceeds previous estimates for plants, providing the most spectacular example of explosive plant species diversification documented to date. Furthermore, it suggests that the high cichlid diversification rates are not unique. Lack of key innovations associated with the Andean *Lupinus* clade suggests that diversification was driven by ecological opportunities afforded by the emergence of island-like habitats after Andean uplift. Data from other genera indicate that lupines are one of a set of similarly rapid Andean plant radiations, continental in scale and island-like in stimulus, suggesting that the high-elevation Andean flora provides a system that rivals other groups, including cichlids, for understanding rapid species diversification.

leguminosae | *Lupinus* | phylogeny | species diversification

Studies of rapid episodes of species diversification or species radiations have provided a continuously rich source of new insights into the evolutionary processes underlying diversification and modern patterns of biodiversity for the last 150 years (1–6). A striking feature of species radiations is the discrepancy between the very high rates of species diversification reported for cichlid fish radiations in east African rift lakes and all bird, arthropod and plant radiations (2, 7, 8). This discrepancy has been attributed in part to the recency of the fish radiations (7). Accurate measurement of peak episodes of diversification embedded within older radiations requires fully sampled and resolved phylogenies (9). In the absence of such phylogenies, measurements of diversification rates for most radiations are less precise because they average out episodes of faster and slower speciation. Furthermore, these approaches measure net diversification rates, and the effects of extinction are not assessed. If speciation is concentrated in the early phases of radiations (3, 7, 10, 11), this could explain the discrepancy between the exceptional rates of species diversification reported for very recent [<2 million years (Myr)] lacustrine fish and other documented bird, arthropod, or plant radiations, which are generally older (>5 Myr) (7). Direct comparisons with the cichlid fish diversification rates have been lacking because few comparably recent radiations have been found in other groups.

To compare plant diversification over a similarly short time period to the very recent fish radiations, we examined patterns of species diversification in the high-altitude flora of the northern Andes, which is by far the most species-rich of any tropical mountain massif (12). It forms part of the tropical Andean biodiversity hotspot, which contains an estimated 45,000 plant species, 44% of which are endemic (13). Endemism is even higher, reaching 60%, in the high-altitude north Andean páramos (14). The cold upland habitats where much of this rich endemic flora is found today have been available for plant colonization only since the late Pliocene or early Pleistocene 2–4 Myr ago (15–18) after final uplift of the northern Andes (15). This recent uplift implies that at least some Andean plant groups must have diversified very recently and rapidly (16, 19). Despite this presumption, there are few reliable estimates of species diversification rates for Andean groups because of uncertainty about species numbers and the lack of robust, well resolved, and densely sampled phylogenies (20, 21).

The prevalence of north and south temperate genera in the high-altitude Andean flora (12) supports the idea that colonization of the emerging high Andes was comparable to colonization of a newly formed island or island archipelago (16, 22) (Fig. 1). On recently formed islands and island-like formations such as lakes and mountains, rapid diversification has been attributed to ecological opportunities afforded by the availability of new habitats and absence of competition (1, 3, 7, 8, 10, 11, 23). Conversely, on continents, where ecological opportunity is harder to demonstrate, rapid episodes of diversification have generally been associated with key morphological or physiological innovations (4, 11). Numerous north temperate plant genera, such as *Alnus*, *Draba*, *Lupinus*, *Quercus*, *Salix*, *Sambucus*, *Valeriana*, and *Viburnum*, are postulated to have arrived in the Andes from North America after uplift of the northern Andes (17, 18). The altitudinal distribution of Andean species diversity for these genera shows that diversification is restricted largely to those growing at high elevations (Fig. 2). Although the monophyly of most of these endemic Andean species flocks has not been tested, this pattern supports the idea that island-like opportunities for diversification were available above the tree line in the high-altitude grassland zone and that plants that were preadapted to cooler conditions were most able to exploit such opportunities (Fig. 2).

The Genus *Lupinus*

To gain insight into colonization of the recently formed high-elevation Andean habitats, we examined species diversification in the legume genus *Lupinus*, which comprises ≈275 species of annual herbs and herbaceous and woody perennials with an amphiatlantic distribution. The majority of *Lupinus* species occur in the New World, with two main centers of species diversity in western North America (≈100 species) and the Andes (≈85 species). Guided by previous estimates of *Lupinus*

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: Myr, million years; ITS, internal transcribed spacer.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ524181–DQ524328 and DQ529744–DQ529979).

*To whom correspondence should be addressed. E-mail: colin.hughes@plants.ox.ac.uk.

© 2006 by The National Academy of Sciences of the USA

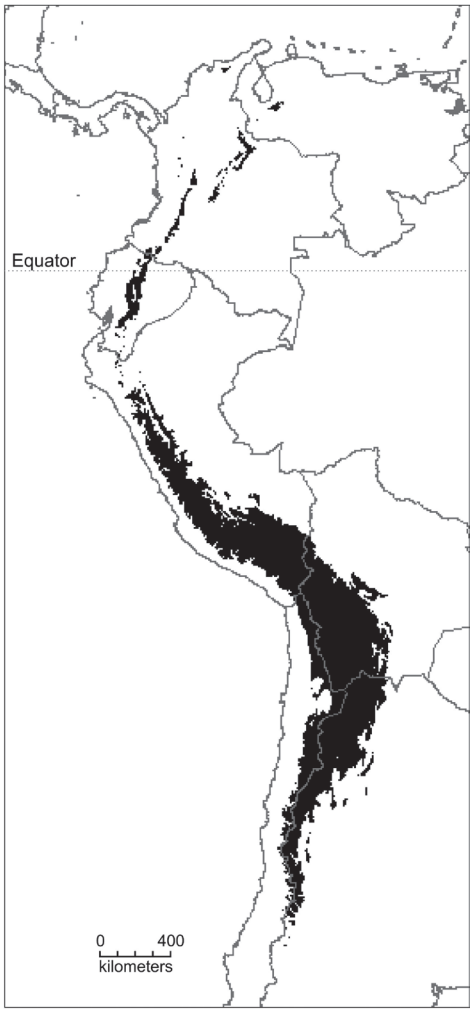


Fig. 1. The Andes as an island archipelago. This map shows the distribution of land over 3,000 m elevation, corresponding to the lower limit of the high-altitude grassland (páramo, jalca, and puna) vegetation zone. Although a few Andean *Lupinus* species occur at lower elevations, the genus is most diverse in the 3,000- to 4,000-m zone. *Lupinus* species are ubiquitous in all of the high-elevation ''islands'' from Venezuela and the Sierra Nevada Santa Marta in northern Colombia in the north to northwest Argentina in the south, and the geographical extent of the Andean *Lupinus* radiation (Fig. 3, k) closely matches the overall area of the Andean ''archipelago'' as delimited here.

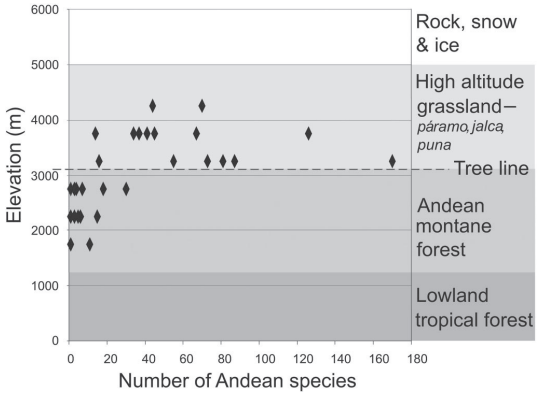


Fig. 2. Altitudinal variation in Andean species diversification. Data points represent plant genera with putative north temperate origins (see *Methods*). Diversification is concentrated above 3,000 m in the high-altitude grassland vegetation zone (Fig. 1).

phylogeny based on internal transcribed spacer (ITS) sequences (24), we generated a densely sampled, well resolved, and robust phylogeny for *Lupinus* based on nuclear ITS and *CYCLOIDEA* gene sequences (Fig. 3). Our analysis shows that the Andean species flock is not monophyletic. Instead, Andean species are present in each of two well supported New World lineages. Nested within a large western New World group (Fig. 3, clade B) that spans North America, Mexico, and the Andes is a well supported clade representing ≈81 Andean species that are distributed from Venezuela to Argentina (Fig. 3, k). Four species distributed in the south-central Andes as far north as Lake Titicaca fall outside this clade. These species are placed in a predominantly lowland eastern New World lineage (Fig. 3, clade A). Species from both clades occur in sympatry in Bolivia and northern Argentina (Fig. 3, k). These distribution patterns suggest that there were at least two independent incursions of *Lupinus* into the Andes.

Species belonging to the main Andean clade exhibit diverse life forms (Fig. 3, a–j) and ecologies. The majority of species grow in the 3,000- to 4,000-m high-elevation grassland (páramo, jalca, and puna) zone spanning wet eastern Andean ranges and drier inter-Andean valley and Pacific slopes (Fig. 1). A subset of prostrate (Fig. 3, b and j) and acaulescent (Fig. 3, e, g, and i) species are adapted to the extreme freeze–thaw conditions found at higher elevations between 4,000 and 5,200 m. Although predominantly a high-elevation group, a small number of mainly annual species adapted to lower elevation (<3,000 m) drier environments are included in the main Andean clade, including one ephemeral annual species, *Lupinus mollendoensis*, that grows in the coastal desert of southern Peru.

Results and Discussion

Independent dates derived from evolutionary rates analysis of DNA sequence data and paleobotanical evidence support the idea that the main Andean *Lupinus* radiation is very recent. The well supported monophyly of the large Andean *Lupinus* clade (parsimony bootstrap 92%, Bayesian posterior probability 1.0) and its placement as sister group to a clade of Mexican species suggest that diversification of this clade followed a single colonization from North America. The low sequence divergence and lack of resolution within the large Andean clade in comparison to the rest of the *Lupinus* phylogeny (Fig. 3) suggest rapid and recent diversification in the Andes. To test this, an evolutionary rates analysis of the combined ITS and *LEGCYCIA* sequence data sets was carried out. This analysis estimated the age of the Andean *Lupinus* radiation to be 1.47 ± 0.29 Myr. Notably, this agrees with extensive palynological evidence (17) that suggests cold upland habitats first appeared 2–4 Myr ago, after the final uplift of the northern Andes 3–5 Myr ago (15). Age estimates for the Andean radiation are relatively insensitive to older calibration dates for the *Lupinus* stem node. For example, using the maximum age of 21.16 Myr (25) in place of the mean of 16.01

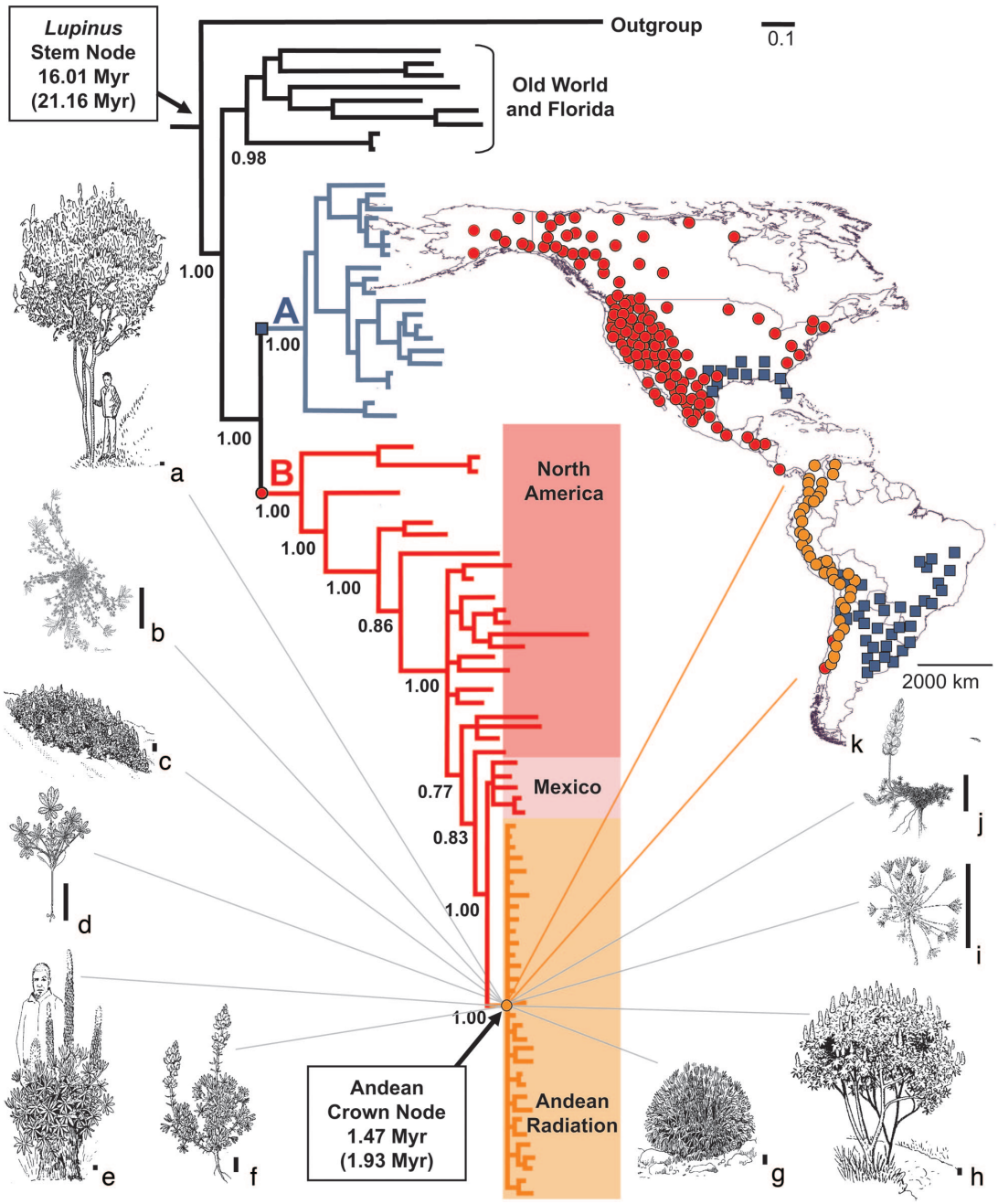


Fig. 3. Phylogeny of *Lupinus*. Fifty percent majority rule Bayesian tree from analysis of the combined ITS/*LEGCYCIA* data sets. Posterior probabilities of major clades are shown below nodes; all nodes had posterior probabilities >0.5. This tree is congruent with the strict consensus derived from parsimony analysis. For clarity, taxon names have been omitted (see supporting information). Branch lengths are proportional to changes on the tree. A and B denote two well supported New World clades: A, eastern New World; B, western North America, Mexico, and the Andes. The map shows the distributions of these two New World clades to be largely allopatric, with limited overlap in the southern U.S. and the south-central Andes. The geographical extent of the main Andean radiation is shown in orange. Chilean accessions of *Lupinus microcarpus* group with their North American counterparts at the base of clade B, but the occurrence of *L. microcarpus* in central Chile (shown in red) is doubtfully native. Line drawings illustrate life forms encompassed by species in the Andean *Lupinus* radiation: a, treelet, *Lupinus semperflorens*; b, prostrate herb, *Lupinus* sp. *nov*; c, perennial woody shrublet, *Lupinus smithianus*; d, ephemeral annual herb, *L. mollendoensis*; e, giant stem rosette, *Lupinus weberbaueri*; f, woody perennial shrub, *Lupinus* sp. *nov*; g, acaulescent rosette, *Lupinus nubigenus*; h, perennial woody shrublet, *Lupinus* sp. *nov*; i, dwarf acaulescent rosette, *Lupinus pulvinaris*; j, prostrate herb, *Lupinus prostratus*. (Scale bars: 5 cm.)

Myr for the stem node only increases the estimated age of the Andean radiation by 0.46 Myr to 1.93 ± 0.35 Myr.

The average per-lineage species diversification rate for the Andean *Lupinus* radiation is 2.50–3.72 species per Myr. Using the older age estimate of 1.93 Myr for the Andean node yields a more conservative estimate of 1.93–2.78 species per Myr. However, even this minimum estimate exceeds reported species diversification rates for plants either on islands (10, 23) or continents (4, 6, 11, 26), providing the most spectacular example of explosive species diversification in plants documented to date. Low sequence divergence, lack of resolution and support, and/or sparse taxon sampling in phylogenetic studies of other high-elevation Andean plant groups (20, 21, 27, 28) mean that the monophyly of other putative Andean plant radiations has not been rigorously demonstrated. Despite these difficulties, two recent studies estimate approximate diversification rates of 0.8–1.34 species per Myr for Andean *Valeriana* (20) and 1.71–3.2 species per Myr for Andean *Gentianella* (21). These high diversification rates, as well as high endemic Andean species diversity in other high-elevation plant genera, such as *Draba*, *Espeletia*, *Huperzia*, *Hypericum*, and *Lysipomia*, suggest that lupines are just one of a set of spectacular, but as yet poorly documented, plant radiations that followed the final uplift of the northern Andes, and that a high proportion of the 3,400 plant species recorded in the northern Andean páramos (14) have evolved in the very recent evolutionary past.

It is notable that there are no novel morphological or physiological traits obviously associated with the Andean *Lupinus* clade that might have precipitated rapid diversification. Although the existence of such a key innovation cannot be ruled out pending more detailed anatomical and physiological analyses, this finding suggests that lupine diversification was driven by ecological opportunities similar to those on islands created by the emergence of largely unoccupied habitats after Andean uplift and subsequent Pleistocene glaciation. This idea is reinforced by the observation that the great diversity of life forms displayed by Andean lupines (Fig. 3, a–j) mirrors that seen in many plant radiations on islands (10, 23). Accelerated morphological divergence of this sort is often associated with island radiations and has been attributed to opportunities provided by depauperate environments where competitors are fewer (1). We identify four factors in the Andes that may explain the exceptional *Lupinus* diversification rate compared with other island radiations. First, the continental scale of the Andean *Lupinus* radiation, spanning >4,000 km from Venezuela to Argentina (Figs. 1 and 3, k), far exceeds typical island radiations, which generally occupy much more confined geographical areas. This scale creates extensive opportunities for geographical isolation, which is thought to be one of the main factors underlying the higher species diversification rates found on larger islands (29). Second, repeated fragmentation of high-altitude habitats due to altitudinal shifts in Andean vegetation zones prompted by Quaternary climate fluctuations are likely to have contributed to geographical isolation and diversification in the Andes (16, 17, 22, 30). Third, the extremely dissected topography in the high Andes and steep tropical montane environmental gradients associated with altitude (31) further enhance opportunities for geographical isolation that is evident from the prevalence of narrowly restricted endemics among the Andean *Lupinus* species today. Finally, resource heterogeneity in terms of habitats and environmental conditions is very high in the mid- to high-elevation Andes.

For plants, the idea that island-type diversifications are faster than mainland radiations has been hotly debated without conclusion (4, 6, 10, 11, 23, 26). The exceptional rates of diversification found for high-elevation Andean plants suggest that island-like diversification driven by ecological opportunity may indeed outpace mainland-type radiations. This result adds to the growing weight of evidence from diverse groups, including fish (2, 7, 8), plants (4, 6), and birds (5), that opportunity, in the form of extrinsic circumstances and events rather than evolutionary innovation, plays a dominant role in driving peak episodes of species diversification.

The age of the Andean *Lupinus* clade and our estimate of diversification rate are comparable to those previously reported for cichlid fish in east African lakes, which show the highest diversification rate estimates across the tree of life (2, 7). This result demonstrates that the high fish diversification rates are not unique and that radiations measured over similarly short time periods can match these rates. Clearly, much more densely sampled and robust phylogenies are needed to accurately measure peak episodes of diversification embedded within older radiations, as, for example, in ref. 32. The overall *Lupinus* phylogeny (Fig. 3) is highly resolved and robustly supported because of the fact that the *LEGCYC1A* locus is among the most rapidly evolving DNA sequence regions used to resolve plant species relationships. However, our gene trees show little resolution among the Andean species despite the diversity of life forms (Fig. 3, a–j), environmental adaptations, and broad geographic span encompassed by the radiation (Figs. 1 and 3, k). This lack of resolution suggests that the branching order for such rapidly evolving plant radiations will be difficult to resolve using conventional DNA sequence data (33), just as it has been for east African cichlid fish (34). As we unravel and document the history of plant diversification in the high Andes we can expect similarly heated debate about the monophyly and exact timing of radiations to that surrounding the cichlid fish radiations (2, 7, 8, 34). However, it is clear that the high-elevation Andean flora, comprising several thousand plant species, provides a system that rivals the most charismatic documented radiations, including cichlid fish, in terms of opportunities for understanding explosive species diversification.

Methods

Andean Species Diversity. Altitudinal patterns of extant Andean species diversity for 31 plant genera with putative north temperate (16–18) or north Andean origins are summarized in Fig. 2. Data on numbers of Andean species were assembled from widely scattered sources, including taxonomic monographs, species checklists, and, in some cases, studies that identify monophyletic groups. Representative Andean altitudinal ranges are derived from the Checklist of Ecuador (35) in the form of frequency distributions of numbers of species present in 500-m elevation bands. The mid-point of the elevation band with maximum species diversity is plotted for each genus. The genera included are *Alnus*, *Ambrosia*, *Bartsia*, *Berberis*, *Boehmeria*, *Buddleja*, *Calamagrostis*, *Cornus*, *Diplostephium*, *Draba*, *Geranium*, *Gentianella*, *Halenia*, *Huperzia*, *Hypericum*, *Lachemilla*, *Lupinus*, *Lysipomia*, *Myrica*, *Niphogeton*, *Prunus*, *Quercus*, *Rhamnus*, *Ribes*, *Salix*, *Sambucus*, *Senecio*, *Vaccinium*, *Valeriana*, and *Viburnum.*

Species Delimitation. Our estimate of 85 Andean species of *Lupinus* is based on comprehensive reexamination of species boundaries as part of taxonomic work that aims to bring consistency of species delimitation across the genus. This work has involved 8 months' fieldwork in the Andes, examination of 8,000 herbarium sheets of Andean material, and observation of greenhouse-grown material. With 480 *Lupinus* species names based on Andean plants and recent estimates of as many as 170 species in the Andes (18), our estimate represents a substantial rationalization of the chaos surrounding the taxonomy of Andean lupines. Furthermore, our estimate is conservative in recognizing variation at infraspecific rank within widespread polymorphic species. Andean species not included in the analysis were assigned to the two New World clades based on morphology and

chromosome number, which differ consistently between these groups (our unpublished data).

Sequence Data and Phylogenetic Analysis. One hundred forty-eight accessions representing 98 species covering previously recognized *Lupinus* clades (24), the geographic range, and eight outgroup species representing five genera from tribe Genisteae were sampled. To test the monophyly of the Andean species, 53 accessions representing 36 species were sampled spanning the range of life forms (Fig. 3, a–j), geography, and altitudes. (Fully annotated parsimony and Bayesian trees, locality and voucher details, and GenBank accession nos. are available in supporting information, which is published on the PNAS web site.) The 5.8S subunit and flanking ITS (ITS1+ITS2) of nuclear ribosomal DNA and one copy (*LEGCYCIA*) of the rapidly evolving regulatory gene *CYCLOIDEA* were sequenced by using standard protocols. Two copies of *CYCLOIDEA*-like *LEGCYCI* genes have been characterized and sequenced in the Genistoid legumes including *Lupinus* (36, 37). These sequences were used to design complementary locus-specific primers located between the conserved TCP and R domains (38), which can be used in combination with general *LEGCYCI* primers to amplify the entire *LEGCYCIA* ORF of ≈1,000 bp in two fragments. An additional pair of locus-specific primers (available upon request) was designed to amplify *LEGCYCIA* for taxa in the lowland eastern New World clade (Fig. 3, clade A), which have a deletion at the original locus-specific primer site. The *LEGCYCIA* paralogue was chosen because it is more variable than *LEGCYCIB* (37, 38) and, unlike *LEGCYCIB*, does not show allelic length variation. ITS and *LEGCYCIA* sequences from Ree *et al.* (37) were included in the analysis. New sequences have been submitted to the GenBank database (see supporting information). Sequences were aligned by using CLUSTALX (39) and manually adjusted. Indels were coded by using the simple gap-coding method in SEQSTATE (40) and included in the parsimony analysis. Parsimony analysis of the combined ITS/*LEGCYCIA* data set was conducted with NONA (41) by using 1,000 random addition sequences, tree bisection, and reconnection, holding 100 trees per replication and attempting to swap to completion (see supporting information). A reduced data set of 89 sequences representing one outgroup and one sequence per species was compiled for Bayesian analysis (42) using the GTR+G and GTR+I models (43) for the ITS and *LEGCYCIA* data partitions, respectively. This analysis involved 3,000,000 Metropolis-coupled Markov Chain Monte Carlo permutations of tree parameters and four chains heated to 0.01. The consensus Bayesian tree was virtually congruent with the parsimony tree but slightly more resolved (see supporting information). Clade stability tests involved Bayesian posterior probabilities and parsimony bootstrap resampling.

Evolutionary Rates Analysis. The relative rate test implemented in MR BAYES rejected a molecular clock for the consensus Bayesian tree (likelihood ratio = 3313.72, df = 142, $P = 0.00$). The program R8S (44) was used to assess variation in substitution rates and incorporate this into the estimation of ages of lineages. Branching order and branch lengths from 100 Bayesian trees sampled every 10,000 generations after stationarity were analyzed to obtain means and standard deviations of ages of clades. Cross-validation and penalized likelihood analysis yielded an average smoothing parameter of 13, which was used in final dating analyses. Relative ages estimated with R8S were converted to absolute ages by fixing the age of the *Lupinus* stem node using mean (16.01 Myr) and maximum (21.16 Myr) dates for the most recent common ancestor of *Spartium* and *Lupinus* from the legume-wide analysis of *matK* sequences calibrated using 13 well studied fossils (25).

Species Diversification Rate. Species diversification rates, assuming an equal rate of random speciation Yule model, were calculated as $SR = (ln\ n_1 - ln\ n_0)/t$, where n_1 is the number of extant species, n_0 is the initial species diversity, here taken as 1, and t is time in Myr. Upper and lower standard deviations of age estimates were used in calculations of speciation rate.

We thank Helene Citerne for primers; Rosemary Wise for artwork; Simon Ho and Matt Lavin for advice; Matt Lavin for age estimates based on but not presented in ref. 25; Donovan Bailey, Stephan Beck, Tim Budden, Ruth Clark, Aniceto Daza, Alfonso Delgado, Chris Fagg, Rob Forrest, Martin Gardner, Greg Kenicer, Bente Klitgaard, Gwil Lewis, Helga Ochoterena, David Neill, Arely Palabral, John Pannell, Terry Pennington, Toby Pennington, Carolyn Proenca, Carlos Reynel, Mario Sousa, Salvador Talavera, John Wood, Marty Wojciechowski, and especially Chris Drummond (University of Idaho, Moscow), Richard Spellenberg (New Mexico State University, Las Cruces), and Silvia Miotto and Maria-Teresa Schifino-Wittman (Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil) for assistance with fieldwork and provision of plant samples; the authorities in Ecuador, Peru, Bolivia, and Brazil for permission to collect material; the B, CAS, CGE, E, F, FHO, G, GH, ISC, K, LPB, M, MO, MOL, MSB, NY, QCNE, UB, UC, US, USM, and USZ herbaria (Index Herbariorum acronyms) for loan of material; and Julie Hawkins, Jane Langdale, John Pannell, Mike Sanderson, and an anonymous reviewer for comments. R.E. created Figs. 1 and 3. This work was supported by the Royal Society, Biotechnology and Biological Sciences Research Council, the Genetics Society, and the Stanley Smith Horticultural Trust.

1. Schluter, D. (2000) *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford).
2. Kocher, T. D. (2004) *Nat. Rev. Genet.* **5,** 288–298.
3. Gravilets, S. & Vose, A. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 18040–18045.
4. Kay, M. K., Reeves, P. A., Olmstead, R. G. & Schemske, D. W. (2005) *Am. J. Bot.* **92,** 1899–1910.
5. Ricklefs, R. E. (2003) *Proc. Biol. Sci.* **270,** 2285–2291.
6. Richardson, J. E., Pennington, R. T., Pennington, T. D. & Hollingsworth, P. M. (2001) *Science* **293,** 2242–2245.
7. McCune, A. R. (1997) in *Molecular Evolution and Adaptive Radiation*, eds. Givinish, T. J. & Sytsma, K. J. (Cambridge Univ. Press, Cambridge, U.K.), pp. 585–610.
8. Verheyen, E., Salzburger, W., Snoeks, J. & Meyer, A. (2003) *Science* **300,** 325–329.
9. Nee, S., Mooers, A. O. & Harvey, P. H. (1994) *Proc. Natl. Acad. Sci. USA* **89,** 8322–8326.
10. Baldwin, B. G. & Sanderson, M. J. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 9402–9406.
11. Klak, C., Reeves, G. & Hedderson, T. (2003) *Nature* **427,** 63–65.
12. Smith, J. M. B. & Cleef, A. M. (1988) *J. Biogeogr.* **15,** 631–645.
13. Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B. & Kent, J. (2000) *Nature* **403,** 853–858.
14. Luteyn, J. L. (1999) *Mem. N.Y. Bot. Gard.* **84,** 1–278.
15. Gregory-Wodzicki, K. M. (2000) *Geol. Soc. Am. Bull.* **112,** 1091–1105.
16. Simpson, B. B. (1975) *Paleobiology* **1,** 273–294.
17. Van der Hammen, T. & Cleef, A. M. (1986) in *High Altitude Tropical Biogeography*, eds. Vuilleumier, F. & Monasterio, M. (Oxford Univ. Press, New York), pp. 153–201.
18. Burnham, R. J. & Graham, A. (1999) *Ann. Mo. Bot. Gard.* **86,** 546–589.
19. Monasterio, M. & Sarmiento, L. (1991) *Trends Ecol. Evol.* **6,** 387–391.
20. Bell, C. D. & Donoghue, M. J. (2005) *Organisms Divers. Evol.* **5,** 147–159.
21. Von Hagen, K. B. & Kadereit, J. W. (2001) *Organisms Divers. Evol.* **1,** 61–79.
22. Vuilleumier, F. (1970) *Am. Nat.* **104,** 373–388.
23. Böhle, U.-T., Hilger, H. H. & Martin, W. F. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 11740–11745.
24. Ainouche, K. & Bayer, R. (1999) *Am. J. Bot.* **86,** 590–607.
25. Lavin, M. T., Herendeen, P. S. & Wojciechowski, M. F. (2005) *Syst. Biol.* **54,** 575–594.
26. Richardson, J. E., Weitz, F. M., Fay, M. F., Cronk, Q. C. B., Linder, H. P., Reeves, G. & Chase, M. W. (2001) *Nature* **412,** 181–183.
27. Koch, M. & Al-Shehbaz, I. A. (2002) *Ann. Mo. Bot. Gard.* **89,** 88–109.
28. Wojciechowski, M. F., Sanderson, M. J. & Hu, J.-H. (1999) *Syst. Bot.* **24,** 409–437.
29. Losos, J. & Schluter, D. (2000) *Nature* **408,** 847–850.

30. Hooghiemstra, H. & Van der Hammen, T. (2004) *Philos. Trans. R. Soc. London B Biol. Sci.* **359,** 173–181.
31. Janzen, D. H. (1967) *Am. Nat.* **101,** 233–249.
32. Kozak, K. H., Weisrock, D. W. & Larson, A. (2005) *Proc. R. Soc. London Biol. Sci.* **273,** 539–546.
33. Hughes, C. E., Eastwood, R. J. & Bailey, C. D. (2006) *Philos. Trans. R. Soc. London B Biol. Sci.* **361,** 211–225.
34. Kocher, T. D. (2003) *Nature* **423,** 489–491.
35. Jørgensen, P. M. & León-Yánes, S., eds. (1999) *Catalogue of the Vascular Plants of Ecuador* (Missouri Botanical Garden Press, St. Louis).
36. Citerne, H., Luo, D., Pennington, R. T., Coen, E. & Cronk, Q. C. B. (2003) *Plant Physiol.* **131,** 1042–1053.
37. Ree, R. H., Citerne, H. L., Lavin, M. T. & Cronk, Q. C. B. (2004) *Mol. Biol. Evol.* **21,** 321–331.
38. Citerne, H. L. (2006) *Edinburgh J. Bot.* **62,** 119–126.
39. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
40. Müller, K. (2005) *Appl. Bioinformatics* **4,** 65–69.
41. Goloboff, P. (2000) NONA: *A Tree Searching Program* (Fundación e Instituto Miguel Lillo Tucumán, Argentina).
42. Huelsenbeck, J. P. & Ronquist, F. (2001) *Bioinformatics* **17,** 754–755.
43. Nylander, J. A. A. (2004) MRMODEL TEST (Uppsala Univ., Uppsala), Version 2.
44. Sanderson, M. J. (2004) R8S *Users Manual* (Univ. of California, Davis), Version 1.7.

Appendix 4 Studies assessed to investigate the status of species-level phylogeny reconstruction

Authors	Year	Journal	Volume (Issue)
Rivadavia, F. *et al.*	2003	American Journal of Botany	90/1
Ford, V.S. & Gottlieb, L.D.	2003	American Journal of Botany	90/2
Schnabel, A. *et al.*	2003	American Journal of Botany	90/2
Andreasen, K. & Baldwin, B.G.	2003	American Journal of Botany	90/3
Hoggard, R.K. *et al.*	2003	American Journal of Botany	90/3
Little, D.P. & Barrington, D.S.	2003	American Journal of Botany	90/3
Samuel, R. *et al.*	2003	American Journal of Botany	90/3
Van den heede, C.J. *et al.*	2003	American Journal of Botany	90/3
Urbatsch, L.E. *et al.*	2003	American Journal of Botany	90/4
Aliscioni, S.S. *et al.*	2003	American Journal of Botany	90/5
Howarth, D.G. *et al.*	2003	American Journal of Botany	90/6
Muschner, V.C. *et al.*	2003	American Journal of Botany	90/8
Coleman, M. *et al.*	2003	American Journal of Botany	90/10
Moore, A.J. & Bohs, L.	2003	American Journal of Botany	90/11
Shaw, A. J. *et al.*	2003	American Journal of Botany	90/12
Hoggard, G.D. *et al.*	2004	American Journal of Botany	91/1
Nickrent, D.L. *et al.*	2004	American Journal of Botany	91/1
Williams, K.J. *et al.*	2004	American Journal of Botany	91/1
Mummenhoff, K. *et al.*	2004	American Journal of Botany	91/2
Sáchez-Baracaldo, P.	2004	American Journal of Botany	91/2
Schuettpelz, E. & Hoot, S.B.	2004	American Journal of Botany	91/2
Beardsley, P.M. *et al.*	2004	American Journal of Botany	91/3
Berry, P.E. *et al.*	2004	American Journal of Botany	91/4
Chat, J. *et al.*	2004	American Journal of Botany	91/5
Goldman, D.H. *et al.*	2004	American Journal of Botany	91/5
Moylan, E.C. *et al.*	2004	American Journal of Botany	91/5
Hoot, S.B. *et al.*	2004	American Journal of Botany	91/6
Mast, A.R. *et al.*	2004	American Journal of Botany	91/6
Otama, R.K. & Baum, D.A.	2004	American Journal of Botany	91/6
Renner, S. *et al.*	2004	American Journal of Botany	91/6
Scheen, A. *et al.*	2004	American Journal of Botany	91/6
Shaw, J. & Small, R.L.	2004	American Journal of Botany	91/6
Weeks, A. & Simpson	2004	American Journal of Botany	91/6
Acevedo-Rosas, R. *et al.*	2004	American Journal of Botany	91/7
Butterworth, C. & Wallace, R.S.	2004	American Journal of Botany	91/7
Carine, M.A. *et al.*	2004	American Journal of Botany	91/7
Graham, S. & Barrett, S.C.H.	2004	American Journal of Botany	91/7
Walker, J.B. *et al.*	2004	American Journal of Botany	91/7
Lihová, J. *et al.*	2004	American Journal of Botany	91/8
Miller, R.E. *et al.*	2004	American Journal of Botany	91/8
Hamzeh,M & Dayanandan,S.	2004	American Journal of Botany	91/9
Dane, F. & Lang, P.	2004	American Journal of Botany	91/11
Griffith, M.P.	2004	American Journal of Botany	91/11
Järvinen, P. *et al.*	2004	American Journal of Botany	91/11
Little, D.P. *et al.*	2004	American Journal of Botany	91/11
Wang, Y. *et al.*	2004	American Journal of Botany	91/11
Mansion, G. & Zeltner	2004	American Journal of Botany	91/12
Becerra, J.X.	2003	Molecular Phylogeny and Evolution	26/2
Ranker, T.A. *et al.*	2003	Molecular Phylogeny and Evolution	26/3
Valcarcel, V. *et al.*	2003	Molecular Phylogeny and Evolution	27/1
Hennequin, S. *et al.*	2003	Molecular Phylogeny and Evolution	27/2
Semerikov, V.L. *et al.*	2003	Molecular Phylogeny and Evolution	27/2
Pelser, P.B. *et al.*	2003	Molecular Phylogeny and Evolution	29/3
Oh, A. & Potter, D.	2003	Molecular Phylogeny and Evolution	29/2
Volkov, R.A. *et al.*	2003	Molecular Phylogeny and Evolution	29/2
Grob, G.B.J. *et al.*	2004	Molecular Phylogeny and Evolution	30/01
Schneeweiss, G.M. *et al.*	2004	Molecular Phylogeny and Evolution	30/2
Rajapakse, S. *et al.*	2004	Molecular Phylogeny and Evolution	30/3
Hao, G. *et al.*	2004	Molecular Phylogeny and Evolution	31/01
Yockteng, R. & Nadot, S.	2004	Molecular Phylogeny and Evolution	31/01
Vanderpoorten, A. *et al.*	2004	Molecular Phylogeny and Evolution	31/2
Wei, X. & Wang, X.	2004	Molecular Phylogeny and Evolution	31/2
Yamashiro, T. *et al.*	2004	Molecular Phylogeny and Evolution	31/2
Lee, C. & Wen, J.	2004	Molecular Phylogeny and Evolution	31/3
Plovanich, A.E. & Panero, J.L.	2004	Molecular Phylogeny and Evolution	31/3
Allan, G.J. *et al.*	2004	Molecular Phylogeny and Evolution	33/1
Beier, B.A. *et al.*	2004	Molecular Phylogeny and Evolution	33/1
Clarkson, J.J. *et al.*	2004	Molecular Phylogeny and Evolution	33/1
Winkworth, R.C. & Donoghue, M.J.	2004	Molecular Phylogeny and Evolution	33/1
Blattner, F.R.	2004	Molecular Phylogeny and Evolution	33/2
Manen, J.M. *et al.*	2004	Molecular Phylogeny and Evolution	33/2
Milne, R.I.	2004	Molecular Phylogeny and Evolution	33/2
Wilson, C.A.	2004	Molecular Phylogeny and Evolution	33/2
Barker, N.P. *et al.*	2004	Molecular Phylogeny and Evolution	33/3
Fan, C. *et al.*	2004	Molecular Phylogeny and Evolution	33/3
Roalson, E.H. & Friar, E.A.	2004	Molecular Phylogeny and Evolution	33/3

Rouhan, G. *et al.*	2004	Molecular Phylogeny and Evolution	33/3
Yi, T. *et al.*	2004	Molecular Phylogeny and Evolution	33/3
Boyd, A.E.	2003	Systematic Botany	28/1
Desrochers, A. M. & Dodge, B.	2003	Systematic Botany	28/1
Gaskin, J.F. & Schaal, B.	2003	Systematic Botany	28/1
Wilson, C.A.	2003	Systematic Botany	28/1
Bailey, C.D. *et al.*	2004	Systematic Botany	29/1
Bohs, L.	2004	Systematic Botany	29/1
Datwyler, S.L. & Wolfe, A.D.	2004	Systematic Botany	29/1
Gravendeel, B. *et al.*	2004	Systematic Botany	29/1
Roberts, R.P. & Urbatsch, L. E	2004	Systematic Botany	29/1
Schultheis, L.M. & Donoghue, M.J.	2004	Systematic Botany	29/1
Clark, J. L. & Zimmer, E.A.	2003	Systematic Botany	28/2
Fritsch, P.W.	2003	Systematic Botany	28/2
Haufler, C.H. *et al.*	2003	Systematic Botany	28/2
McDowell, T. *et al.*	2003	Systematic Botany	28/2
Zomlefer, W.B. *et al.*	2003	Systematic Botany	28/2
Allen, G.A. *et al.*	2003	Systematic Botany	28/3
Arias, S. *et al.*	2003	Systematic Botany	28/3
Beilstein, M.A. & Windham, M.D.	2003	Systematic Botany	28/3
Doyle, J.J. *et al.*	2003	Systematic Botany	28/3
Roalson, E.H. *et al.*	2003	Systematic Botany	28/3
Saar, D.F. *et al.*	2003	Systematic Botany	28/3
Downie, S.R. *et al.*	2004	Systematic Botany	29/3
Ingram, A.L. & Doyle, J.J.	2004	Systematic Botany	29/3
Lohwasser, U. *et al.*	2004	Systematic Botany	29/3
Mort, M.E. *et al.*	2004	Systematic Botany	29/3
Pelser, P.B. *et al.*	2004	Systematic Botany	29/3
Riley-Hulting, E.T. *et al.*	2004	Systematic Botany	29/3
Crisp, M.D. & Cook, L	2003	Systematic Botany	28/4
Freeman, C. E. *et al.*	2003	Systematic Botany	28/4
Gernandt, D.S. *et al.*	2003	Systematic Botany	28/4
Plana, V.	2003	Systematic Botany	28/4
Tate, J. & Simpson, B.B.	2003	Systematic Botany	28/4
Bunsawat, J. *et al.*	2004	Systematic Botany	29/4
Ickert-Bond, S.M. & Wojciechowski, M.F.	2004	Systematic Botany	29/4
Kita, Y. & Kato, M.	2004	Systematic Botany	29/4
Petersen, G. & Seberg, O.	2004	Systematic Botany	29/4
Simpson, B.B. *et al.*	2004	Systematic Botany	29/4
Wagstaff, S.J. & Breitwieser	2004	Systematic Botany	29/4
Guillon, J.	2004	Systematic Botany	29/2
Koch, M. & Al-Shehbaz, I.A.	2004	Systematic Botany	29/2
Koontz, J.A. *et al.*	2004	Systematic Botany	29/2
Schneider, H. *et al.*	2004	Systematic Botany	29/2
Small, R.L.	2004	Systematic Botany	29/2
Xia, Y. *et al.*	2004	Systematic Botany	29/2
Jamzad, Z. *et al*	2003	Taxon	52/1
Senters, A.E. & Soltis, D.E.	2003	Taxon	52/1
Seo, K.S. *et al.*	2003	Taxon	52/1
Roberts, R.P. & Urbatsch, L. E	2003	Taxon	52/2
Swenson, U. & Manns, U.	2003	Taxon	52/3
Långström, E. & Oxelman, B.	2003	Taxon	52/4
Bakker, F.T. *et al.*	2004	Taxon	53/1
Bolmgren, K. & Oxelman, B.	2004	Taxon	53/3
Guo, Y. *et al.*	2004	Taxon	53/3
Fuse, S. *et al.*	2004	Taxon	53/4
Gottschling, M. & Hilger, H.H.	2004	Taxon	53/4
Oberlander, K.C. *et al.*	2004	Taxon	53/4
Wu, M. & Hunag, T.	2004	Taxon	53/4

Appendix 5 Sequence data utilised, vouchers and Genbank numbers

Species	Origin	Collector Name & Number	Herbaria where vouchers are housed (including seedling vouchers)	ITS	GPAT1¤	GPAT2	LEGCYC1A 1st half	LEGCYC1A 2nd half	ITS & LEGCYC1A (#=BA)	3-locus (*=missing data)
Adenocarpus anagyrifolius	-	-	-	AF330639«	-	-	-	-	n	n
Amphithalea ericifolia	-	-	-	AJ310728¦	-	-	-	-	n	n
Anarthrophyllum capitata	-	M. Lavin 769	-	unpublished	-	-	-	-	n	n
Argyrocytisus battandieri	-	-	-	AF330638«	-	-	-	-	n	n
Argyrolobium zanonii	-	-	-	Z72274-5¥	-	-	-	-	n	n
Aspalathus cephalotes	-	-	-	Z72308-9¶	-	-	-	-	n	n
Chamaecytisus mollis	-	-	-	AF007472‡	-	-	-	-	n	n
Crotalaria podocarpa	-	-	-	AF007469‡	-	-	-	-	n	n
Genista anglica	Cornwall, UK	CEH2339	FHO, K	DQ524181	DQ852360¤		DQ529977	-	Y	Y*
Genista hirsuta	Algarve, Portugal	CEH2183	FHO, K	DQ524182	DQ852363¤		-	-	n	n
Genista pilosa	Cornwall, UK	CEH2340	FHO, K	DQ524183	DQ852361, DQ852362¤		QG529976	-	Y	Y*
Genista tenera	UBC Botanical Garden	Anon. BG11	UBC	AY338950†	-	-	AY338924†		Y#	n
Genista tinctoria	-	-	-	AF007471‡	-	-	-	-	n	n
Genista umbellata	Malaga, Spain	CEH1980	FHO, K	DQ524184	-	-	DQ529783	DQ529784	Y	n
L. 'Russell hybrid'	Sussex, UK	RJE s.n.		DQ524185	-	-	-	-	n	n
L. aff. piurensis	Peru	RJE60	FHO, K, MOL, USM	DQ524186	DQ852478	-	-	-	n	n
L. aff. ramossisimus	Cajamarca, Peru	CEH2010	CAS, FHO, K, MOL, NY	DQ524289	DQ852448	-	DQ529761	DQ529762	Y#	Y
L. aff. sarmentosus	Cajamarca, Peru	CEH2015	CAS, FHO, K, MO, MOL, NY	DQ524187	DQ852450	-	DQ529805	DQ529806	Y#	Y
L. affinis	Oregon, USA	-	-	AF007487‡	-	-	-	-	n	n
L. albescens	Brasil	CEH2509	FHO, ICN	DQ524190	DQ852503, DQ852504	-	DQ529974	DQ529975	Y#	Y
L. albescens	Rio Grande do Sul, Brasil	-	-	Z72212-3¶	-	-	-	-	n	n
L. albifrons	California, USA	-	-	AF007454-5‡	-	-	-	-	n	n
L. albifrons	-	-	-	Z72164-5¶	-	-	-	-	n	n
L. albus	Algrave, Portugal	CEH1970	FHO, K	DQ524191	DQ852367	DQ852536	DQ529787	DQ529788	Y#	Y
L. albus	UBC, Botanical Garden	USDA PI 170528§	UBC	AY338946†	-	-	AY338920†		Y	n
L. albus	Algeria	-	-	AF007481‡	-	-	-	-	n	n
L. albus subsp. albus	-	-	-	Z72200-1¶	-	-	-	-	n	n
L. albus var. graecus	Greece	-	-	AF007450-1‡	-	-	-	-	n	n
L. albus var. graecus	-	-	-	Z72198-9¶	-	-	-	-	n	n
L. anatolicus	-	-	-	AF108085¢	-	-	-	-	n	n
L. andersonii	Nevada, USA	M. Lavin 815985	MONT	AY338934†	-	-	AY338908†		Y#	n
L. angustifolius	Algarve, Portugal	CEH1972	FHO, K	DQ524192	DQ852368	DQ852541	DQ529978		Y#	Y
L. angustifolius	Sevilla, Spain	CEH1976	FHO, K	DQ524193	-	DQ852539, DQ852540	DQ529757	DQ529758	Y	n
L. angustifolius	Algeria	-	-	AF007448-9‡	-	-	-	-	n	n
L. angustifolius	Algeria	-	-	AF007477‡	-	-	-	-	n	n

L. angustifolius	-	-	-	Z72202-3¶	-	-	-	-	n	n
L. arboreus	Oxford, USA	CEH s.n.		DQ524194	-	DQ852548, DQ852549	DQ529753	DQ529754	Y	Y
L. arboreus	California, USA	G. Kenicer120	FHO	DQ524195	DQ852383, DQ852384	-	DQ529897	DQ529898	Y	Y
L. arboreus	California, USA	R. Spellenberg13488	FHO, NMC	DQ524196	DQ852380, DQ852381	-	DQ529972	DQ529973	Y#	Y
L. arboreus	UBC Botanical Garden	UBC Botanical Garden s.n.	UBC	AY338944†	-	-	AY338918†		Y	n
L. arboreus	-	-	-	Z72166-7¶	-	-	-	-	n	n
L. arcticus	Yukon, Canada	-	-	AF007495‡	-	-	-	-	n	n
L. arcticus	-	-	-	Z72156-7¶	-	-	-	-	n	n
L. argenteus	Montana, USA	A. Delgado & M. Lavin s.n.	MEXU	AY338929†	-	-	AY338903†		Y#	n
L. argenteus	-	M. F. Wojciechowski6905	ASU	-	DQ852379	-	-	-	n	n
L. argenteus	Nevada, USA	M. Lavin 815982	MONT	AY338931†	-	-	AY338905†		Y	n
L. argenteus	Montana, USA	M. Lavin s.n.	MONT	AY338936†	-	-	AY338910†		Y	n
L. argenteus	Washington, USA	-	-	AF007458-9‡	-	-	-	-	n	n
L. argenteus var. argenteus	New Mexico, USA	R. Spellenberg 13448	FHO, NMC	DQ524197	DQ852377, DQ852378	-	DQ529970	DQ529971	Y	Y
L. aridus	Oregon, USA	-	-	AF007446-7‡	-	-	-	-	n	n
L. arizonicus	Arizona, USA	USDA PI577285§		DQ524198	DQ852385, DQ852386	DQ852555	DQ529968	DQ529969	Y#	Y
L. arizonicus	USA	-	-	AF007483‡	-	-	-	-	n	n
L. arvensis	Bolivar, Ecuador	RJE188	FHO, QCNE	DQ524199	DQ852490	-	DQ529877	DQ529878	Y#	Y
L. aschenbornii	Mexico	N. Oropeza s.n.	-	unpublished	-	-	-	-	n	n
L. aschenbornii	Mexico	N. Oropeza s.n.	-	unpublished	-	-	-	-	n	n
L. aschenbornii	Mexico	-	-	Z72190-1¶	-	-	-	-	n	n
L. atlanticus	Morocco	-	-	AF007432-3‡	-	-	-	-	n	n
L. atlanticus	-	-	-	Z72216-7¶	-	-	-	-	n	n
L. aurenonitens	-	-	-	Z72210-1¶	-	-	-	-	n	n
L. ballianus	Ancash, Peru	CEH2038	CAS, FHO, K, MO, MOL NY	DQ524201	DQ852455, DQ852456	-	DQ529914	DQ529915	Y	Y
L. bandelierae	Cochabamba, Bolivia	CEH2301	FHO, K, LPB	DQ524202	DQ852500	-	DQ529966	DQ529967	Y	Y
L. bandelierae	Chuquisaca, Bolivia	CEH2310	FHO, K, LPB	DQ524203	DQ852501	-	DQ529820	DQ529821	Y	Y
L. bandelierae	Potosi, Bolvia	CEH2318	FHO, K, LPB	DQ524204	DQ852502	-	DQ529861	DQ529862	Y	Y
L. bandelierae	La Paz, Bolivia	CEH2321	FHO, K, LPB	DQ524205	-	-	DQ529859	DQ529860	Y#	n
L. bandelierae	La Paz, Bolivia	CEH2327	FHO, K, LPB	DQ524206	-	-	DQ529964	DQ529965	Y	n
L. bangii	La Paz, Bolivia	CEH2250	FHO (and RJE13 and RJE25), K, LPB	DQ524256	DQ852462	-	-	DQ529911	Y#	n
L. bangii	La Paz, Boliva	CEH2444	FHO, K, LPB	DQ524208	DQ852475, DQ852476	-	DQ529826	DQ529827	Y#	Y
L. benthamii	-	-	-	Z72168-9¶	-	-	-	-	n	n
L. bicolor	California, USA	USDA PI284713§	-	DQ524209	DQ852387, DQ852388	DQ852552	DQ529962	DQ529963	Y#	Y
L. bogotensis	-	-	-	Z72192-3¶	-	-	-	-	n	n

L. bracteolaris	Santa Catarina, Brasil	S. T. S. Miotto & M. Pinheiro 1810	ICN, FHO (RJE235)	DQ524210	DQ852505	-	DQ529816	DQ529817	Y#	Y
L. bracteolaris	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro 1819	ICN	DQ524211	DQ852506	-	DQ529857	DQ529858	Y	Y
L. bracteolaris	Brasil	-	-	AF007473‡	-	-	-	-	n	n
L. brevicaulis	New Mexico, USA	C. D. Bailey 243	FHO, NMC	DQ524213	-	-	-	-	n	n
L. breweri	California	M. Lavin 7031	MONT	AY338938†	-	-	AY338912†		Y	n
L. breweri	Nevada, USA	M. Lavin 8515983	MONT	AY338932†	-	-	AY338906†		Y#	n
L. chachas	Cajamarca, Peru	CEH2027	CAS, F, FHO, K, MO, MOL, NY	DQ524216	DQ852451, DQ852452	-	DQ529763	DQ529764	Y	Y
L. chamissonis	California, USA	G. Kenicer121	FHO	DQ524217	DQ852415, DQ852416	-	DQ529895	DQ529896	Y#	Y
L. chamissonis	California, USA	G. Kenicer122	FHO	DQ524218	-	-	DQ529893	DQ529894	Y	n
L. chamissonis	Oxford, UK	RJE s.n.		DQ524219	DQ852391, DQ852392	-	-	-	n	n
L. chrysanthus	La Paz, Bolivia	CEH2264	FHO, K, LPB	DQ524220	DQ852463, DQ852464	-	DQ529793	DQ529794	Y#	Y
L. chrysanthus	Apurimac, Peru	CEH2390	FHO, K, MOL	DQ524221	DQ852474	-	DQ529901	DQ529902	Y#	Y
L. concinnus	New Mexico, USA	C. D. Bailey242	FHO, NMC	DQ524212	DQ852389	-	DQ529960	DQ529961	Y	Y
L. concinnus	Arizona, USA	M. Baker10768	ASU	DQ524214	DQ852390	-	DQ529958	DQ529959	Y#	Y
L. concinnus	California, USA	-	-	AF007438-9‡	-	-	-	-	n	n
L. cosentinii	Algarve, Portugal	CEH1967	FHO, K	DQ524222	DQ852372	-	DQ529744	-	Y	Y*
L. cosentinii	-	USDA W617967	UBC	AY338947†	-	-	AY338921†		Y#	n
L. cosentinii	-	-	-	AF007464-5‡	-	-	-	-	n	n
L. cosentinii	-	-	-	Z72218-9¶	-	-	-	-	n	n
L. crotalarioides	Distrito Federal, Brasil	CEH2481	FHO, ICN, UB	DQ524223	DQ852507, DQ852508, DQ852509, DQ852510	-	DQ529863	DQ529864	Y#	Y
L. crotalarioides	Goias, Brasil	CEH2482	FHO, ICN, UB	DQ524224	DQ852511	-	DQ529855	DQ529856	Y	Y
L. crotalarioides	Goias, Brasil	CEH2483	FHO, ICN, UB	DQ524225	-	-	DQ529853	DQ529854	Y	n
L. crunkshanskii	-	-	-	Z72194-5¶	-	-	-	-	n	n
L. cumulicola	Florida, USA	CEH1983	FHO, K	DQ524226	DQ852373	-	DQ529745	DQ529746	Y#	Y
L. cumulicola	Florida, USA	CEH1988	FHO, K	DQ524227	DQ852374	DQ852537	DQ529747	DQ529748	Y	Y
L. densiflorus	UBC Botanical Garden	UBC Botanical Garden s.n.	UBC	AY338940†	-	-	AY338914†		Y	n
L. densiflorus	UBC Botanical Garden	UBC Botanical Garden s.n.	UBC	AY338941†	-	-	AY338915†		Y	n
L. densiflorus	UBC Botanical Garden	UBC Botanical Garden s.n.	UBC	AY338942†	-	-	AY338916†		Y	n
L. densiflorus	UBC Botanical Garden	UBC Botanical Garden s.n.	UBC	AY338943†	-	-	AY338917†		Y#	n
L. densiflorus	-	-	-	Z72186-7¶	-	-	-	-	n	n
L. digitatus	UBC Botanical Garden	USDA W611995§	UBC	AY338948†	-	-	AY338922†		Y#	Y*
L. digitatus	Egypt	-	-	AF007430-1‡	-	-	-	-	n	n
L. digitatus	-	-	-	Z72220-1¶	-	-	-	-	n	n

L. duranii	California, USA	-	-	AF007493‡	-	-	-	-	n	n
L. elegans	Mexico	-	-	AF007462-3‡	-	-	-	-	n	n
L. elegans		-	-	Z72170-1¶	-	-	-	-	n	n
L. ellsworthianus	Junin, Peru	RJE158	FHO, K, MOL, USM	DQ524215	DQ852483, DQ852484	-	DQ529885	DQ529886	Y	Y
L. excubitus	California, USA	-	-	AF007492‡	-	-	-	-	n	n
L. formosus	-	-	-	Z72172-3¶	-	-	-	-	n	n
L. gibertianus	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro1835	ICN	DQ524229	DQ852512	-	DQ529851	DQ529852	Y#	Y
L. grayi	California, USA	M. Lavin7033	MONT	AY338939†	-	-	AY338913†		Y#	n
L. guaraniticus	Rio Grande do Sul, Brasil	CEH2488	FHO, ICN	DQ524230	DQ852513, DQ852514	-	DQ529954	DQ529955	Y	Y
L. guaraniticus	Argentina	G. Kenicer s.n.	E, FHO	DQ524231	DQ852515	-	DQ529953	DQ529953	Y#	Y
L. havardii	Texas, USA	USDA DLEG910433§	-	DQ524232	DQ852394	-	DQ529950	DQ529951	Y	Y
L. havardii	Texas, USA	USDA DLEG930123§	-	DQ524233	DQ852395, DQ852396	-	DQ529948	DQ529949	Y#	Y
L. hirsutissimus	California, USA	S & S Seeds Ltd s.n.	-	DQ524234	DQ852397, DQ852398	DQ852554	DQ529946	DQ529947	Y	Y
L. hirsutissimus	California, USA	USDA DLEG920120§	FHO (RJE239, RJE240)	DQ524235	DQ852399, DQ852400	DQ852553	DQ529944	DQ529945	Y#	Y
L. hirsutissimus	California, USA	-	-	AF007486‡	-	-	-	-	n	n
L. hispanicus	Chaves, Portugal	J.R. Pannell1	FHO	DQ524236	DQ852369, DQ852370	DQ852542	DQ529780	DQ529781	Y#	Y
L. hispanicus	Portugal	-	-	AF007466-7‡	-	-	-	-	n	n
L. hispanicus	-	-	-	Z72204-5¶	-	-	-	-	n	n
L. huaronensis	Ancash, Peru	CEH2241	FHO, K, MOL	DQ524237	DQ852460	-	DQ529770	DQ529771	Y	Y
L. huigrensis	Loja, Ecuador	RJE217	FHO, QCNE	DQ524304	DQ852495	-	DQ529871	DQ529872	Y	Y
L. lactus	New Mexico, USA	R. Spellenberg12415	FHO, NMC	DQ524240	Genbank#	-	DQ529979		Y#	n
L. lanatus	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro1806	ICN	DQ524241	DQ852517	-	DQ529849	DQ529850	Y#	Y
L. lanatus	Santa Catarina, USA	S. T. S. Miotto & M. Pinheiro1837	ICN	DQ524242	DQ852516	-	DQ529847	DQ529848	Y	Y
L. latifolius	Arizona, USA	M. Baker10938	ASU	DQ524243	DQ852402	-	DQ529942	DQ529943	Y#	Y
L. latifolius	California, USA	-	-	AF007456-7‡	-	-	-	-	n	n
L. latifolius	-	-	-	Z72174-5¶	-	-	-	-	n	n
L. lepidus	New Mexico, USA	C.D. Bailey254	FHO, NMC	DQ524244	DQ852401	-	unpublished	unpublished	Y	Y
L. lepidus	New Mexico, USA	C.D. Bailey255	FHO, NMC	DQ524245	-	-	-	-	n	n
L. lepidus	Nevada, USA	M. Lavin815984	MONT	AY338933†	-	-	AY338907†		Y	n
L. lepidus	USA	USDA PI504439§	-	DQ524246	-	DQ852551	DQ529939	-	Y#	n
L. lepidus	Wyoming, USA	-	-	AF007485‡	-	-	-	-	n	n
L. leptophyllus	Mexico	N. Oropeza s.n.	-	unpublished	-	-	-	-	n	n
L. leptophyllus	Mexico	N. Oropeza s.n.	-	unpublished	-	-	-	-	n	n
L. leptophyllus	Mexico	N. Oropeza s.n.	-	unpublished	-	-	-	-	n	n
L. leucophyllus	Oregon, USA	-	-	AF007442-3‡	-	-	-	-	n	n
L. lindleyanus	Lima, Peru	RJE141	FHO, K, MOL, USM	DQ524247	DQ852479, DQ852480	-	DQ529889	DQ529890	Y#	Y
L. linearis	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro1798	ICN	DQ524248	DQ852519	-	DQ529828	DQ529829	Y#	Y

L. linearis	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro1813	ICN	DQ524239	DQ852518	-	DQ529846	-	Y	Y*
L. littoralis	Washington, USA	-	-	AF007452-3‡	-	-	-	-	n	n
L. luteolus	California, USA	-	-	AF007490‡	-	-	-	-	n	n
L. luteus	Sevilla, Spain	CEH1973	FHO, K	DQ524249	DQ852371	-	DQ529782	-	Y#	Y*
L. luteus	Algeria	-	-	AF007478‡	-	-	-	-	n	n
L. luteus	-	-	-	Z72206-7¶	-	-	-	-	n	n
L. magnistipulatus	Rio Grande do Sul, Brasil	M. Pinheiro164	ICN	DQ524251	DQ852521	-	DQ529937	DQ529938	Y	Y
L. magnistipulatus	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro1840	ICN	DQ524250	DQ852520	-	DQ529830	DQ529831	Y#	Y
L. magnistipulatus	Brasil	Zachia, N. R. B. 5842	FHO, ICN	-	-	-	unpublished	unpublished	n	n
L. mantaroensis	Cajamarca, Peru	CEH2000	CAS, FHO, K, MO, MOL, NY	DQ524200	DQ852444	-	DQ529759	DQ529760	Y#	Y
L. mexicanus	Mexico	N. Oropeza s.n.	-	unpublished	-	-	-	-	n	n
L. mexicanus	USA	-	-	AF007444-5‡	-	-	-	-	n	n
L. micranthus	Malaga, Spain	CEH1982	FHO, K	DQ524252	-	DQ852538	DQ529813	-	Y#	Y*
L. micranthus	Morocco	M. Carine278	BM	-	-	-	unpublished	-	n	n
L. micranthus	Algeria	-	-	AF007480‡	-	-	-	-	n	n
L. micranthus	-	-	-	Z72208-9¶	-	-	-	-	n	n
L. microcarpus	California	C. Drummond0331A02	not yet deposited, FHO (RJE243)	DQ524228	DQ852393	DQ852547	DQ529956	DQ529957	Y	Y*
L. microcarpus	Valparaiso, Chile	Darwin Chilean Initiative1513	FHO	DQ524253	DQ852404, DQ852405	-	DQ529935	DQ529936	Y#	Y
L. microcarpus	-	-	-	Z72188-9¶	-	-	-	-	n	n
L. microcarpus var. densiflorus	USA	-	-	AF007489‡	-	-	-	-	n	n
L. microcarpus var. microcarpus	Maryland, USA	-	-	AF007488‡	-	-	-	-	n	n
L. microphyllus	Ancash, Peru	CEH2238	FHO, K, MOL	DQ524254	DQ852458	-	DQ529789	DQ529790	Y	Y
L. microphyllus	Cochabamba, Bolivia	CEH2272	FHO, K, LPB	DQ524255	DQ852465	-	DQ529774	DQ529775	Y#	Y
L. minimus	Oregon, USA	-	-	AF007497‡	-	-	-	-	n	n
L. misticola	Lima, Peru	RJE144	FHO, K, MOL, USM	DQ5242657	DQ852481, DQ852482	-	DQ529887	DQ529888	Y#	Y
L. mollendoensis	Arequipa, Peru	M. Weigend97/703	USM	DQ524258	DQ852496, DQ852497	-	DQ529933	DQ299934	Y#	Y
L. montanus	Veracruz, Mexico	CEH2174	FHO, MEXU	DQ524259	DQ852427	DQ852559	DQ529809	DQ529810	Y#	Y
L. montanus	Mexico	N. Oropeza77s	-	unpublished	-	-	-	-	n	n
L. multiflorus	Brasil	CEH2489	FHO, ICN	-	-	-	unpublished	unpublished	n	n
L. multiflorus	Santa Catarina, Brasil	S. T. S. Miotto & M. Pinheiro1825	ICN	DQ524260	DQ852522	DQ852545, DQ852546	DQ529844	DQ529845	Y#	Y
L. multiflorus	Brasil	-	-	AF007475‡	-	-	-	-	n	n
L. mutabilis	Cajamarca, Peru	CEH1999	CAS, FHO, K, MOL, NY	DQ524261	DQ852428	DQ852568	-	-	n	n
L. mutabilis	Cajamarca, Peru	CEH2009	FHO (and RJE11, RJE18), K, MOL, NY	DQ524262	DQ852429	DQ852564, DQ852565, DQ852566, DQ852567	DQ529799	DQ529800	Y#	Y

L. mutabilis	Peru	CEH2039	FHO, K, MOL, NY	DQ524263	DQ852430	-	-	-	n	n
L. mutabilis	Peru	CEH2220	FHO, K, MOL	DQ524264	-	-	-	-	n	n
L. mutabilis	Peru	CEH2237	FHO, K, MOL	DQ524265	DQ852431	DQ852562, DQ852563	-	-	n	n
L. mutabilis	Peru	CEH2244	FHO, K, MOL	DQ524266	-	-	-	-	n	n
L. mutabilis	Boliva	CEH2253	FHO, K, LPB	DQ524267	DQ852432	-	-	-	n	n
L. mutabilis	Boliva	CEH2279	FHO, K, LPB	DQ524268	-	-	-	-	n	n
L. mutabilis	Boliva	CEH2280	FHO, K, LPB	DQ524269	-	-	-	-	n	n
L. mutabilis	Cochabamba, Bolivia	CEH2291	FHO, K, LPB	DQ524270	DQ852433	-	DQ529778	DQ529779	Y	Y
L. mutabilis	Boliva	CEH2309	FHO, K, LPB	DQ524271	-	-	-	-	n	n
L. mutabilis	Boliva	CEH2317	FHO, K, LPB	DQ524272	-	-	-	-	n	n
L. mutabilis	Boliva	CEH2323	FHO, K, LPB	DQ524273	-	-	-	-	n	n
L. mutabilis	Peru	-	-	AF007484‡	-	-	-	-	n	n
L. mutabilis	-	-	-	Z72196-7¶	-	-	-	-	n	n
L. nanus	-	Sutton Seeds s.n.	-	AY338928†	-	-	AY382155†		n	n
L. nanus	California, USA	-	-	AF007440-1‡	-	-	-	-	n	n
L. nanus	-	-	-	Z72176-7¶	-	-	-	-	n	n
L. neomexicanus	New Mexico, USA	R. Spellenberg13422	FHO, NMC	DQ524274	DQ852406, DQ852407	-	DQ529931	DQ529932	Y#	Y
L. nootkatensis	-	-	-	Z72160-1¶	-	-	-	-	n	n
L. nubigenus	Chimborazo, Ecuador	RJE200	FHO, QCNE	DQ529875	DQ852491, DQ852492	-	DQ529875	DQ529876	Y#	Y
L. odoratus	California, USA	C. Drummond0311A01	not yet deposited	DQ524276	DQ852403	-	DQ529930	-	Y#	Y*
L. palaestinus	Middle-East	-	-	AF007479‡	-	-	-	-	n	n
L. paraguariensis	Rio Grande do Sul, Brasil	CEH2495	FHO, ICN	DQ524277	DQ852524, DQ852525	-	DQ529928	DQ529929	Y	Y
L. paraguariensis	Rio Grande do Sul, Brasil	CEH2496	FHO, ICN	DQ524278	DQ852526, DQ852527, DQ852528	-	DQ529926	DQ529927	Y#	Y
L. paraguariensis	Brasil	CEH2497	FHO, ICN	-	-	-	unpublished	unpublished	n	n
L. paraguariensis	Brasil	-	-	AF007476‡	-	-	-	-	n	n
L. paraguariensis	-	-	-	Z72214-5¶	-	-	-	-	n	n
L. paranensis	Brasil	CEH2490	FHO, ICN	-	-	-	unpublished	unpublished	n	n
L. paranensis	Brasil	CEH2504	FHO, ICN	-	-	-	unpublished	unpublished	n	n
L. paranensis	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro ind 2_18	ICN	DQ524279	DQ852523	-	DQ529924	DQ529925	Y#	Y
L. parvifolius	Minas Gerais, Brasil	CEH2480	FHO, ICN, UB	DQ524280	DQ852529	-	DQ529836	DQ529837	Y#	Y
L. perennis	-	-	-	Z72162-3¶	-	-	-	-	n	n
L. pilosus	-	-	-	AF007434‡	-	-	-	-	n	n
L. pilosus	-	-	-	Z72222-3¶	-	-	-	-	n	n
L. piurensis	Cajamarca, Peru	CEH1997	FHO (and RJE16), K, MOL, NY	DQ524281	DQ852434, DQ852435	-	DQ529814	DQ529815	Y#	Y
L. piurensis	Cajamarca, Peru	RJE110	FHO, K, MOL, USM	DQ524282	DQ852436	-	DQ529891	DQ529892	Y	Y
L. piurensis	Piura, Peru	RJE122	FHO, K, MOL	DQ524283	DQ852437, DQ852438	-	-	-	n	n
L. piurensis	Piura, Peru	RJE124	FHO, K, MOL, USM	DQ524284	DQ852439	-	-	-	n	n

L. polycarpus	-	-	-	Z72180-1¶	-	-	-	-	n	n
L. polyphyllus	Nevada, USA	M. Lavin815981	MONT	AY338930†	-	-	AY338904†		Y#	n
L. polyphyllus	UBC Botanical Garden	UBC Botanical Garden s.n.	UBC	AY338945†	-	-	AY338919†		Y	n
L. polyphyllus	Washington, USA	-	-	AF007496‡	-	-	-	-	n	n
L. polyphyllus	-	-	-	Z72154-5¶	-	-	-	-	n	n
L. praestabilis	Arequipa, Peru	CEH2353	FHO, K, MOL	DQ524285	DQ852471	DQ852561	DQ529822	DQ529823	Y#	Y
L. princei	-	-	-	Z72224-5¶	-	-	-	-	n	n
L. prostratus	Ancash, Peru	CEH2239	FHO, K, MOL	DQ524286	DQ852459	-	DQ529791	DQ529792	Y#	Y
L. pubescens	Cajamarca, Peru	CEH2037	CAS, FHO, K, MO, MOL, NY	DQ524187	DQ852454	-	DQ529766	DQ529767	Y#	Y
L. pubescens	Pichincha, Ecuador	RJE176	FHO, QCNE	DQ524287	DQ852488, DQ852489	-	DQ529879	DQ529880	Y	Y
L. pubescens	-	-	-	Z72178-9¶	-	-	-	-	n	n
L. pulvinaris	Cochabamba, Boliva	CEH2273	FHO, K, LPB	DQ524288	DQ852466	-	DQ529776	DQ529777	Y#	Y
L. purosericeus	Pasco, Peru	CEH2248	FHO, K, MOL	DQ524238	DQ852461	-	DQ529772	DQ529773	Y#	Y
L. pusillus	Oregon, USA	-	-	AF007491‡	-	-	-	-	n	n
L. reitzii	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro1846	ICN	DQ524290	DQ852530	-	DQ529832	DQ529833	Y#	Y
L. rivularis	California, USA	M. Lavin7040	MONT	AY338937†	-	-	AY338911†		Y	n
L. rivularis	-	-	-	Z72182-3¶	-	-	-	-	n	n
L. rubriflorus	Brasil	CEH2485	FHO, ICN	-	-	-	unpublished	unpublished	n	n
L. rubriflorus	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro1838	ICN	DQ524292	DQ852531	-	DQ529842	DQ529843	Y#	Y
L. sarmentosus	Pichincha, Ecuador	R. J. Eastwood174	FHO	DQ524291	DQ852487	-	DQ529881	DQ529882	Y	Y
L. semperflorens	Cajamarca, Peru	CEH2012	CAS, F, FHO, K, MO, MOL, NY	DQ524293	DQ852441	-	DQ529801	DQ529802	Y	Y
L. semperflorens	Cajamarca, Peru	CEH2030	CAS, FHO, K, MOL, NY	DQ524188	DQ852453	-	DQ529765		Y	n
L. semperflorens	Cajamarca, Peru	RJE109	FHO, K, MOL, USM	DQ524295	-	-	unpublished		n	n
L. semperflorens	Loja, Ecuador	RJE219	FHO, K, QCNE	DQ524296	DQ852440	-	DQ529922	DQ529923	Y#	Y
L. semperflorens	Loja, Ecuador	RJE220	FHO, K, QCNE	DQ524297	DQ852442, DQ852443	DQ852569, DQ852570, DQ852571	DQ529869	DQ529870	Y	Y
L. sericeus	Montana, USA	M. Lavin s.n.	MONT	AY338935†	-	-	-	-	Y#	n
L. sericeus	Utah, USA	-	-	AF007436-7‡	-	-	-	-	n	n
L. sierrae-blancae	New Mexico, USA	R. Spellenberg13409	FHO, NMC	DQ524298	DQ852408, DQ852409	DQ852550	DQ529920	DQ529921	Y#	Y
L. solangorum	Cajamarca, Peru	CEH2014	CAS, FHO, K, MOL, NY	DQ524294	DQ852448	-	DQ529803	DQ529804	Y	Y
L. sp. NOH0-s	Mexico	-	-	unpublished	-	-	-	-	n	n
L. sp. NOH105	Mexico	-	-	unpublished	DQ852420	-	-	-	n	n
L. sp. NOH135C	Mexico	-	-	unpublished	-	-	-	-	n	n
L. sp. NOH155	Mexico	-	-	unpublished	-	-	-	-	n	n
L. sp. NOH1-s	Mexico	-	-	unpublished	-	-	-	-	n	n
L. sp. NOH61	Mexico	-	-	unpublished	DQ852417	DQ852556, DQ852557	-	-	n	n
L. sp. NOH67	Mexico	-	-	unpublished	-	-	-	-	n	n
L. sp. NOH80	Mexico	-	-	unpublished	DQ852419	DQ852558	-	-	n	n
L. sp. NOH82	Mexico	-	-	unpublished	DQ852418	-	-	-	n	n

L. sp1	Cajamarca, Peru	CEH2002	CAS, F, FHO, K. MEXU, MO, MOL, NY	DQ524299	DQ852445	-	DQ529795	DQ529796	Y#	Y
L. sp1	Cajamarca, Peru	CEH2005	CAS, FHO, K, MOL, NY	DQ524189	DQ852447	-	DQ529785	DQ529786	Y	Y
L. sp2	Cajamarca, Peru	CEH2004	CAS, F, FHO, K, MO, MOL, NY	DQ524305	DQ852446	DQ852560	DQ529797	DQ529798	Y#	Y
L. sp3	Cajamarca, Peru	CEH2016	FHO, MEXU	DQ524307	-	-	-	-	n	n
L. sp4	Oaxaca, Mexico	CEH2107	FHO, MEXU	DQ524308	DQ852421	-	DQ529807	DQ529808	Y#	Y
L. sp4	Oaxaca, Mexico	CEH2118	FHO, MEXU	DQ524309	DQ852424	-	DQ529768	DQ529769	Y#	Y
L. sp5	Oaxaca, Mexico	CEH2117	FHO, MEXU	DQ524310	DQ852422, DQ852423	-	DQ529912	DQ529913	Y#	Y
L. sp6	Mexico, Mexico	CEH2160	FHO, MEXU	DQ524311	DQ852425, DQ852426	-	DQ529909	DQ529910	Y#	Y
L. sp7	Imbabura, Ecuador	RJE59	FHO, K, MOL	DQ524301	DQ852477	-	DQ529867	DQ529868	Y#	Y
L. sp8	Imbabura, Ecuador	RJE168	FHO, QCNE	DQ524302	DQ852485, DQ852486	-	DQ529883	DQ529884	Y#	Y
L. sp9	Azuay, Ecuador	RJE206	FHO, QCNE	DQ524303	DQ852493, DQ852494	-	DQ529873	DQ529874	Y	Y
L. sparsiflorus	Arizona, USA			AF007482‡	-	-	-	-	n	n
L. subacaulis	Potosi, Boliva	CEH2319	FHO, K, LPB	DQ524312	DQ852468, DQ852469	-	DQ529907	DQ529908	Y	Y
L. subsessilis	Goias, Brasil	CEH2479	FHO, ICN, UB	DQ524313	DQ852532	DQ852543, DQ852544	DQ529840	DQ529841	Y#	Y
L. succulentus	California, USA	-	-	AF007494‡	-	-	-	-	n	n
L. succulentus	-	-	-	Z72184-5¶	-	-	-	-	n	n
L. sulphureus	Washington, USA	-	-	AF007460-1‡	-	-	-	-	n	n
L. tarapacensis	Tarapaca, Chile	M. Gardner6296	FHO	DQ524314	DQ852498, DQ852499	-	DQ529899	DQ529900	Y#	Y
L. tegeticulatus	-	-	-	AF007493	-	-	-	-	n	n
L. texensis	-	Fothergills Seeds s.n.	-	DQ524315	-	-	-	-	n	n
L. texensis	-	M. F. Wojciechowski244	ASU	-	-	-	-	-	n	n
L. texensis	Texas, USA	S&S Seeds Ltd s.n.	-	DQ524316	DQ852411, DQ852412, DQ852413	-	DQ529918	DQ529919	Y	Y
L. texensis	Texas, USA	USDA PI577291§	UBC	AY338949†	DQ852410	-	AY338923†	-	Y#	Y
L. texensis	Texas, USA	-	-	AF007474‡	-	-	-	-	n	n
L. tomentosus	La Paz, Bolivia	CEH2322	FHO, K, LPB	DQ524317	DQ852470	-	DQ529905	DQ529906	Y#	Y
L. tominensis	Cochabamba, Bolivia	CEH2294	FHO (and RJE24), K, LPB	DQ524207	DQ852467	-	DQ529824	DQ529825	Y	Y
L. tominensis	La Paz, Bolivia	CEH2370	FHO, K, MOL	DQ524300	DQ852472, DQ852473	-	DQ529903	DQ529904	Y	Y
L. truncatus	Texas, USA	S&S Seeds Ltd s.n.	FHO (RJE237, RJE238, RJE242)	DQ524318	DQ852414	-	DQ529916	DQ529917	Y#	Y
L. uleanus		CEH2506	FHO, ICN		-	-	unpublished	unpublished	n	n
L. uleanus	Rio Grande do Sul, Brasil	S. T. S. Miotto & M. Pinheiro1848	ICN	DQ524319	DQ852533	-	DQ529838	DQ529839	Y#	Y
L. velutinus	Distrito Federal, Brasil	CEH2477	FHO, ICN, UB	DQ524320	DQ852534	-	DQ529834	DQ529835	Y	Y
L. velutinus	Distrito Federal, Brasil	CEH2478	FHO, ICN, UB	DQ524321	DQ852535	-	DQ529818	DQ529819	Y#	Y
L. villosus	Florida, USA	CEH1986	FHO, K	DQ524322	DQ852375	-	DQ529749	DQ529750	Y#	Y
L. villosus	Florida, USA	CEH1987	FHO, K	DQ524323	DQ852376	-	DQ529751	DQ529752	Y	Y

L. weberbaueri	Ancash, Peru	CEH2233	FHO, K, MOL	DQ524324	DQ852457	-	DQ529811	DQ529812	Y#	Y
Melolobium obcordatum	-	-	-	Z95561-2	-	-	-	-	n	n
Petteria ramenta	-	-	-	Z72232-3	-	-	-	-	n	n
Pterospartum tridentata	Alentejo, Portugal	CEH2338	FHO, K	DQ524325	-	-	DQ529866	-	Y	n
Retama monosperma	Algarve, Portugal	CEH2182	FHO, K	DQ524326	DQ852364, DQ852365¤		DQ529755	-	Y	Y*
Spartium junceum	Algarve, Portugal	CEH2190	FHO, K	DQ524327	DQ524327¤		DQ529756	-	Y	n
Stauracanthus genistoides	Algarve, Portugal	CEH2335	FHO, K	DQ524328	DQ852366¤		DQ529865	-	Y	n
Thermopsis rhombifolia	Idaho, USA	-	-	AF007468‡	-	-	-	-	n	n
Ulex parviflorus	France	-	-	AF007470‡	-	-	-	-	n	n

¤ Copy type has not been ascertained for all outgroups.

Sources

¢Ainouche, A. 1998. Diversity and evolution of the genus *Lupinus* L. (Fabaceae), University of Rennes.

‡: Aïnouche, A. and Bayer, R. (1999) Phylogenetic relationships in *Lupinus* (Fabaceae: Papiliondeae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. American Journal of Botany 86: 590-607

¥ Kaess, E. (1995) Thesis, Ubiversitaet Heidelberg, Heidelberg

¶: Käss, E. and Wink, M.(1997) Molecular phylogeny and phylogeography of *Lupinus* (Leguminosae) inferred from nucleotide sequences of the rbcL gene and ITS 1 + 2 regions of rDNA. Plant Systematics and Evolution 208: 139-167

Appendix 6 Protocols and Solutions

Protocol 1 - Leaf grinding Method 1

1) Label set of 1.5 ml tubes with sample accession numbers.
2) Harvest 1-3 leaflets per sample.
3) Chill empty tube by repeatedly adding liquid nitrogen. Add leaf material and more liquid nitrogen.
4) Chill narrow pestle in liquid nitrogen for a few seconds then use to grind leaf material to a fine powder.
5) Store in freezer until needed.

Protocol 2 - Leaf grinding Method 2

1) Label cluster tubes (ABGene tubes - AB0672, caps – AB0981) with sample accession numbers.
2) Place 1-3 leaflets per sample into tubes.
3) Place one tungsten carbide bead (Quiagen No. 69997) into each tube.
4) Divide tubes equally between two racks and place into grinding machine making sure the distribution of the tubes in each rack is balanced.
5) Grind for 30 seconds at 20 beats per second. Remove racks and replace in reverse orientation. Repeat grinding step.
6) Check samples to ensure the material has been ground to a fine powder. Repeat step 5 if clumps of leaflet are visible.
7) Store in freezer until needed. The bead can be easily removed during subsequent protocols.

Protocol 3 – Cleaning beads

1) Rinse in water.
2) Swirl for 1 minute in 0.4M hydrochloric acid.
3) Rinse in deionised water using sieve.
4) Air dry.

Protocol 4 - DNA Extraction (Modified CTAB method (Doyle & Doyle, 1987))

1) Preheat a mix of 2x CTAB buffer (1 ml per sample) and mercapto-ethanol (10 µl per sample) to 65 °C in a water bath.
2) Grind leaf material following protocol 1 or 2.
3) Pipette 1 ml of CTAB/mercapto-ethanol into each tube of powdered leaf material and continue to grind until the tissue forms a homogeneous mix. Incubate for 30 minutes in a 65 °C water bath.
4) Remove tubes from water bath and allow to cool to room temperature.
5) In a fume cupboard add 200 µl 'wet' chloroform to each tube. Mix gently by inverting tube.
6) Centrifuge for 10 minutes at 13,000 rpm.
7) Remove the aqueous upper phase to a new and labelled tube. Repeat steps 5 to 7.
8) Add 600 µl (or equivalent to 2/3 volume of the aqueous layer) of cold (-20 °C) propan-2-ol to each tube and mix gently. Leave to stand for 10-60 minutes until plenty of white precipitate is visible.
9) Centrifuge for 15 minutes at 13,000 rpm.
10) Remove the supernatant with a pipette and add 1 ml of wash buffer to each tube. Vortex each tube and leave for at least 1 hour (this can be left all night).
11) Centrifuge for 2 minutes at 13,000 rpm.
12) Remove as much supernantant as possible without losing or disturbing the DNA pellet. Place tubes into an incubator at 37 °C to dry.
13) Dissolve dry pellet in 200 µl of TE buffer.
14) The DNA suspension can be stored for up to 6 months at 4 °C or longer at -20 °C.

Protocol 5 – RNA removal

1) Add 1 µl of RNAse per 100 µl DNA solution and incubate at 37 °C for 30 minutes.
2) Add 100 µl of cold 7.5M ammonium acetate and then 750 µl cold 100% ethanol. Incubate on ice for 15 minutes.
3) Centrifuge for 10 minutes at 13,000 rpm.
4) Pour of liquid and wash with 200 µl of 70% ethanol.
5) Centrifuge for 2 minutes at 13,000.
6) Pour off liquid and air dry.
7) Resuspend DNA in 50-200 µl or TE buffer or water.
8) The DNA suspension can be stored for up to 6 months at 4 °C or longer at -20 °C.

Protocol 6 DNA Extraction (Quiagen DNeasy Plant Mini Kit No, 69104)

1) Add 400 µl of Buffer AP1 and 4 µl RNase A stock solution to ground leaf material and vortex vigorously.
2) Incubate the mixture for 10 minutes at 65 °C. Mix 2-3 times during incubation by inverting tube.
3) Add 130 µl of Buffer AP2 to the lysate, mix, and incubate for 5 minutes on ice.
4) Centrifuge the lysate for 5 minutes at 13,000 rpm.
5) Apply the lysate to the QIAshredder spin column sitting in a 2ml collection tube and centrifuge for 2 minutes at 13,000 rpm. If beads were used during the leaf grinding process collect at this stage and clean following protocol 3.
6) Transfer flow-through fraction from step 5 to a new tube without disturbing the cell-debris pellet.
7) Add 1.5 volumes of Buffer AP3/E to the cleared lysate and mix by pipetting.
8) Apply 650 µl of the mixture from step 7, including any precipitate, to the DNeasy mini spin column sitting in a 2 ml collection tube. Centrifuge for 1 minute at 8000 rpm and discard flow-through.
9) Repeat step 8 with remaining sample reusing the same collection tube.
10) Place DNAeasy column in a new 2 ml collection tube, add 500 µl Buffer AW to the DNAeasy column and centrifuge for 1 minute at 8000 rpm. Discard flow-through and reuse the collection tube in step 11.
11) Add 500 ul Buffer AW to the DNAeasy column and centrifuge for 2 minute at maximum speed to dry the membrane.
12) Transfer the DNAeasy column to a 1.5 ml tube and pipet 50 µl of preheated (65 °C) Buffer AE directly onto the DNAeasy membrane. Incubate for 5 minutes at room temperature and then centrifuge for 1 minute at 8000 rpm to elute.
13) Repeat step 12 once.
14) The DNA suspension can be stored for up to 6 months at 4 °C or longer at -20 °C.

Protocol 7 – Agar gel (1%)

1) Add 0.3 g of agar to 30 ml of TE buffer.
2) Microwave for 1 minute.
3) Add 5 µl ethidium bromide.
4) Cool to room temperature.
5) Pour into gel rig with baffles present and leave to harden.

Protocol 8 – DNA concentration and quality determination

1) Aliquot 1 µl of stop buffer onto gel-loading plate.
2) Add 2.5 µl of DNA solution, mix and load onto 1% gel. In a separate lane load 1 µl of 1 kb DNA ladder (New England Biolabs #N3232S).
3) Run gel at 100mV and 70mA for 20 minutes in an electrophoresis machine.
4) View gel under UV light. Assess quality of DNA and amount present by comparison with the ladder.

Protocol 9 - Amplification of ITS

Recipe

Water	34 µl
10x PCR buffer	5 µl
Q solution/betaine	5 µl
dNTP mix (i.e. 10mM of each)	2 µl
Each primer in 20 µM solution	1.25 µl
taq	0.5 µl
DNA	1 µl
Total volume	50 µl

PCR programme

	3 minutes at 94 °C	pre-treatment
34 x	(1 minute at 94 °C	denaturation
	1 minute at 52 °C	annealing
	1 minute at 72 °C)	extension
	10 minutes at 72 °C	post-treatment

1) Make up PCR mixture containing appropriate volumes of water, 10x PCR buffer, Q/betaine, dNTP mix and primers sufficient for all samples. Primer pairs used in this study were ITS4 & ITS5 and ITS4 & 17SE. Primer sequences are listed in Appendix 5.
2) Label tubes and incubate on ice.
3) Aliquot 48.5 µl of PCR mix into each tube including a negative control which will not receive any DNA.
4) Add 1 µl taq to each tube.
5) Add 1 µl DNA to each tube except the negative control.
6) Flick to mix and centrifuge for 20 seconds.
7) Place tubes in thermocycler and run programme listed above.
8) Run out 2.5 µl on an agar gel to visualise products.

Protocol 10 - Amplification of GPAT

Recipe

Water	9.3 µl
10x PCR buffer	2 µl
Q solution/betaine	2 µl
dNTP mix (i.e. 10mM of each)	1 µl
Each primer in 20 µM solution	1 µl
taq	0.2 µl
DNA	3.5 µl *
Total volume	20 µl

PCR programme

	4 minutes at 94 °C	pre-treatment
25-35 x	(45 seconds at 94 °C	denaturation
	45 seconds at 50-60 °C **	annealing
	1-3 minutes at 72 °C)	extension
	10 minutes at 72 °C	post-treatment

Recipe with PCR beads

Water	20 µl
Each primer in 20 µM solution	1 µl
DNA	2.5 µl *
Total volume	25 µl

PCR programme with PCR beads

	5 minutes at 95 °C	pre-treatment
25-35 x	(45 seconds at 95 °C	denaturation
	45 seconds at 50-60 °C **	annealing
	1-3 minutes at 72 °C)	extension
	10 minutes at 72 °C	post-treatment

* Although this quantity of DNA seems excessive poorer results were obtained with less.
** Extension times were varied according to product length following the general rule of 1 minute per 1 kilobase.

For normal PCR.
1) Make up PCR mixture containing appropriate volumes of water, 10x PCR buffer, Q/betaine, dNTP mix and primers sufficient for all samples. Primer pairs and appropriate annealing temperatures and extension times are listed in Appendix 5. Primer sequences are listed in Appendix 5.
2) Label tubes and incubate on ice.
3) Aliquot 48.5 µl of PCR mix into each tube including a negative control which will not receive any DNA.
4) Add 1 µl taq to each tube.
5) Add DNA to each tube except the negative control.
6) Flick to mix and centrifuge for 20 seconds.
7) Place tubes in thermocycler and run appropriate programme.
8) Run out 2.5 µl on an agar gel to visualise products.

For PCR with beads (Amersham Biosciences pure Taq Ready-To-Go PCR Beads No. 27-9558-01).
1) Make up PCR mixture containing appropriate volumes of water and primers sufficient for all samples. Primer pairs and appropriate annealing temperatures and extension times are listed in Appendix 5. Primer sequences are listed in Appendix 5.
2) Label tubes and incubate on ice.
3) Aliquot 16.3 µl of water and primer mix into each tube including a negative control which will not receive any DNA.
4) Follow from step 5 as per normal PCR.

Protocol 11 – Amplication of plasmid and insert

Recipe

Water	7.12 µl
10x PCR buffer	1 µl
dNTP mix (i.e. 10mM of each)	0.5 µl
Each primer in 20 µM solution	0.25 µl
taq	0.13 µl
DNA	1 µl
Total volume	10 µl

PCR programme

	3 minutes at 94 °C	pre-treatment
35 x	(30 seconds at 94 °C	denaturation
	30 seconds at 54 °C	annealing
	1 minute at 72 °C)	extension
	10 minutes at 72 °C	post-treatment

1) Make up PCR mixture containing appropriate volumes of water, 10x PCR buffer, dNTP mix and primers sufficient for all samples. M13F and M13R primers are used. Primer sequences are listed in Appendix 5.
2) Label tubes and incubate on ice.
3) Aliquot 48.5 µl of PCR mix into each tube including a negative control which will not receive any DNA.
4) Add 1 µl taq to each tube.
5) Add DNA to each tube except the negative control.
6) Flick to mix and centrifuge for 20 seconds.
7) Place tubes in thermocycler and run appropriate programme.
8) Run out 2.5 µl on an agar gel to visualise products.

Protocol 12 – Amplification of ISSRs

Recipe

Water	11.2 µl
10x PCR buffer	2 µl
dNTP mix (i.e. 10mM of each)	1.6 µl
Each primer in 20 µM solution	1 µl
taq	0.2 µl
DNA	3 µl
Total volume	20 µl

PCR programme

	4 minutes at 94 °C	pre-treatment
40 x	(30 seconds at 94 °C	denaturation
	45 seconds at 45 °C	annealing
	1 minute at 72 °C)	extension
	10 minutes at 72 °C	post-treatment

1) Make up PCR mixture containing appropriate volumes of water, 10x PCR buffer, dNTP mix and primers sufficient for all samples. One 3' and one 5' ISSR primer used. Primer sequences are listed in Appendix 8.
2) Label tubes and incubate on ice.
3) Aliquot 18 µl of PCR mix into each tube including a negative control which will not receive any DNA.

4) Add 2 µl taq to each tube.
5) Add DNA to each tube except the negative control.
6) Flick to mix and centrifuge for 20 seconds.
7) Place tubes in thermocycler and run appropriate programme.
8) Run out 4 µl on an agar gel to visualise products.

Protocol 13 – Amplification of specific loci identified using ISSR primers

Recipe

Water	15.8µl
10x PCR buffer	2 µl
dNTP mix (i.e. 10mM of each)	1.0 µl
Each primer in 20 µM solution	0.4 µl
taq	0.4 µl
DNA	0.4 µl
Total volume	20 µl

PCR programme

	4 minutes at 94 °C	pre-treatment
30 x	(30 seconds at 94 °C	denaturation
	30 seconds at 50 °C	annealing
	1 minute at 72 °C)	extension
	10 minutes at 72 °C	post-treatment

1) Make up PCR mixture containing appropriate volumes of water, 10x PCR buffer, dNTP mix and primers sufficient for all samples. For locus 7:22-2 primers 7:22-2F1 and 7:22-2R1 used. For locus E1 primers 3:6-F1 and 3:6-R2 were used with an annealing temperature of 55 °C. Primer sequences are listed in Appendix 8.
2) Label tubes and incubate on ice.
3) Aliquot 19.6 µl of PCR mix into each tube including a negative control which will not receive any DNA.
4) Add 0.4 µl taq to each tube.
5) Add DNA to each tube except the negative control.
6) Flick to mix and centrifuge for 20 seconds.
7) Place tubes in thermocycler and run appropriate programme.
8) Run out 2.5 µl on an agar gel to visualise products.

Protocol 14 – QIAquick PCR purification (Quiagen No. 28104)
1) Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix.
2) Place a QIAquick spin column in a provided 2 ml collection tube
3) To bind DNA, apply the sample to the QIAquick column and centrifuge for 30 seconds at 13,000 rpm.
4) Discard flow-through. Place the QIAquick column back into the same tube.
5) Wash column with 0.75 ml Buffer PE and centrifuge for 30 seconds at 13,000 rpm.
6) Discard flow-through and place the QIAquick column back into the same tube. Centrifuge for an additional minute at 13,000 rpm.
7) Place the QIAquick column in a clean 1.5 ml microcentrifuge tube.
8) To elute DNA, add 30 µl Buffer EB or water to the centre of the membrane. Let the membrane stand for one minute and centrifuge for one minute.

Protocol 15 – Bioline Quick-Clean (Bioline BIO-37046)
1) Add 1x volume of Quick-Clean to PCR sample and vortex thoroughly.
2) Incubate at room temperature for 15 minutes.
3) Centrifuge for 15 minutes at 13,000 rpm and discard supernatant.
4) Add 100 µl of 70% ethanol and vortex thoroughly.
5) Centrifuge for 10 minutes and discard supernatant.
6) Repeat steps 5 and 6.
7) Air dry. Resuspend pellet in 30 µl water.

Protocol 16 – Quiagen QIAquick gel extraction (Quiagen No. 28704)

1) Prepare a large 2% gel with wide wells sufficient to hold the entire sample.
2) Mix entire PCR sample with 3 µl of loading buffer and load into well.
3) Run gel at 200mV, 175mA for an hour in an electrophoresis machine.
4) Prepare labeled tubes.
5) Cut out bands as closely as possible with a sterile scalpel using a UV transilluminator for visualization. Sterilize scalpel in hypochlorite and water between each band.
6) Weigh gel slice and add 3 volumes of Buffer QG to 1 volume of gel.
7) Incubate at 50 °C for 10 minutes, vortexing every 2-3 minutes.
8) Add 1 gel volume of isopropanol to the sample and mix.
9) Apply the sample to a QIAquick spin column in a 2 ml collection tube. Centrifuge for 1 minute at 13,000.
10) Discard flow-through and place QIAquick column back into same tube.
11) Add 0.5 ml of Buffer QG to QIAquick column and centrifuge for 1 minute at 13,000.
12) Add 0.75 ml of Buffer PE to QIAquick column and centrifuge for 1 minute at 13,000.
13) Discard flow-through and centrifuge QIAquick column for an additional minute at 13,000.
14) Place QIAquick column into a clean 1.5 ml tube.
15) Add 30 µl of Buffer EB or water onto the centre of the membrane and let the column stand for 1 minute. Centrifuge at 13,000 for 1 minute.

Protocol 17 - Sequencing

Recipe			PCR Programme
Water	1 to 4 µl	25 x	(30 seconds at 96 ° C
Primer	0.33 µl		15 seconds at 50 °C
BigDye 3.1	4 µl		4 minutes at 60 °C)
DNA	*Volume was dependent on DNA concentration*		
Total volume	10 µl		

1) Judge volume of DNA to include using protocol 8 and aiming to include at least 30ng of DNA.
2) Label tubes and incubate on ice.
3) Mix primer and water (one volume per sample) and aliquot into tubes.
4) Add BigDye and then DNA.
5) Flick tubes gently to mix and centrifuge for 20 seconds at 13,000 rpm.
6) Place tubes in thermocycler and run sequencing programme given above.

Protocol 18 - Ethanol precipitation

1) Remove tubes from centrifuge and centrifuge for 20 seconds at 13,000 rpm.
2) Transfer reaction contents to an appropriately labelled 0.5 ml tube which already contains 100% ethanol and 1 µl of 3M sodium acetate.
3) Vortex tubes briefly immediately after adding DNA and leave for 10 minutes.
4) Centrifuge at for 20 minutes at 13,000 rpm.
5) Pipette away as much of the liquid as possible.
6) Add 250 µl 70% ethanol.
7) Mix gently and then centrifuge for 5 minutes at 13,000.
8) Remove as much liquid as possible without disturbing the pellet. Place in oven to dry.

Protocol 19 – Cloning modified from Promega (pGEM®-T Vector System II No. A3610)

(Once the E. coli have been transformed all products coming into direct contact with the transformants need to be disposed of according to category 1 GM guidelines. Contaminated pipette tips/glassware/and other items (accept plates and liquid cultures) should be sterilised. This can be done with a GM approved autoclave or by soaking items in a 10% hypochlorite solution for at least 2 hours. Contaminated work surfaces should be sterilised with 10% hypochlorite solution or 70% alcohol (not 100%). All contaminated culture plates should be autoclaved in a GM approved autoclave and disposed off in a yellow biohazard bag. Liquid cultures must be sterilised (bring up to 10% hypochlorite) or autoclave in a GM approved autoclave. Less heavily contaminated dispoable items (gloves/tips/etc) can be placed directly into a yellow biohazard bag for incineration or autoclaving.)

Ligation stage
1) Centrifuge vector (pGEM-T).
2) In a 0.5 µl epindorph tube mix:
- 2.5 µl 2x Rapid ligation buffer,
- 0.5 µl vector
- 1.5 µl PCR product
- 0.5 µl DNA ligase

3) Mix by pipetting using an aerosol free tip or swirling.
4) Incubate overnight at 4°C for maximum transformants. To save time it is possible to incubate for 1 hour at room temperature or use a combination with some hours in the fridge and 45 minutes on the bench but the number of transformations is less this way.

Transformation stage
5) Centrifuge ligation reaction.
6) Label 1.5 ml tubes and keep on ice.
7) Add 2 µl of each ligation reaction to appropriate tube.
8) Remove necessary tube of JM109 competent cells from -80 °C. Cells cannot be refrozen so work out the appropriate volume to use for the number of reactions being carried out. Thaw the cells on ice, mix gently by tapping/flicking the tube. These are sensitive so do not centrifuge.
9) Aliquot 25 µl (for 8 reactions) or appropriate volume (i.e. 20 µl for 10 reactions) of cells into each tube using a filter tip.
10) Mix cells and plasmids gently by swirling the pipette. Do not pipette up and down. Incubate on ice for 20 minute.
11) Thaw SOC buffer (stored at -20 °C).
12) Place LB/X-gal/Amp plates in the 37 °C incubator to equilibrate.
13) Heat
shock for 45-50 sec in a 42 °C water bath.
14) Immediately return to ice for 2 minutes.
15) Add 150 µl of SOC buffer.
16) Incubate for 45 minutes in a shaking incubator at 37°C. This stage allows the bacteria to repair their cell walls.
17) Plate out 100 µl of transformant onto LB/X-gal/Amp plate and incubate overnight at 37 °C.
18) Prepare 10 (or more) 0.5 ml epindorph tubes for each transformation with roughly 200 µl (exact amount unimportant) of LB + amp solution (REM to add ampilcillin to LB. Once amp is added this will keep in fridge for a couple of weeks. It is worth looking at this medium carefully for bacterial colonies before use as it is an easy source of contamination.)
19) Prepare a negative containing just the LB + amp solution.
20) Sterilise a toothpick/cocktail stick with 70% ethanol and flame between picking each colony and use a new toothpick/cocktail stick for each transformant.
21) To pick a colony just touch the colony with the sterile toothpick/cocktail stick and then just dip the tip into the LB + amp solution in the epindorph. Pick each colony into a separate epindorph. Each epindorph holds a single clone isolate.
22) Grow up the colonies for 3-12 hours in a shaking incubator at 37 °C. Longer is better for this stage because PCR amplifications are increased with greater product and thus easier to visualise.
23) Screen the cloned isolates to check for the correct size insert by carrying out a PCR using 10 µl total volume. Preferably with the primers used for amplification of the insert or the M13F and M13R primers (if these primers are used the product will be ~200bp larger in size).
24) To grow greater amounts of product, larger (5-6 ml) liquid cultures can be set up.
25) Label 20 ml screwtop tubes and add 5-6 ml of LB + amp solution.
26) Add 2 µl of each cloned isolate from the small liquid culture. The cells will grow quicker if taken from a warm culture. If the product has been in the fridge for a long time it is a good idea to grow up another small culture for 6 hours and then use this to inoculate a large culture.
27) Isolate plamid DNA using Qiagen, QIAprep Spin Miniprep Kit.

Protocol 20 – QIAprep® Spin Miniprep Plasmind isolation Kit (Qiagen No. 27104)

1) Centrifuge for 10 minutes.
2) Discard liquid.
3) Resuspend pellet in 250 µl Buffer P1 and transfer to 1.5 ml tube.
4) Add 250 µl Buffer P2 and gently invert the tube 4-6 times to mix.
5) Add 350 µl Buffer N3 and invert the tube immediately but gently 4-6 times.
6) Centrifuge for 10 minutes at 13,000 rpm.
7) Apply supernatant to the QIAprep spin column by decanting or pipetting.
8) Centrifuge 30 seconds. Discard flow-through.
9) Wash the spin column by adding 0.5 ml Buffer PB and centrifuging for 30 seconds. Discard flow-through.
10) Wash QIAprep spin column by adding 0.75 ml Buffer PE and centrifuging for 30 seconds.
11) Discard the flow-through and centrifuge for one minute.
12) Place the QIAprep column in a new 1.5 ml tube. Add 50 µl of Buffer EB or water to the centre of each column and let stand for one minute. Centrifuge for one minute.

Solutions

CTAB buffer
2% cetyltrimethylammonium bromide
20 mM EDTA
100 mM Tris-HCl (pH 8.0)
1.4 M NaCl

TE buffer (1 litre)
108g tris
55g boric acid
7.44g Na_2EDTA (pH 8)

LB medium (1/2 litre)
5g peptone
2.5g yeast extract
2.5g NaCl
500ml water

LB medium for plates (1/2 litre) ~ 22 plates
5g peptone
2.5g yeast extract
2.5g NaCl
500ml water
7.5g agar
After autoclaving add
0.05g ampicillin
1 ml IPTG
0.75 X-gal

IPTG (store at -4°C)
1.2g IPTG
50ml water

X-Gal (store at -20°C)
100mg X-gal
2ml N'N'dimethyl-formamide

From Quiagen kits
Buffer AP1
Buffer AP2
Buffer AP3/E
RNase A stock solution
Buffer AW
Buffer AE
Buffer EB
Buffer PB
Buffer PE
Buffer QG

SOC medium
2g Peptone
0.5g yeast extract
1ml 1M NaCl
0.25ml 1M KCl
97ml distilled water
After autoclaving add
1ml 2M Mg2+
1ml 2M glucose
Additional water until 100ml

References

Doyle, J.J. & Doyle, J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11-15.

Appendix 7 Primer sequences and PCR protocols

Primer List

Primer	Sequence 5' to 3'
ITS4	TCC TCC GCT TAT TGA TAT GC
ITS5	GGA AGT AAA AGT CGT AAC AAG G
17SE	ACG AAT TCA TGG TCC GGT GAA GTG TTC G
GAF1	TTT GGY CAA AAT TAT ATT CGK CC
GAR1	CCA CCA CTK GGT GCA ATC CA
GPAT_5F	TTC TTA TGT CGC CAA CAT GC
GPAT_5R	GCA TGT TGC CGA CAT AAG AA
GPAT_6F	AAA CTG AAG CAG ATC CAG CAA
GPAT_6R	TTG CTG GAT CTG CTT CAG TTT
GPAT_7F	ATC CTG TGT GCA AAC CCT TC
GPAT_7R	ATC CTG TGT GCA AAC CCT TC
GPAT1_I5R4	GTG GCA SGA AGA ACT TAA CAG
GPAT1_I5R6	CAA GTT TGA CAT TTC ATG AAA
GPATOG_I5F1	TAT TGT RAA GTT ACT TGT TTG
GPATOG_I5R2	CAA GTT TGA CAT TTC ATA A
GPAT1_I5R4ALLBUT2n=36	GTG GCA GGA AGA ACT TAA CAG
GPAT1_I5R4_2n=36	GTG GCA CGA AGA ACT TAA CAG
M13F	TGT AAA ACG ACG GCC AGT
M13R	GGA AAC AGC TAT GAC CAT
7:22-2-F1	GACCATGCAATACACCTCAAGA
7:22-2-F2	CTCCTCAGAGAATACAAG
7:22-2-R1	CTCCTATAATAGCCAGTTA
7:22-2-R2	CAACAAATCTGTCAAGGG
3:6-F1	CTTGTTCTCATGACCCATGTC
3:6-F2	CTAGCCTACCACATCTAAAG
3:6-R1	AACCATTCTCACCACCC
3:6-R2	TCGATCACACCGAGGGAA

PCR temperatures for GPAT primer pair combinations

Primer pair	Annealing Temperature °C	Extension time (mins)
Without beads		
GAF1 and GAR1	55	2.5
GPAT_5F and GPAT_R1	53	3
GPAT_5F and GPAT_6R	55	1.5
GPAT_5F and GPAT_7R	54	1.5
With beads		
GPAT_F1 and GPAT_6R	53	1.5
GPAT_5F and GPAT1_15R4	55	1
GPAT_5F and GPAT_I5R6	55	1
GPAT_5F and GPAT_6R	55	1.5
GPAT1_I5R4ALLBUT2n=36 and GPAT_5F	55	1
GPAT1_I5R4_2n=36 and GPAT_5F	55	1
GPAT0G_EX5F1 and GPATOG_I5R2	53	1

Appendix 8 Bayes and r8s Blocks

Bayes Block 1 – ITS dataset

```
#NEXUS
BEGIN DATA;
DIMENSIONS NTAX=281 NCHAR=655;
FORMAT MISSING=? GAP=-  DATATYPE=DNA;
MATRIX
(alignment should be inserted here)
end;
begin mrbayes;
set autoclose=yes nowarn=yes;
log start filename=log.out;
lset nst=6 rates=gamma;
mcmc nruns=1 ngen=10000000 printfreq=100 temp=0.01 nchain=4 savebrlens=yes samplefreq=100
file=allitsresults_a.nex1;
mcmc file=allitsresults_b.nex2;
mcmc file=allitsresults_c.nex3;
sump filename=allitsresults_a.nex1 burnin=10000 printtofile=yes;
sump filename=allitsresults_b.nex2 burnin=10000 printtofile=yes;
sump filename=allitsresults_c.nex3 burnin=10000 printtofile=yes;
sumt filename=allitsresults_a.nex1 burnin=10000 contype=halfcompat;
sumt filename=allitsresults_b.nex2 burnin=10000 contype=halfcompat;
sumt filename=allitsresults_c.nex3 burnin=10000 contype=halfcompat;
end;
```

Bayes Block 2 - GPAT1 dataset

```
#NEXUS
BEGIN DATA;
DIMENSIONS NTAX=175 NCHAR=635;
FORMAT MISSING=? GAP=-  DATATYPE=DNA interleave=no;
MATRIX
(alignment should be inserted here)
end;
begin mrbayes;
set autoclose=yes nowarn=yes;
log start filename=log.out;
exclude 1 - 77;
lset nst=6 rates=gamma;
mcmc nruns=1 ngen=3000000 printfreq=100 temp=0.001 nchain=4 savebrlens=yes samplefreq=100
file=gpat1_a.nex1;
mcmc file=gpat1_b.nex2;
mcmc file=gpat1_c.nex3;
sump filename=gpat1_a.nex1 burnin=10000 printtofile=yes;
sump filename=gpat1_b.nex2 burnin=10000 printtofile=yes;
sump filename=gpat1_c.nex3 burnin=10000 printtofile=yes;
sumt filename=gpat1_a.nex1 burnin=10000 contype=halfcompat;
sumt filename=gpat1_b.nex2 burnin=10000 contype=halfcompat;
sumt filename=gpat1_c.nex3 burnin=10000 contype=halfcompat;
end;
```

Bayes Block 3 - Combined 2-locus dataset

```
#NEXUS
BEGIN DATA;
DIMENSIONS NTAX=89 NCHAR=1934;
FORMAT MISSING=? GAP=-  DATATYPE=DNA;
MATRIX
```

(alignment should be inserted here)
END;
begin mrbayes;
set autoclose=yes nowarn=yes;
log start filename=log.out;
charset its = 1-682;
charset legcyc1a = 683-1934;
partition favored = 2: its, legcyc1a;
set partition = favored;
lset applyto=(1) nst=6 rates=gamma;
lset applyto=(2) nst=6 rates=invgamma;
unlink statefreq=(all) revmat=(all) shape=(all) pinvar=(all);
prset applyto=(all) ratepr=variable;
mcmc ngen=3000000 printfreq=100 temp=0.01 nchain=4 savebrlens=yes samplefreq=100 file=itslegcyc1a_
a.nex1;
mcmc file=itslegcyc1a_b.nex2;
sump filename=itslegcyc1a_a.nex1 burnin=10000 printtofile=yes;
sump filename=itslegcyc1a_b.nex2 burnin=10000 printtofile=yes;
sumt filename=itslegcyc1a_a.nex1 burnin=10000 contype=halfcompat;
sumt filename=itslegcyc1a_b.nex2 burnin=10000 contype=halfcompat;
end;

Bayes Block 4 - Combined 3-locus dataset
#NEXUS
BEGIN DATA;
DIMENSIONS NTAX=108 NCHAR=2478;
FORMAT MISSING=? GAP=- DATATYPE=DNA interleave=no;
MATRIX
(alignment should be inserted here)
end;
begin mrbayes;
set autoclose=yes nowarn=yes;
log start filename=log.out;
charset its = 1-679;
charset legcyc1a = 680-1925;
charset gpat1 = 1926-2478;
exclude 1226 - 1252 1898-1925; [legcyc1a exclude 1226-1252]
partition favored = 3: its, legcyc1a, gpat1;
set partition = favored;
lset applyto=(1,3) nst=6 rates=gamma;
lset applyto=(2) nst=6 rates=invgamma;
unlink statefreq=(all) revmat=(all) shape=(all) pinvar=(all);
prset applyto=(all) ratepr=variable;
mcmc nruns=1 ngen=3000000 printfreq=100 temp=0.001 nchain=4 savebrlens=yes samplefreq=100
file=3loci_a.nex1;
mcmc file=3loci_b.nex2;
mcmc file=3loci_c.nex3;
sump filename=3loci_a.nex1 burnin=10000 printtofile=yes;
sump filename=3loci_b.nex2 burnin=10000 printtofile=yes;
sump filename=3loci_c.nex3 burnin=10000 printtofile=yes;
sumt filename=3loci_a.nex1 burnin=10000 contype=halfcompat;
sumt filename=3loci_b.nex2 burnin=10000 contype=halfcompat;
sumt filename=3loci_c.nex3 burnin=10000 contype=halfcompat;
end;

r8s Block 1 – Cross-validation example

```
#NEXUS
Begin TREES;
TREE    1rep.2100000 = (((((L._micranthus_CEH1982:0.50569700,(L._angustifolius_
CEH1976:0.34350800,(L._luteus_CEH1973:0.10521300,L._hispanicus_Pan1:0.1243050
0):0.45648600):0.07435400):0.19293300,(L._albus_CEH1970:0.57980100,(L._digitatus_
AY338948:0.09684000,L._cosentinii_AY338947:0.11555700):0.59758000):0.08052000):0.0602540
0,(((((((L._magnistipulatus_MP1840:0.00118300,L._reitzii_MP1846:0.02407100):0.07078600,(L.
_paraguariensis_CEH2496:0.21979300,((L._guaraniticus_GregBlue:0.12524700,L._parvifolius_CEH
2480:0.19386000):0.03711100,(L._crotalarioides_CEH2481:0.10689200,(L._velutinus_CEH2478_
2:0.10125500,L._subsessilis_CEH2479:0.00144400):0.08155600):0.03065600):0.05814300):0.1136160
0):0.03480700,L._lanatus_MP1806:0.15699200):0.04815300,(L._albescens_CEH2509:0.04456400,L._
multiflorus_MP1825:0.01857600):0.03301000):0.11323100,L._linearis_MP1798:0.17696500):0.057
68700,((((L._rubriflorus_MP1838:0.02624300,L._paranensis_MPind2_18:0.01098800):0.00118600,L
._uleanus_MP1848:0.01876500):0.03408600,L._gilbertiannus_MP1835:0.06259300):0.00398800,(L.
_bracteolaris_MP1810:0.07758000,L._bandelierae_CEH2321:0.11061300):0.02673900):0.13918200
):0.01191800,(L._harvardii_USDA930123:0.04086900,L._texensis_AY338949:0.08559800):0.181181
00):0.09081300,((L._hirsutissimus_USDA920120:0.27617600,(((L._sierra_blancae_Spell13409:0.204
42300,((((L._arboreus_Spell13488:0.03256900,L._rivularis__AY338937:0.05802100):0.06527000,(L.
_bicolor_USDA284713:0.31002200,(L._lepidus_USDA_PI504439:0.03903100,L._lactus_Spell13415_:
0.02723400):0.10592800):0.02825900):0.02424900,L._sericeus_Lavin777:0.10760500):0.0393180
0,((L._latifolius_Baker10938:0.16371500,L._polyphyllus_Lavin762:0.25255500):0.02406400,(((L._
argenteus_Lavin761:0.21651200,(L._grayi_AY338939_Lavin793:0.11788800,L._breweri_Lavin764:0.28
755900):0.01554000):0.02466500,(L._sp._Kenicer121:0.12161700,((L._sp._CEH2160:0.04117500,L._
montanus_CEH2174:0.07047100):0.02416000,(((L._sp._CEH2107:0.02505100,L._sp._CEH2118:
0.13237100):0.07360800,(((L._proculaustrinus_CEH2353:0.01503800,L._ecuadorian_ann_RJE217:
0.01328200):0.00577900,(L._sp._RJE206:0.05270400,((L._pulvinaris_CEH2273:0.04244700,(((L._
nubigenus_200:0.03872300,L._silver_tomentose_CEH2322:0.02762700):0.00840400,((L._carpapaticus_
RJE158:0.05710900,L._semperflorens_RJE219:0.02238600):0.03534000,L._dillioni_CEH2002:0.04667
500):0.03080500):0.00078700,(((L._altiplani_CEH2264:0.05747900,L._altiplani_CEH2390:0.0213010
0):0.07493000,L._chapra_RJE144:0.05694300):0.00715900,(((L._piurensis_CEH1997:0.02673800,L._
vert._cajamarca_CEH2000:0.00459800):0.02051500,((L._celedin_CEH2004:0.01925700,L._amazonas_
CEH2027:0.04324100):0.01993200,L._sp._CEH2250:0.01583800):0.00515000):0.03016500,(L.
_aff._revolutus_CEH2015:0.05587600,L._aff._huaronensis_CEH2248:0.09388600):0.01678200):0.0
2001500):0.00325200):0.00156600):0.00814300,((((L._sp._RJE188:0.03610300,L._microphyllus_C
EH2272:0.00197900):0.04362100,(L._long_pedicels_RJE59:0.01044100,L._ancash_dwarf_CEH2
239:0.01828000):0.00475900):0.03282900,L._san_marcos_RJE168:0.03688000):0.00070200,((
L._Bolivian_ann_CEH2444:0.00342000,L._weberbaueri_CEH2233:0.11140400):0.02151300,L.
_western_slope_ann_RJE141:0.02696500):0.00456600):0.01561100):0.00527600):0.00166200):0.0
0872700,(L._mutabilis_CEH2009:0.01165500,(L._mollendoensis_Weigend97_703:0.03874800,(L._
tarapacensis_Gardener220A:0.01078400,(L._Porculla_CEH2037:0.03945600,L._ramossisimus_CEH
2010:0.01514500):0.00783000):0.00287600):0.00134600):0.01412400):0.10656200):0.00547300,
L._sp._CEH2117:0.04196000):0.02950500):0.04402100):0.03413900):0.10685100,(L._andersonii_
AY338934:0.25155600,L._neomexicanus_Spell13422:0.04858500):0.05017100):0.00613400):0.03068
400):0.02158200):0.24569400,(L._brevicaulis_Baker10768_W1993:0.07868400,L._truncatus_SS:0.09
827900):0.08791600):0.07761700,L._arizonicus_USDA577285:0.49275800):0.15655200):0.0463170
0,((L._densiflorus_AY338943:0.01959900,L._microcarpus_Gardener1513:0.01493400):0.33518600,L.
_odoratus_CD0311A01:0.14042200):0.31738300):0.08162600):0.25560200):0.05491200,(L._villosus_
CEH1986:0.03743400,L._cumulicola_CEH1983:0.00762500):0.56718800):0.13389320,Genista_tenera_
AY338950:1.20738780);
end;

begin rates;
blformat nsites=1934 lengths=persite;
collapse;
mrca LupinusStem Genista_tenera_AY338950 L._nubigenus_200;
mrca LupinusCrown L._villosus_CEH1986 L._sp._RJE188;
```

mrca 2n36 L._harvardii_USDA930123 L._paranensis_MPind2_18;
mrca 2n36Dispersal L._guaraniticus_GregBlue L._paranensis_MPind2_18;
mrca 2n48 L._densiflorus_AY338943 L._nubigenus_200;
mrca Mexican L._montanus_CEH2174 L._sp._CEH2118;
mrca Andean L._mutabilis_CEH2009 L._celedin_CEH2004;
fixage taxon=LupinusStem age=16.008759;
set verbose=0;
divtime method=pl algorithm=TN crossv=yes cvstart=-2 cvinc=0.1 cvnum=40;
end;

r8s Block 2 – Penalised likelihood example

#NEXUS
Begin TREES;
TREE 1rep.2100000 = (((((L._micranthus_CEH1982:0.50569700,(L._angustifolius_
CEH1976:0.34350800,(L._luteus_CEH1973:0.10521300,L._hispanicus_Pan1:0.1243050
0):0.45648600):0.07435400):0.19293300,(L._albus_CEH1970:0.57980100,(L._digitatus_
AY338948:0.09684000,L._cosentinii_AY338947:0.11555700):0.59758000):0.08052000):0.0602540
0,(((((((L._magnistipulatus_MP1840:0.00118300,L._reitzii_MP1846:0.02407100):0.07078600,(L.
_paraguariensis_CEH2496:0.21979300,((L._guaraniticus_GregBlue:0.12524700,L._parvifolius_CEH
2480:0.19386000):0.03711100,(L._crotalarioides_CEH2481:0.10689200,(L._velutinus_CEH2478_
2:0.10125500,L._subsessilis_CEH2479:0.00144400):0.08155600):0.03065600):0.05814300):0.1136160
0):0.03480700,L._lanatus_MP1806:0.15699200):0.04815300,(L._albescens_CEH2509:0.04456400,L._
multiflorus_MP1825:0.01857600):0.03301000):0.11323100,L._linearis_MP1798:0.17696500):0.057
68700,((((L._rubriflorus_MP1838:0.02624300,L._paranensis_MPind2_18:0.01098800):0.00118600,L
._uleanus_MP1848:0.01876500):0.03408600,L._gilbertiannus_MP1835:0.06259300):0.00398800,(L.
_bracteolaris_MP1810:0.07758000,L._bandelierae_CEH2321:0.11061300):0.02673900):0.13918200
):0.01191800,(L._harvardii_USDA930123:0.04086900,L._texensis_AY338949:0.08559800):0.181181
00):0.09081300,((L._hirsutissimus_USDA920120:0.27617600,(((L._sierra_blancae_Spell13409:0.204
42300,((((L._arboreus_Spell13488:0.03256900,L._rivularis__AY338937:0.05802100):0.06527000,(L.
_bicolor_USDA284713:0.31002200,(L._lepidus_USDA_PI504439:0.03903100,L._lactus_Spell13415_:
0.02723400):0.10592800):0.02825900):0.02424900,L._sericeus_Lavin777:0.10760500):0.0393180
0,((L._latifolius_Baker10938:0.16371500,L._polyphyllus_Lavin762:0.25255500):0.02406400,(((L._
argenteus_Lavin761:0.21651200,(L._grayi_AY338939_Lavin793:0.11788800,L._breweri_Lavin764:0.28
755900):0.01554000):0.02466500,(L._sp._Kenicer121:0.12161700,((L._sp._CEH2160:0.04117500,L._
montanus_CEH2174:0.07047100):0.02416000,(((L._sp._CEH2107:0.02505100,L._sp._CEH2118:
0.13237100):0.07360800,(((L._proculaustrinus_CEH2353:0.01503800,L._ecuadorian_ann_RJE217:
0.01328200):0.00577900,(L._sp._RJE206:0.05270400,((L._pulvinaris_CEH2273:0.04244700,(((L._
nubigenus_200:0.03872300,L._silver_tomentose_CEH2322:0.02762700):0.00840400,((L._carpapaticus_
RJE158:0.05710900,L._semperflorens_RJE219:0.02238600):0.03534000,L._dillioni_CEH2002:0.04667
500):0.03080500):0.00078700,(((L._altiplani_CEH2264:0.05747900,L._altiplani_CEH2390:0.0213010
0):0.07493000,L._chapra_RJE144:0.05694300):0.00715900,(((L._piurensis_CEH1997:0.02673800,L._
vert._cajamarca_CEH2000:0.00459800):0.02051500,((L._celedin_CEH2004:0.01925700,L._amazonas_
CEH2027:0.04324100):0.01993200,L._sp._CEH2250:0.01583800):0.00515000):0.03016500,(L.
_aff._revolutus_CEH2015:0.05587600,L._aff._huaronensis_CEH2248:0.09388600):0.01678200):0.0
2001500):0.00325200):0.00156600):0.00814300,((((L._sp._RJE188:0.03610300,L._microphyllus_C
EH2272:0.00197900):0.04362100,(L._long_pedicels_RJE59:0.01044100,L._ancash_dwarf_CEH2
239:0.01828000):0.00475900):0.03282900,L._san_marcos_RJE168:0.03688000):0.00070200,((
L._Bolivian_ann_CEH2444:0.00342000,L._weberbaueri_CEH2233:0.11140400):0.02151300,L.
_western_slope_ann_RJE141:0.02696500):0.00456600):0.01561100):0.00527600):0.00166200):0.0
0872700,(L._mutabilis_CEH2009:0.01165500,(L._mollendoensis_Weigend97_703:0.03874800,(L._
tarapacensis_Gardener220A:0.01078400,(L._Porculla_CEH2037:0.03945600,L._ramossisimus_CEH
2010:0.01514500):0.00783000):0.00287600):0.00134600):0.01412400):0.10656200):0.00547300,
L._sp._CEH2117:0.04196000):0.02950500):0.04402100):0.03413900):0.10685100,(L._andersonii_
AY338934:0.25155600,L._neomexicanus_Spell13422:0.04858500):0.05017100):0.00613400):0.03068
400):0.02158200):0.24569400,(L._brevicaulis_Baker10768_W1993:0.07868400,L._truncatus_SS:0.09
827900):0.08791600):0.07761700,L._arizonicus_USDA577285:0.49275800):0.15655200):0.0463170
0,((L._densiflorus_AY338943:0.01959900,L._microcarpus_Gardener1513:0.01493400):0.33518600,L.

_odoratus_CD0311A01:0.14042200):0.31738300):0.08162600):0.25560200):0.05491200,(L._villosus_CEH1986:0.03743400,L._cumulicola_CEH1983:0.00762500):0.56718800):0.13389320,Genista_tenera_AY338950:1.20738780);
end;
begin rates;
blformat nsites=1934 lengths=persite;
collapse;
mrca LupinusStem Genista_tenera_AY338950 L._nubigenus_200;
mrca LupinusCrown L._villosus_CEH1986 L._sp._RJE188;
mrca 2n36 L._harvardii_USDA930123 L._paranensis_MPind2_18;
mrca 2n36Dispersal L._guaraniticus_GregBlue L._paranensis_MPind2_18;
mrca 2n48 L._densiflorus_AY338943 L._nubigenus_200;
mrca Mexican L._montanus_CEH2174 L._sp._CEH2118;
mrca Andean L._mutabilis_CEH2009 L._celedin_CEH2004;
fixage taxon=LupinusStem age=16.008759;
set penalty=add;
set smoothing=5;
divtime method=pl algorithm=TN;
set checkgradient=yes;
showage;
describe plot=chronogram;
set num_time_guesses=3;
end;

Appendix 9 Age estimates

ITS/ LEGCYC1a dating			*Lupinus* Stem fixed at 16.00 MYR					
Tree	Smoothing Value	*Lupinus* Crown	2n=36	2n=36 dispersal	2n=48	*L. microcarpus*	Mexican	Andean
1	13	13.35	5.60	5.36	8.08	0.22	2.03	1.11
2	13	14.74	5.06	4.67	7.92	0.27	2.02	1.73
3	13	14.64	4.98	4.62	8.07	0.25	2.31	1.61
4	13	14.57	6.71	5.56	9.39	0.30	2.96	1.62
5	13	14.29	5.80	5.22	8.59	0.32	2.25	1.67
6	13	15.20	11.58	10.84	13.22	9.86	1.67	0.61
7	13	14.56	6.02	5.70	9.16	0.28	1.70	1.49
8	13	14.47	5.58	5.37	8.75	0.04	1.76	1.62
9	13	14.77	6.40	5.73	10.37	0.54	1.90	1.65
10	13	15.07	7.22	5.39	10.58	0.25	2.19	1.58
11	13	14.09	7.97	5.81	9.07	0.86	2.08	1.43
12	13	13.94	6.37	4.82	8.84	0.38	2.16	1.46
13	13	14.07	6.82	5.94	9.17	0.41	1.53	1.33
14	13	14.59	6.18	4.95	9.31	0.30	2.13	1.73
15	13	13.61	5.74	4.81	9.69	0.49	1.80	1.72
16	13	13.49	6.83	6.06	8.63	0.17	1.57	1.45
17	13	14.35	5.70	4.88	8.38	0.43	1.73	1.72
18	13	14.89	6.86	5.34	8.96	0.70	1.75	1.69
19	13	13.81	6.21	4.58	9.42	0.19	1.96	1.58
20	13	14.00	3.60	5.22	8.33	0.48	2.25	1.53
21	13	13.08	5.36	4.49	8.42	0.36	1.69	1.20
22	13	14.73	6.33	5.10	9.37	0.75	1.92	1.33
23	13	14.19	7.98	6.94	9.84	1.04	2.23	1.45
24	13	14.09	6.72	5.36	9.35	0.44	2.27	1.62
25	13	14.74	6.08	5.27	9.55	0.90	2.00	1.80
26	13	14.41	6.04	5.10	8.42	0.46	1.74	1.52
27	13	14.12	6.55	5.40	8.17	0.28	1.92	1.70
28	13	15.17	6.53	6.07	9.34	0.46	1.77	1.57
29	13	13.72	5.18	5.14	8.87	1.08	1.72	1.35
30	13	12.99	5.10	4.27	7.69	0.27	1.77	1.34
31	13	14.10	6.47	5.46	10.51	0.54	1.62	0.99
32	13	15.08	6.37	5.37	10.40	0.27	2.11	1.45
33	13	14.03	5.92	4.90	8.90	0.64	2.31	1.34
34	13	13.68	5.35	4.33	8.19	0.59	1.43	1.82
35	13	14.69	6.62	4.88	8.77	0.27	2.96	1.88
36	13	14.34	7.31	6.31	9.12	0.19	1.67	1.38
37	13	13.69	6.35	5.66	9.13	0.31	2.29	1.34
38	13	11.92	4.71	3.88	7.17	0.46	1.74	1.41
39	13	15.42	8.52	7.04	10.61	0.38	2.50	1.77
40	13	13.34	6.19	5.31	7.80	0.49	1.27	1.21
41	13	14.06	6.69	5.92	9.30	0.37	1.59	1.73
42	13	14.77	5.66	4.96	8.50	0.63	1.54	1.57
43	13	13.50	6.99	6.26	8.83	0.24	1.64	1.59
44	13	14.43	4.66	4.28	8.33	0.06	2.00	1.38
45	13	15.19	11.59	9.27	13.20	n/a	2.58	0.52
46	13	14.30	7.78	6.20	12.32	4.13	0.56	0.41
47	13	13.88	6.42	5.72	9.68	0.35	1.87	1.33

48	13	14.39	6.41	5.59	9.09	0.14	2.01	1.86
49	13	13.28	5.59	5.00	9.03	0.24	1.76	1.46
50	13	12.05	4.57	4.16	6.98	0.28	1.77	1.27
51	13	13.72	5.02	4.34	8.26	0.26	1.51	1.47
52	13	14.49	6.17	5.50	10.15	0.49	1.58	1.48
53	13	13.92	6.20	4.96	9.39	0.64	1.76	1.42
54	13	13.13	5.85	4.51	7.99	0.41	1.86	1.71
55	13	13.98	6.41	5.71	8.9	0.39	1.59	1.42
56	13	14.60	5.02	4.65	9.69	0.30	1.97	1.38
57	13	13.34	5.84	4.88	7.67	0.42	1.40	1.42
58	13	11.91	5.58	4.58	7.91	0.29	1.41	1.35
59	13	12.48	5.66	5.13	8.58	0.77	1.73	1.51
60	13	15.30	8.14	5.01	8.51	0.33	2.32	1.54
61	13	7.65	3.68	3.19	6.32	0.21	0.98	0.93
62	13	14.34	6.20	5.44	10.76	0.59	2.37	1.67
63	13	14.20	5.22	4.82	9.57	0.64	2.67	1.59
64	13	13.56	6.35	4.93	8.3	0.36	1.96	1.40
65	13	14.65	5.54	5.12	9.03	0.17	2.25	1.34
66	13	13.89	5.95	4.83	9.17	0.44	2.10	1.55
67	13	14.51	7.17	5.59	8.91	1.08	2.39	1.89
68	13	13.41	5.32	4.73	8.22	0.11	1.84	1.31
69	13	14.70	5.51	4.53	7.84	0.52	2.35	0.98
70	13	14.43	6.48	5.70	9.54	0.30	1.65	1.77
71	13	13.65	6.64	5.84	8.69	0.33	2.76	1.61
72	13	14.06	6.37	5.71	8.95	0.32	2.47	1.38
73	13	14.64	6.26	5.19	9.77	0.24	1.78	1.72
74	13	13.79	5.57	4.72	8.59	0.48	2.14	1.86
75	13	14.63	6.46	5.15	9.02	0.30	2.12	1.28
76	13	14.05	6.29	5.60	9.10	1.15	1.97	1.63
77	13	15.10	8.06	6.92	10.36	0.74	2.33	2.08
78	13	14.38	5.77	5.04	9.07	0.61	1.52	1.34
79	13	14.32	6.04	4.57	9.18	0.17	1.82	1.20
80	13	14.36	7.75	6.04	10.53	0.45	2.31	1.44
81	13	11.35	5.43	4.36	7.10	0.36	1.25	1.13
82	13	14.79	6.14	5.52	9.75	0.79	1.97	2.09
83	13	14.39	6.57	5.73	9.77	0.35	2.17	1.79
84	13	14.07	5.57	4.72	8.65	0.34	1.13	1.58
85	13	15.02	6.71	5.27	9.99	0.82	1.72	1.39
86	13	14.18	5.69	4.53	9.05	0.4	2.2	1.32
87	13	13.51	7.23	6.12	7.16	0.22	1.85	1.85
88	13	12.83	5.59	3.99	8.58	0.29	1.59	1.38
89	13	11.96	4.90	4.38	6.79	0.57	1.76	1.30
90	13	13.03	4.98	4.39	7.56	0.51	1.80	1.17
91	13	13.56	7.22	5.34	9.09	0.54	1.34	1.44
92	13	12.62	5.01	4.71	8.03	0.37	2.07	1.07
93	13	14.04	5.78	4.98	9.15	0.53	2.35	1.64
94	13	15.36	6.64	5.63	8.51	0.33	1.79	1.46
95	13	12.47	5.82	4.92	9.13	0.29	1.68	1.45
96	13	14.27	6.44	4.86	8.53	0.41	1.94	1.95
97	13	11.76	5.11	4.47	8.14	0.19	1.81	1.51
98	13	13.63	5.61	5.61	8.46	0.40	2.07	1.21
99	13	12.89	6.31	4.83	8.46	0.67	1.88	1.83
100	13	8.65	4.33	3.47	5.67	0.23	1.48	1.04

Mean		13.87	6.19	5.25	8.94	0.56	1.91	1.47
Mode		14.74	6.37	4.88	9.39	0.30	1.76	1.34
Min.		7.65	3.60	3.19	5.67	0.04	0.56	0.41
Max.		15.42	11.59	10.84	13.22	9.86	2.96	2.09
Size of range		7.77	7.99	7.65	7.55	9.82	2.40	1.68
Standard deviation		1.18	1.19	0.97	1.16	1.04	0.39	0.29
		Lupinus **Crown**	**2n=36**	**2n=36 dispersal**	**2n=48**	***L . microcarpus***	**Mexican**	**Andean**

ITS/ LEGCYC1a dating		***Lupinus* Stem fixed at 21.15 MYR**						
Tree	**Smoothing Value**	***Lupinus* Crown**	**2n=36**	**2n=36 dispersal**	**2n=48**	***L. microcarpus***	**Mexican**	**Andean**
1	13	17.67	7.46	7.11	10.48	0.29	2.66	1.45
2	13	19.53	6.74	6.21	10.50	0.35	2.67	2.28
3	13	19.39	6.62	6.14	10.72	0.34	3.05	2.13
4	13	19.34	8.95	7.43	12.46	0.39	3.90	2.12
5	13	18.95	7.72	6.95	11.38	0.42	2.95	2.18
6	13	19.39	9.09	5.23	10.41	0.76	3.43	1.85
7	13	19.29	8.00	7.57	12.13	0.37	2.24	1.96
8	13	19.16	7.42	7.14	11.59	0.48	2.35	2.12
9	13	19.55	8.52	7.63	13.75	0.72	2.50	2.17
10	13	19.94	9.59	7.17	14.02	0.33	2.88	2.07
11	13	18.71	10.63	7.75	12.04	1.14	2.74	1.87
12	13	18.46	8.46	6.41	11.71	0.50	2.83	1.94
13	13	18.64	9.08	7.91	12.15	0.54	2.01	1.74
14	13	19.32	8.21	6.57	12.33	0.40	2.80	2.27
15	13	18.04	7.64	6.41	12.85	0.65	2.38	2.26
16	13	17.87	9.09	8.06	11.75	0.22	2.07	1.91
17	13	19.01	7.57	6.48	11.11	0.57	2.29	2.27
18	13	19.71	9.13	7.11	11.87	0.93	2.31	2.22
19	13	18.29	8.24	6.07	12.48	0.25	2.57	2.08
20	13	18.55	8.38	6.94	11.05	0.63	2.97	2.02
21	13	17.32	7.13	5.97	11.16	0.47	2.23	1.58
22	13	19.49	8.41	6.78	12.41	0.99	2.52	1.74
23	13	18.79	10.59	9.21	13.04	1.38	2.94	1.91
24	13	18.66	8.93	7.12	12.39	0.58	2.97	2.12
25	13	19.51	8.08	7.01	12.64	1.20	2.62	2.35
26	13	19.07	8.02	6.77	11.16	0.61	2.30	2.01
27	13	18.69	8.69	7.17	10.81	0.36	2.53	2.23
28	13	20.08	8.69	8.08	12.36	0.62	2.31	2.05
29	13	18.17	6.89	6.83	11.75	1.43	2.27	1.78
30	13	17.20	6.76	5.67	10.18	0.36	2.33	1.76
31	13	18.68	8.63	7.29	13.95	0.72	2.13	1.31
32	13	19.96	8.49	7.17	13.80	0.37	2.78	1.90
33	13	18.60	7.89	6.53	11.80	0.86	3.04	1.76
34	13	18.16	7.12	5.77	10.89	0.78	1.89	2.41
35	13	19.44	8.79	6.48	11.61	0.36	3.88	2.45
36	13	19.01	9.71	8.38	12.09	0.26	2.20	1.82
37	13	18.16	8.45	7.54	12.11	0.42	3.02	1.76
38	13	15.78	6.26	5.15	9.51	0.60	2.30	1.86
39	13	20.40	11.32	9.35	14.03	0.50	3.28	2.33
40	13	17.68	8.21	7.04	10.31	0.64	1.67	1.58
41	13	18.61	8.86	7.87	12.31	0.49	2.09	2.28
42	13	19.57	7.55	6.62	11.27	0.84	2.03	2.07
43	13	17.89	9.29	8.34	11.71	0.31	2.15	2.08
44	13	19.11	6.20	5.70	11.05	0.08	2.64	1.81
45	13	18.31	6.74	5.88	11.49	0.29	3.09	1.37
46	13	19.34	8.41	6.48	11.68	0.40	3.09	2.02
47	13	18.41	8.58	7.64	12.86	0.47	2.46	1.75
48	13	19.07	8.53	7.45	12.06	0.18	2.64	2.44

49	13	17.61	7.44	6.65	11.97	0.32	2.33	1.92
50	13	15.96	6.07	5.53	9.25	0.37	2.33	1.60
51	13	18.16	6.66	5.76	10.93	0.34	1.99	1.93
52	13	19.18	8.19	7.30	13.44	0.65	2.07	1.94
53	13	18.44	8.25	6.60	12.47	0.86	2.31	1.86
54	13	17.40	7.77	5.99	10.60	0.55	2.46	2.25
55	13	18.48	8.49	7.57	11.78	0.52	2.09	1.87
56	13	19.33	6.68	6.19	12.84	0.40	2.59	1.81
57	13	17.67	7.76	9.49	10.15	0.56	1.84	1.86
58	13	15.77	7.41	6.09	10.47	0.39	1.84	1.77
59	13	16.53	7.51	6.81	11.37	1.02	2.27	1.98
60	13	20.23	10.80	6.66	11.28	0.44	3.05	2.02
61	13	10.11	4.87	4.22	8.35	0.28	1.29	1.22
62	13	19.00	8.24	7.24	14.24	0.78	3.09	2.18
63	13	18.80	7.33	6.41	12.66	0.85	3.50	2.08
64	13	17.96	8.43	6.55	10.99	0.48	2.59	1.85
65	13	19.40	7.36	6.80	11.96	0.22	2.97	1.77
66	13	18.39	7.90	6.41	12.14	0.57	2.75	2.03
67	13	19.23	9.55	7.44	11.80	1.44	3.15	2.49
68	13	17.83	7.11	6.33	10.93	0.14	2.43	1.72
69	13	19.47	7.31	6.01	10.39	0.69	3.09	1.28
70	13	19.11	8.61	7.57	12.64	0.39	2.34	2.18
71	13	18.11	8.86	7.79	11.51	0.44	3.63	2.11
72	13	18.64	8.47	7.60	11.86	0.43	3.24	1.81
73	13	19.37	8.32	6.90	12.97	0.31	2.36	2.28
74	13	18.28	7.41	6.28	11.39	0.63	2.81	2.44
75	13	19.36	8.58	6.84	11.95	0.40	2.79	1.68
76	13	18.63	8.37	7.46	12.08	1.53	2.60	2.15
77	13	20.00	10.72	9.21	13.74	0.98	3.06	2.72
78	13	20.07	16.82	14.79	15.21	3.50	0.98	0.82
79	13	17.55	7.05	5.31	10.69	0.20	1.97	1.21
80	13	n/a	n/a	n/a	n/a	n/a	n/a	n/a
81	13	15.02	7.20	5.79	9.39	0.47	1.65	1.48
82	13	19.59	8.16	7.34	12.92	1.04	2.59	2.75
83	13	19.06	8.73	7.62	12.94	0.46	2.84	2.35
84	13	18.64	7.40	6.28	11.47	0.45	1.49	2.08
85	13	20.02	9.56	6.99	10.58	5.19	3.65	0.90
86	13	18.76	7.55	9.01	11.97	0.53	2.90	1.73
87	13	17.93	9.65	8.17	9.51	0.29	2.44	2.44
88	13	16.98	7.42	5.30	11.37	0.38	2.09	1.81
89	13	15.82	6.49	5.81	8.98	0.75	2.31	1.71
90	13	17.25	6.60	5.83	9.99	0.67	2.35	1.53
91	13	18.01	9.62	7.12	12.07	0.71	1.77	1.90
92	13	16.71	6.65	6.25	10.63	0.48	2.72	1.40
93	13	18.61	7.68	6.62	12.12	0.71	3.09	2.15
94	13	20.30	8.82	7.47	11.28	0.44	2.36	1.92
95	13	16.55	7.75	6.56	12.14	0.38	2.22	1.91
96	13	18.88	8.56	6.46	11.30	0.55	2.55	1.91
97	13	15.58	6.79	5.93	10.78	0.25	2.38	1.98
98	13	18.07	7.47	7.47	11.20	0.54	2.72	1.58
99	13	17.05	8.36	6.41	11.18	0.89	2.46	2.39
100	13	11.43	5.73	4.59	7.48	0.30	1.94	1.36

Mean		18.35	8.16	6.93	11.64	0.64	2.54	1.93
Mode		18.64	6.74	6.41	11.71	0.40	3.09	1.91
Min		10.11	4.87	4.22	7.48	0.08	0.98	0.82
Max		20.40	16.82	14.79	15.21	5.19	3.90	2.75
Size of range		10.29	11.95	10.57	7.73	5.11	2.92	1.93
Standard deviation		1.57	1.43	1.26	1.27	0.62	0.52	0.35
		Lupinus **Crown**	**2n=36**	**2n=36 dispersal**	**2n=48**	***L . microcarpus***	**Mexican**	**Andean**

Appendix 10 ISSR primers

Name	Sequence	Source	Anchor position	Anchor Length	Repeat Length	Length	G/C content
R1	$(AG)_8CG$	Modified from UBC835, Talhinas *et al* (2003)	3'	2	2	18	10/18
R2	$CCCA(AG)_7$	MF21,Yang *et al* (2001)	5'	4	2	18	7/18
R3	$GAG(TC)_8$	MF121,Yang *et al* (2001)	5'	3	2	17	9/17
R4	$GTCT(AGC)_5$	Modified from MF42, Yang *et al* (2001)	5'	4	3	19	12/19
R5	$GGA(GCG)_5$	New, RJE	5'	3	3	18	17/18
R6	$(CTC)_5GC$	Modified from MAO, Wolfe Lab Website (2000)	3'	2	3	17	12/17
R7	$CA(CT)_8$	Modified from K7, Wu *et al* (1994)	5'	2	2	19	9/18
R8	$GGG(AAC)_5$	MF51,Yang *et al* 2001	5'	3	3	18	8/18
R9	$CC(AGA)_5$	Modified from MF43, Yang *et al* (2001)	5'	2	3	17	7/17
R10	$(GAG)_5TCG$	Modified from OMAR, Wolfe Lab Website (2000)	3'	3	3	18	12/18
R11	$(CAC)_5TG$	Modified from MANNY, Wolfe Lab Website (2000)	3'	2	3	17	11/17
R12	$(AC)_7GAT$	Modified from UBC856, Talhinas *et al* (2003)	3'	3	2	17	8/17
R13	$(GT)_7GC$	Modified from UBC850, Talhinas *et al* (2003)	3'	2	2	18	10/18
R14	$(AC)_8CG$	Modified from UBC856, Talhinas *et al* (2003)	3'	2	2	18	10/18
R15	$AAA(TC)_7$	New, RJE	5'	3	2	17	7/17
R16	$CG(GT)_8$	New, RJE	5'	3	2	18	10/18
R17	$GTT(AGC)_6$	New, RJE	5'	3	3	21	13/21
R18	$CA(GCG)_5$	New, RJE	5'	2	3	17	16/17
R19	$GCAT(GCG)_5$	New, RJE	5'	4	3	19	17/19
R20	$(AC)_5TTGC$	New, RJE	3'	4	2	18	9/18
R21	$(AGC)_5GTCA$	New, RJE	3'	4	3	19	12/19
R22	$(CTC)_5AG$	New, RJE	3'	3	3	17	11/17
R23	$(TC)_8GAG$	New, RJE	3'	3	2	17	9/17
R24	$(GCG)_6GT$	New, RJE	3'	2	3	20	19/20

References

Talhinhas, P., J. M. Neves-Martins, AND J. Leitão. (2003) AFLP, ISSR and RAPD markers reveal high levels of genetic diversity among *Lupinus* spp. *Plant Breeding* **122**, 507-510.

Yang, H., M. W. Sweetingham, W. A. Cowling, AND P. M. C. Smith. (2001) DNA fingerprinting based on microsatellite-anchored fragment length polymorphisms, and isolation of sequence-specific PRC markers in lupin (*L. angustifolius* L.). Molecular Breeding **7**, 203-209.

Wolfe Lab website. (2000) http://www.biosci.ohio-state.edu/~awolfe/ISSR/protocols.ISSR.html

Wu, K.-S., R. Jones, L. Danneberger, AND P. A. Scolnik. (1994) Detection of microsatellite polymorphisms without cloning. *Nucleic Acids Research* **22**, 3257-3258.

Appendix 11 ISSR primer combinations

5'horizontal, 3' vertical	R2	R3	R4	R5	R7	R8	R9	R15	R16	R17	R18	R19
	G: bright amplification but bands <500bp	**C:** smear	**H:** faint, <500bp	**i:** faint, few bands	**AL**: smear	**Y**: poor amplification	**AJ**: no amplification	**1:15**	**1:16**	**1:17**	**1:18**	**1:19**
R6	**D:** some bands but similar size ones faint	**E:** good amp candidate band **E1** 900bp. Repeatable.	**B:** several bands but small	**f:** no amplification	**B**: bands <500bp	**AB**:- no amplification	**X:** bands <500bp	**6:15**	**6:16:**	**6:17**	**6:18**	**6:19**: poor amplification
R10	**J**: no amplification	**K**: no amplification	**l**: no amplification	**m**: no amplification	**AY**: faint amplification	**AZ**: no amplification	**BA**: 1.2 kb band faint repeatable	**10:15**	**10:16**	**10:17:**	**10:18**	**10:19**: poor amplification
R11	**AK**: faint 650bp band	**P**: no amplification	**AL**: bands <500bp	**AM**: no amplification	**BC**: 800bp band repeatable	**BD**: poor amplification	**AD**: faint bands <500bp	**11:15**	**11:16**	**11:17**	**11:18**	**11:19**
R12	**O**: faint band at 600bp in 2 taxa	**AE**: one 1kb band - not repeated	**AR**: 700bp band	**AS**: bands faint	**U**: nice amplification but bands <600bp	**BB**: 1.1 kb band	**AC**: poor amplification	**12:15**	**12/16**: X too many bands	**12:17**: no amplification	**12/18**: 1kb band repeatable	**12/19**: X small bands
R13	**AG**: smear	**AF**: smear	**AT**: no amplification	**AU**: 800bp band but amplification smeary	**AO**: smear	**AP**: smear	**S**: Band S1 1.2 kb	**13/15**: poor amplification	**13:16**	**13:17**	**13:18**	**13:19**
R14	**AH**: poor amplification	**W**: poor amplification	**AW**: 800bp band repeatable	**AX**: faint amplification	**BE**: poor amplification	**T**: bands in Andean taxa <500bp	**R**: band R1 870bp	**14:15**	**14:16**	**14:17**	**14:18**	**14:19**
R20	**02:20**	**03:20**	**04:20**	**05:20**	**07:20**	**08:20**	**09/20:** 800bp band repeatable	**15/20**: smear	**16:20**	**17:20**	**18:20**	**19:20**
R21	**02:21**	**03:21**	**04:21**	**05:21**	**07:21**	**08:21**	**09/21:** faint	**15/21**: smear	**16:21**	**17:21**: poor amplification	**18:21**	**19:21**
R22	**02:22**: 750bp + 1.4kb bands	**03:22**	**04:22**	**05:22**	**07/22**: 750bp band repeatable	**08:22**	**09/22:** 900bp repeatable	**15/22:**	**16:22**	**17:22**	**18:22**	**19/22**: 850bp + 1kb repeatable in some
R23	**02:23**	**03:23**	**04:23**	**05:23**	**07:23**	**08:23**	**09/23:** 600bp	**15:23**	**16:23**	**17:23**	**18:23**	**19:23**
R24	**02:24**	**03:24**	**04:24**	**05:24**	**07:24**	**08:24**	**09/24:** faint bands	**15:24**	**16:24**	**17:24**	**18:24**	**19:24**

Appendix 12 Comparison of ISSR-based SCAR loci

	BC	E1	R1	7/22-1	7/22-2	9/20	9/22	12/18
No. species	5	10	5	4	6	3	2	4
Poss copy No. issue	no	yes	no	no	no	yes	no	yes
GenBank Match	*Pisum sativum* mtDNA tRNA-Ser	none	*Solanum demissum*, *Arabidopsis thaliana* uncharacterised nDNA	none	none	none	none	mt-nad2
Aligned Length	677	720	838	828	843	797	910	1002
No var. cha rs.	12	22	19	9	13	19	0	n/a*
No. PIC	6	7	1	0	1	0	n/a	n/a*
No. indels	0	10	4	3	1	2	0	n/a*
No. informative indels	0	1	3	0	0	0	0	n/a*
% PIC	0.9	1	0.1	0	0.1	0	n/a	n/a*
ITS aligned length	641	641	641	641	641	641	641	641
ITS No. var. chars.	25	25	5	6	7	1	1	2
ITS No. PIC	2	4	2	2	2	0	n/a	0
ITS % PIC	0.3	0.6	0.3	0.3	0.3	0	n/a	0
ITS No. indels	0	5	2	0	0	4	0	0
ITS No. informative indels	0	0	0	0	0	0	0	0
GPAT aligned length	549	549	549	549	549	549	549	549
GPAT1 No. var. chars.	21	25	13	23	18	1	3	18
GPAT1 No. PIC	3	5	3	1	4	0	n/a	2
GPAT % PIC	0.5	0.9	0.5	0.2	0.7	0	n/a	0.4
GPAT1 No. indels	6	7	9	7	7	1	1	4
GPAT1 No. informative indels	0	1	6	0	0	0	0	0
LEG aligned length	1279	1279	1279	1279	1279	1279	1279	1279
LEGCYC1A No. var. chars.	25	63	32	29	22	4	6	20
LEGCYC1A No. PIC	3	5	3	3	3	1	n/a	0
LEGCYC1A % PIC	0.2	0.4	0.2	0.2	0.2	0.1	n/a	0
LEGCYC1A No. indels	6	7	4	4	7	0	3	3
LEGCYC1A No. informative indels	0	0	0	0	0	0	0	0

Appendix 13 Seed data

Species	Region*	Collector Name & Number	Voucher material (Herbaria) (including seedling vouchers)	Weight (g per 25 seeds)	Mean width (mm)	Mean length (mm)	Mean depth (mm)	Shape	Colour	Pattern	Rim Aril	Seed coat surface pattern
L. albifrons	NA	Kew Seed database	-	0.69	-	-	-	-	-	-	-	-
L. albifrons spp. abramsii	NA	Kew Seed database	-	0.76	-	-	-	-	-	-	-	-
L. albifrons var. douglasii	NA	Kew Seed database	-	0.86	-	-	-	-	-	-	-	-
L. albus	OW	J. Lopez and P. D. Guerrero26/02/1992	-	5.7	8.92	9.65	4.22	flattened cuboide	cream	plain	none	granular
L. albus	OW	Kew Seed database	-	7	-	-	-	-	-	-	-	-
L. albus 'Dieta'	OW	Soya UK s.n.	-	6.2	-	-	-	flattened cuboide	cream	plain	none	-
L. alopecuroides	AND	CEH2241	FHO, K, MOL	0.27	2.49	3.51	1.9	flattened cuboide, flattened oval	black, brown, white	marbled with eyebrown	none	-
L. andersoni	NA	Kew Seed database	-	0.88	-	-	-	-	-	-	-	-
L. angustifolius	OW	J. Lopez19/03/1991	-	2.14	4.6	5.75	4.16	flattened cuboide	browns, creams, greys	marbled	none	papillose
L. angustifolius	OW	Kew Seed database	-	3.99	-	-	-	-	-	-	-	-
L. angustifolius	OW	Soya UK s.n.	-	-	-	-	-	cuboide	cream, pale pink, white	plain	none	papillose
L. arboreus	NA	Kew Seed database	-	0.68	-	-	-	-	-	-	-	-
L. arbustus	NA	Kew Seed database	-	0.63	-	-	-	-	-	-	-	-
L. argenteus var. heteranthus	NA	Kew Seed database	-	0.03	-	-	-	-	-	-	-	-
L. arizonicus	NA	Kew Seed database	-	0.18	-	-	-	-	-	-	-	-
L. arizonicus	NA	USDA PI 577285	FHO (RJE241)	0.12	2.08	2.61	1.42	flattened oval	brown, cream	marbled	none	rugulate
L. arvensis	AND	RJE221	FHO, QCNE	-	-	-	-	-	-	-	-	-
L. austrobicolor	AND	CEH2000	CAS, FHO, K, MO, MOL, NY	0.32	-	-	-	-	-	-	-	-
L. austrobicolor	AND	CEH2001	FHO, H, MOL, NY	0.35	2.64	3.76	2	flattened cuboide	browns, cream, black	marbled with eyebrow	none	-
L. bandelierae	AND	CEH2301	FHO, K, LPB	0.42	-	-	-	-	-	-	-	-
L. bandelierae	AND	CEH2318	FHO, K, LPB	0.63	3.45	4.5	2.05	flattened cuboide, flattened sphere	browns, cream, grey, reds	marbled	none	-
L. bandelierae	AND	CEH2321	FHO, K, LPB	-	3.24	4.15	1.85	flattened cuboide	black, brown, cream	marbled	none	-
L. benthamii	NA	Kew Seed database	-	0.23	-	-	-	-	-	-	-	-

L. bicolor	NA	Kew Seed database	-	0.25	-	-	-	-	-	-	-	-
L. bicolor	NA	USDA PI284713	-	0.18	-	-	-	flattened cuboide, flattened oval	brown, cream	marbled	none	rugulate
L. bracteolaris	SA	s.n.10/12/2002	ICN	0.23	2.41	2.86	1.86	flattened cuboide	black, browns, cream, greys	marbled	none	-
L. bracteolaris	SA	S. T. S. Miotto1684	ICN	0.32	-	-	-	flattened cuboide	black, browns, cream, greys	marbled	none	-
L. bracteolaris	SA	S. T. S. Miotto & M. Pinheiro1797	ICN	0.28	-	-	-	flattened cuboide	black, browns, cream, greys	marbled	none	-
L. bracteolaris	SA	S. T. S. Miotto & M. Pinheiro1810	ICN, FHO (RJE235)	0.3	-	-	-	flattened cuboide	black, browns, cream, greys	marbled	none	-
L. bracteolaris	SA	S. T. S. Miotto & M. Pinheiro1819	ICN	0.24	-	-	-	flattened cuboide	black, browns, cream, greys	marbled	none	rugulate
L. brevicaulis	NA	C. D. Bailey242	FHO	-	-	-	-	flattened cuboide, flattened oval	black, browns, cream, greys	marbled	none	rugulate
L. chachas	AND	CEH2027	CAS, F, FHO, K, MO, MOL, NY	0.88	-	-	-	flattened cuboide	browns, creams	marbled	none	-
L. chamissonis	NA	Kew Seed database	-	0.5	-	-	-	-	-	-	-	-
L. chrysanthus	AND	CEH2264	FHO, K, LPB	0.21	2.48	3.55	1.48	flattened cuboide, flattened oval	black, browns, cream, greys	marbled	none	-
L. consentinii	OW	USDA s.n.	-	-	-	-	-	-	-	-	-	fascicular
L. croceus	NA	Kew Seed database	-	1.04	-	-	-	-	-	-	-	-
L. culbertsonii ssp.culbertsonii	NA	Kew Seed database	-	0.12	-	-	-	-	-	-	-	-
L. cumulicola	NA	CEH1983	FHO, K	0.46	-	-	-	-	-	-	-	simple-foveolate
L. cumulicola	NA	CEH1984	FHO, K	0.5	-	-	-	-	-	-	-	simple-foveolate
L. cumulicola	NA	CEH1985	FHO, K	0.61	3.86	4.53	1.96	flattened cuboide, flattened sphere	brown, cream, grey, white	marbled	none	-
L. densiflorus	NA	Drummond0331A02	not yet deposited, FHO (RJE243)	0.64	-	-	-	flattened oval	browns	marbled with eyebrow	none	-
L. digitatus	OW	Kew Seed database	-	12.03	-	-	-	-	-	-	-	-
L. excubitus	NA	Kew Seed database	-	0.02	-	-	-	-	-	-	-	-
L. excubitus ssp. astromontanus	NA	Kew Seed database	-	0.85	-	-	-	-	-	-	-	-
L. formosus	NA	Kew Seed database	-	0.65	-	-	-	-	-	-	-	-
L. gibertianus	SA	S. T. S. Miotto & M. Pinheiro1835	ICN	0.58	-	-	-	flattened cuboide	olive brown	plain	yes	simple-foveolate
L. guaraniticus	SA	S. T. S. Miotto2111	ICN	0.49	3.22	4.6	1.87	flattened cuboide	browns	marbled	yes	-

L. guaraniticus	SA	S. T. S. Miotto & M. Pinheiro1841	ICN	0.63	-	-	-	flattened cuboide	browns	marbled	yes	simple-foveolate
L. havardii	NA	USDA DLEG910433	-	0.27	-	-	-	flattened cuboide	black, brown, cream, grey, red	marbled	none	-
L. havardii	NA	USDA DLEG930123	-	0.39	-	-	-	flattened cuboide	black, browns, cream, greys	marbled	none	simple-foveolate
L. hirsutissimus	NA	Kew Seed database	-	0.58	-	-	-	-	-	-	-	-
L. hirsutissimus	NA	Kew Seed database	-	0.62	-	-	-	-	-	-	-	-
L. hirsutissimus	NA	S&S Seeds Ltd s.n.	-	0.57	3.23	4	2.83	flattened cuboide, flattened oval	browns	marbled with eyebrow	none	substraite simple-foveolate
L. hirsutissimus	NA	USDA DLEG920120	FHO (RJE239, RJE240)	0.55	-	-	-	flattened cuboide, flattened oval	browns	marbled with eyebrow	none	substraite simple-foveolate
L. hispanicus	OW	Kew Seed database	-	2.33	-	-	-	-	-	-	-	-
L. hispanicus	OW	S. Castroviejo21/06/1996	-	1.83	-	-	-	flattened cuboide	cream, pink, yellows	marbled	none	simple-foveolate
L. hispanicus var. bicolor	OW	Kew Seed database	-	2.28	-	-	-	-	-	-	-	-
L. honoratus	AND	Montiel57	GH	-	-	-	-	-	-	-	-	rugulate
L. lanatus	SA	Anon. 30/12/02-2	ICN	1.56	-	-	-	flattened cuboide	brown, cream, dark green	plain or marbled	yes	-
L. lanatus	SA	Anon. 30/12/02-3	ICN	1.53	-	-	-	flattened cuboide	brown, cream, dark green	plain or marbled	yes	simple-foveolate
L. lanatus	SA	S. T. S. Miotto1688	ICN	1.28	4.76	6.19	2.23	flattened cuboide	browns	plain or marbled	yes - beaked	-
L. lanatus	SA	S. T. S. Miotto1714	ICN	1.53	-	-	-	flattened cuboide	browns	plain or marbled	yes - beaked	-
L. lanatus	SA	S. T. S. Miotto & M. Pinheiro1796	ICN	1.14	-	-	-	flattened cuboide	dark green, brown, cream	plain or marbled	yes	-
L. lanatus	SA	S. T. S. Miotto & M. Pinheiro1806	ICN	1.25	4.48	5.89	2.53	flattened cuboide	browns	plain or marbled	yes - beaked	-
L. lanatus	SA	S. T. S. Miotto & M. Pinheiro1837	ICN	1.02	-	-	-	flattened cuboide	dark green, brown, cream	plain or marbled	yes	-
L. lanatus	SA	S. T. S. Miotto & M. Pinheiro1839	ICN	0.74	3.8	5	2.04	flattened cuboide	browns	plain or marbled	yes - beaked	-
L. latifolius	NA	Kew Seed database	-	0.63	-	-	-	-	-	-	-	-
L. latifolius	NA	Kew Seed database	-	0.65	-	-	-	-	-	-	-	-
L. latifolius ssp. latifolius	NA	Kew Seed database	-	0.34	-	-	-	-	-	-	-	-
L. latifolius ssp. parishii	NA	Kew Seed database	-	0.73	-	-	-	-	-	-	-	-

L. lepidus	NA	USDA PI504439	-	0.12	2.1	2.64	1.42	flattened oval	browns, cream	marbled	none	rugulate
L. lepidus spp. confertus	NA	Kew Seed database	-	0.17	-	-	-	-	-	-	-	-
L. linearis	SA	S. T. S. Miotto & M. Pinheiro1798	ICN	0.23	-	-	-	flattened cuboide	browns, creams, greys	marbled	none	-
L. linearis	SA	S. T. S. Miotto & M. Pinheiro1804	ICN, FHO (RJE236)	0.39	-	-	-	flattened cuboide	browns, creams, greys	marbled	none	-
L. linearis	SA	S. T. S. Miotto & M. Pinheiro1813	ICN	0.2	-	-	-	flattened cuboide	browns, creams, greys	marbled	none	-
L. longifolius	NA	Kew Seed database	-	0.57	-	-	-	-	-	-	-	-
L. luteolus	NA	Kew Seed database	-	0.67	-	-	-	-	-	-	-	-
L. luteolus	NA	Kew Seed database	-	0.71	-	-	-	-	-	-	-	-
L. luteus	OW	J. Lopez and P. D. Guerrero23/03/1991	-	2.63	6.39	7.28	3.12	flattened cuboide	browns, cream	marbled with half moon	none	simple-foveolate
L. luteus	OW	Kew Seed database	-	2.25	-	-	-	-	-	-	-	-
L. magnistipulatus	SA	M. Pinheiro164	ICN	0.69	-	-	-	flattened cuboide	browns, creams, greys	marbled	yes - beaked	-
L. magnistipulatus	SA	S. T. S. Miotto1633	ICN	0.67	-	-	-	flattened cuboide	browns, creams, greys	marbled	yes - beaked	-
L. magnistipulatus	SA	S. T. S. Miotto & M. Pinheiro1840	ICN	0.42	-	-	-	flattened cuboide	browns, creams, greys	marbled	yes - beaked	-
L. magnistipulatus	SA	S. T. S. Miotto & M. Pinheiro1842	ICN	0.39	-	-	-	flattened cuboide	browns, creams, greys	marbled	yes - beaked	-
L. mexicanus	NA	Kew Seed database	-	0.48	-	-	-	-	-	-	-	-
L. micranthus	OW	Falavera29/05/2001	FHO	-	-	-	-	flattened oval, flattened sphere	brown, orange	marbled	none	papillose
L. micranthus	OW	Kew Seed database	-	1.19	-	-	-	-	-	-	-	-
L. microcarpus	NA	C. Drummond12109	not yet deposited	0.6	-	-	-	flattened oval, flattened sphere	cream, sandy	plain	none	rugulate and protuberant
L. microcarpus	NA	Kew Seed database	-	0.6	-	-	-	-	-	-	-	-
L. microcarpus spp. densiflorus	NA	Kew Seed database	-	0.61	-	-	-	-	-	-	-	-
L. microcarpus spp. microcarpus	NA	Kew Seed database	-	0.46	-	-	-	-	-	-	-	-
L. microphyllus	AND	CEH2046	CAS, FHO, K, MOL, NY	0.1	2.14	2.67	1.32	flattened cuboide, flattened oval	black, browns, cream, greys	marbled	none	-
L. microphyllus	AND	CEH2272	FHO, K, LPB	0.15	-	-	-	flattened cuboide, flattened oval	black, browns, cream,	marbled with eyebrow	none	-

L. misticola	AND	CEH2250	FHO (and RJE13 and RJE25), K, LPB	0.17	2.27	3.28	1.39	flattened cuboide, flattened oval	black, browns, cream, greys	marbled with eyebrow	none	-
L. misticola	AND	CEH2350	FHO, K, MOL	0.23	2.23	3.52	1.02	flattened oval	browns	marbled sometimes with eyebrow	none	-
L. misticola	AND	CEH2354	FHO, K, MOL	0.33	2.17	3.67	1.46	flattened oval	black, browns, cream, greys	marbled	none	-
L. misticola	AND	CEH2365	FHO, K, MOL	0.18	2.3	3.35	2.49	flattened cuboide, flattened oval	browns	marbled	none	-
L. montanus	NA	CEH2174	FHO, MEXU	0.39	3.2	4.43	1.74	flattened cuboide	black, brown, white	marbled	none	rugulate
L. multiflorus	SA	S. T. S. Miotto & M. Pinheiro1825	ICN	0.29	-	-	-	flattened cuboide	black, browns, cream, greys	marbled	yes	-
L. multiflorus	SA	S. T. S. Miotto & M. Pinheiro1851	ICN	0.68	-	-	-	-	-	-	-	-
L. multiflorus	SA	S. T. S. Miotto & M. Pinheiro1857	ICN	0.61	-	-	-	flattened cuboide	black, browns, cream, greys	marbled	yes	-
L. mutabilis	AND	CEH1999	CAS, FHO, K, MOL, NY	5.47	7.2	8.95	5.02	cuboide	cream	plain	none	-
L. mutabilis	AND	CEH2009	FHO (and RJE11, RJE18), K, MOL, NY	6.79	7.86	9.26	5.29	cuboide, spherical	black, dark brown or white	plain	none	-
L. mutabilis	AND	CEH2302	FHO, K, LPB	6.75	8.48	10.06	4.98	cuboide, spherical	cream or white	plain	none	-
L. mutabilis	AND	CEH2317	FHO, K, LPB	4.81	7.16	8.86	4.39	cuboide	cream or white	plain	none	-
L. mutabilis	AND	CEH2323	FHO, K, LPB	7.19	8.24	10.48	4.9	cuboide, oval, spherical	cream	plain	none	-
L. mutabilis	AND	CEH2326	FHO (and RJE14), K, LPB	4.99	6.9	9.04	4.59	cuboide, oval, spherical	cream, cream and black or cream and grey	plain or with a moustache	none	-
L. mutabilis	AND	RJE151	FHO, K, MOL, USM	8.79	9.67	11.54	5.38	cuboide, oval	cream	plain	none	-
L. nanus	NA	Anon. s.n.6/6/00	-	0.29	2.3	3.08	2.23	flattened cuboide	brown, cream, grey, white	marbled	none	rugulate
L. nanus	NA	Kew Seed database	-	0.26	-	-	-	-	-	-	-	-
L. odoratus	NA	C. Drummond0311A01	not yet deposited	0.21	-	-	-	flattened cuboide, flattened oval	cream	plain	none	rugulate and protuberant
L. onustus	NA	Kew Seed database	-	0.83	-	-	-	-	-	-	-	-
L. palaestinus	OW	Kew Seed database	-	7.86	-	-	-	-	-	-	-	-
L. paranensis	SA	Anon. ind 2, 18	-	0.42	-	-	-	flattened cuboide	browns	marbled	yes	-

L. peirsonii	NA	Kew Seed database	-	0.7	-	-	-	-	-	-	-	-
L. perennis	NA	Kew Seed database	-	0.52	-	-	-	-	-	-	-	-
L. piurensis	AND	CEH1997	FHO (and RJE16), K, MOL, NY	0.81	-	-	-	flattened cuboide	black, browns, greys	marbled with eyebrow	none	-
L. piurensis	AND	RJE113	FHO, K, MOL, USM	0.77	3.43	4.91	2.74	flattened cuboide	browns, greys	marbled	none	rugulate
L. piurensis var. jamesonianus	AND	RJE196	FHO, K, QCNE	0.71	3.42	5.11	2.7	flattened oval	browns	marbled with eyebrow	none	-
L. plattensis	NA	Kew Seed database	-	1.14	-	-	-	-	-	-	-	-
L. polyphyllus	NA	Kew Seed database	-	0.46	-	-	-	-	-	-	-	-
L. polyphyllus var. cermineus	NA	Kew Seed database	-	0.53	-	-	-	-	-	-	-	-
L. praestabilis	AND	CEH2353	FHO, K, MOL	1.3	4.86	6.41	2.37	flattened cuboide, flattened oval	browns, cream, grey	marbled often with eyebrow	none	-
L. praestabilis	AND	RJE142	FHO, K, MOL, USM	-	-	-	-	flattened cuboide	light browns	marbled	none	rugulate
L. princei	OW	Kew Seed database	-	2.08	-	-	-	-	-	-	-	-
L. prostratus	AND	CEH2239	FHO, K, MOL	0.34	3.22	4.26	1.67	flattened cuboide	olive green, brown	marbled	none	-
L. pubescens	AND	CEH1989	CAS, FHO, K, MOL, NY	1.35	5.3	6.91	2.2	flattened cuboide, flattened oval	browns, cream, grey	marbled often with eyebrow	none	-
L. pulvinaris	AND	CEH2333	FHO, K, LPB	0.11	1.82	2.34	1.39	flattened oval	browns, greys, yellows	marbled	none	-
L. purosericeus	AND	CEH2248	FHO, K, MOL	0.33	2.94	4.07	1.72	flattened cuboide, flattened oval	browns	marbled	none	-
L. pusillus	NA	Kew Seed database	-	0.44	-	-	-	-	-	-	-	-
L. ramossisimus	AND	CEH2037	CAS, F, FHO, K, MO, MOL, NY	0.46	3.07	4.37	2.23	flattened cuboide	browns, cream, black	marbled with eyebrow	none	-
L. ramossisimus	AND	CEH2038	CAS, F, FHO, K, MO, MOL, NY	0.81	3.98	5.32	2.18	flattened cuboide	black, browns, greys	marbled with eyebrow	none	-
L. reitzii	SA	M. Pinheiro310	ICN	0.37	-	-	-	flattened cuboide	browns, cream, grey	marbled	yes - beaked	-
L. reitzii	SA	S. T. S. Miotto & M. Pinheiro1833	ICN	1.05	-	-	-	flattened cuboide	grey	marbled	yes - beaked	-
L. reitzii	SA	S. T. S. Miotto & M. Pinheiro1844	ICN	0.55	-	-	-	flattened cuboide	browns, cream, grey	marbled	yes - beaked	-
L. reitzii	SA	S. T. S. Miotto & M. Pinheiro1846	ICN	0.56	-	-	-	flattened cuboide	browns	marbled	yes - beaked	-
L. rivularis	NA	Kew Seed database	-	0.9	-	-	-	-	-	-	-	-
L. rubriflorus	SA	S. T. S. Miotto2117	ICN	0.17	-	-	-	flattened cuboide	beige, browns, grey	marbled	yes	-

L. rubriflorus	SA	S. T. S. Miotto & M. Pinheiro1838	ICN	0.4	-	-	-	flattened cuboide	beige, browns, grey	marbled	yes	-
L. semperflorens	AND	CEH2012	CAS, F, FHO, K, MO, MOL, NY	0.69	3.31	4.88	2.46	flattened cuboide, flattened oval	browns, creams, greys	marbled with eyebrow	none	-
L. semperflorens	AND	CEH2032	CAS, F, FHO (and RJE15), K. MEXU, MO, MOL, NY	1.55	4.99	6.71	2.91	flattened cuboide, flattened oval	browns, creams	marbled	none	-
L. semperflorens	AND	RJE219	FHO, K, QCNE	1.33	-	-	-	flattened cuboide	brown, grey	marbled with eyebrow	none	-
L. semperflorens	AND	RJE220	FHO, K, QCNE	0.49	3.44	4.67	2.38	flattened oval	brown, grey	marbled with eyebrow	none	rugulate
L. smithanus	AND	RJE201	FHO, K, QCNE	-	-	-	-	cuboide, flattened cuboide	purple	-	none	rugulate
L. spW	AND	Muller827	OXF	-	-	-	-	-	-	-	-	simple-foveolate
L. spX	AND	J. I. Wood s.n.	FHO	-	-	-	-	flattened cuboide, flattened oval	dark brown	-	yes- small beak	simple-foveolate
L. spA	NA	CEH1664	E, FHO, K, MEXU, MO, NY	-	-	-	-	flattened cuboide, flattened oval	black, brown, cream	marbled with eyebrow	none	ocellate
L. spA	NA	CEH2099	FHO, MEXU	0.35	-	-	-	-	-	-	-	ocellate
L. spE	AND	CEH2218	FHO, K, MOL	0.26	2.59	3.68	1.8	flattened cuboide, flattened oval	greys, browns, black, cream	marbled with eyebrow	none	-
L. spF	AND	CEH2247	FHO, K, MOL	0.26	-	-	-	-	-	-	-	-
L. spH	AND	CEH2256	FHO, K, LPB	0.14	-	-	-	flattened cuboide, flattened oval	brown, black	marbled	none	-
L. spI	AND	CEH2261	FHO, K, LPB	0.11	2	2.58	1.18	flattened cuboide, flattened oval	black, brown, cream	marbled with eyebrow	none	-
L. spJ	AND	CEH2274	FHO, K, LPB	0.28	-	-	-	-	-	-	-	-
L. spK	AND	CEH2276	FHO, K, LPB	0.19	-	-	-	-	-	-	-	-
L. sp16	AND	CEH2294	FHO (and RJE24), K, LPB	0.15	2.26	2.59	1.11	flattened cuboide, flattened oval	black, browns, cream, grey	marbled with eyebrow	none	-
L. spO	AND	CEH2352	FHO, K, MOL	2.17	2.18	3.52	1.69	flattened cuboide, flattened oval	brown, cream, grey, pink	marbled often with eyebrow	none	-
L. sp1	AND	CEH2002	CAS, F, FHO, K. MEXU, MO, MOL, NY	0.73	3.9	5.33	2.26	flattened cuboide, flattened oval	browns, creams, greys	marbled with eyebrow	none	-
L. spP	AND	CEH2386	FHO, K, MOL	2.17	-	-	-	flattened cuboide, flattened oval	white, brown	marbled often with eyebrow	none	-
L. spT	AND	RJE141	FHO, K, MOL, USM	-	-	-	-	flattened cuboide	brown, cream, greys	marbled with eyebrow	none	rugulate
L. sp2	AND	CEH2004	CAS, F, FHO, K, MO, MOL, NY	0.58	3.52	4.68	2.05	flattened cuboide, flattened oval	black, browns	marbled	none	-

L. spC	AND	CEH2045	FHO, K, MOL, NY	0.33	-	-	-	-	-	-	-	-
L. spG	AND	CEH2255	FHO, K, LPB	0.33	-	-	-	-	-	-	-	-
L. spD	NA	CEH2106	FHO, MEXU	0.44	-	-	-	-	-	-	-	rugulate
L. sp6	NA	CEH2160	FHO, MEXU	0.55	-	-	-	-	-	-	-	rugulate
L. sparsiflorus	NA	Kew Seed database	-	0.13	-	-	-	-	-	-	-	-
L. stipulatus	NA	Kew Seed database	-	0.37	-	-	-	-	-	-	-	-
L. stiversii	NA	Kew Seed database	-	0.09	-	-	-	-	-	-	-	-
L. subacaulis	AND	CEH2319	FHO, K, LPB	0.16	-	-	-	flattened cuboide, flattened oval	black, browns, cream, greys	marbled	none	rugulate
L. subacaulis	AND	CEH2324	FHO, K, LPB	0.08	1.75	2.96	1.36	flattened cuboide, flattened oval	browns, cream, grey, oranges,	marbled often with eyebrow	none	-
L. succulentus	NA	Kew Seed database	-	0.63	-	-	-	-	-	-	-	-
L. tarapacensis	AND	Gardener6296	E	-	-	-	-	-	-	-	-	rugulate
L. texensis	NA	Kew Seed database	-	0.81	-	-	-	-	-	-	-	-
L. texensis	NA	Kew Seed database	-	0.83	-	-	-	-	-	-	-	-
L. texensis	NA	S&S Seeds Ltd s.n.	-	0.83	4.31	4.33	2.35	flattened cuboide, flattened sphere	black, brown, grey, cream	marbled	none	-
L. texensis	NA	USDA DLEG900334	-	0.9	-	-	-	flattened cuboide	black, brown, grey, cream	marbled	none	-
L. texensis	NA	USDA PI577291	-	0.82	4.01	3.9	2.64	flattened cuboide, flattened sphere	black, brown, grey, cream	marbled	none	simple-foveolate
L. tidestromii	NA	Kew Seed database	-	0.94	-	-	-	-	-	-	-	-
L. tomentosus	AND	CEH2325	FHO, K, LPB	0.46	3.13	4.42	2.02	flattened cuboide, flattened oval	browns, creams, greys	marbled with eyebrow	none	-
L. tominensis	AND	CEH2296	FHO, K, LPB	0.21	2.47	3.42	1.63	flattened cuboide, flattened oval	black, brown, cream	marbled sometimes with moustache	none	-
L. tominensis	AND	CEH2332	FHO, K, LPB	0.28	2.15	3	1.44	flattened cuboide, flattened oval	black, brown, greys	marbled	none	-
L. truncatus	NA	S&S Seeds Ltd s.n.	FHO (RJE237, RJE238, RJE242)	0.28	2.97	3.61	2.03	flattened cuboide, flattened oval	brown, cream	marbled	none	substraite simple-foveolate
L. uleanus	SA	S. T. S. Miotto & M. Pinheiro1848	ICN	0.38	-	-	-	flattened cuboide	brown, cream, greys	marbled	none	-
L. varicolor	NA	Kew Seed database	-	0.33	-	-	-	-	-	-	-	-
L. varius	NA	Kew Seed database	-	7.19	-	-	-	-	-	-	-	-
L. verbasciformis	AND	Cuatrecasas28080	F, US	-	-	-	-	-	-	-	-	rugulate
L. villosus	NA	CEH1986	FHO, K	0.36	3.22	3.6	1.53	flattened cuboide, flattened oval	brown, cream, grey, red	marbled	none	simple-foveolate
L. villosus	NA	CEH1988	FHO, K	0.46	3.49	3.99	2.26	flattened cuboide, flattened oval	browns, greys	marbled	none	simple-foveolate
L. weberbaueri	AND	CEH2233	FHO, K, MOL	0.25	2.89	4.01	1.54	flattened cuboide	black, brown, cream	marbled with eyebrow	none	rugulate

* Regions

AND=Andes, NA= North America, OW=Old World, SA=Non-Andean South America

Appendix 14 Observations of articulation and pulvinule/pulvinules in non-digitate species

Species	Specimen	Sessile	Evidence of pulvinus, articulation or abscission at petiole base	Evidence of pulvinule, articulation or abscission at blade base	Notes
L. arenarius	Gardner 4500	No	Leaf scars.	Blade at an angle to the petiole.	
L. arenarius	Clussen 1839	No	Leaf scars.	Blade at an angle to the petiole.	
L. arenarius	Clussen 1840	No	Leaf scars.	Blade at an angle to the petiole.	Estipulate.
L. crotalarioides	CEH2482	No	Leaf scars.	Persistent petiole with blade abscised. Blades often at an angle to the petiole.	Stipules fused to petiole and stem.
L. crotalarioides	CEH2483	No	None.	None.	
L. crotalarioides	CEH2481	No	None.	Persistent petiole with blade abscised.	
L. cumulicola	CEH1988	No	None.	Blades often at an angle to the petiole. Slight ridge across petiole apex.	
L. cumulicola	CEH1984	No	Leaf scars.	Blades often at an angle to the petiole.	
L. cumulicola	CEH1983	No	Leaf scars.	Persistent petioles with blade abscised. Blades often at an angle to the petiole.	
L. guaraniticus	Krapovickas 36808	No	Leaf scars.	Persistent petiole with blade abscised.	
L. guaraniticus	Quarin 3402	No	None.	None.	
L. guaraniticus	CEH2488	No	Leaf scars.	None.	
L. guaraniticus	Kenicer s.n.	No	Leaf scars.	None.	Young blades at right angles to stem. Mature blades held against stem. Estipulate.
L. guaraniticus	Keller 92	No	None.	Possible fractures between petiole apex and blade.	
L. paraguariensis	CEH2493	No	None.	Persistent petioles with blade abscised.	Non-digitate and digitate leaves
L. paraguariensis	CEH2495	No	None.	Persistent petioles with blade abscised.	Non-digitate and digitate leaves
L. paraguariensis	CEH2497	No	None.	Persistent petioles with blade abscised. Blades of non-digitate leaf often at an angle to petiole.	Non-digitate and digitate leaves
L. parvifolius	CEH2480	Yes	Leaf scars.	None.	
L. sp.	Rieddel	No	Leaf scars.	Blade at an angle to the petiole.	
L. subsessilis	CEH2479	Nearly	Leaf scars.	Persistent petioles with blade abscised. Blades often at an angle to the petiole.	
L. velutinus	CEH2478	No	None.	Blade at an angle to the petiole.	
L. velutinus	CEH2444	No	None.	None.	
L. villosus	CEH1986	No	Leaf scars.	Persistent petioles with blade abscised. Blade at an angle to the petiole. Petiole swollen at apex.	
L. villosus	CEH1987	No	Leaf scars.	Blades often at an angle to the petiole. Petiole swollen at apex.	
L. villosus	Nutall 1821	No	None.	Blades often at an angle to the petiole. Petiole swollen at apex.	

Appendix 15 *Lupinus* seed coat surface pattern character states and representative species

Seed Coat Type	Species	Figure	Distribution and Infrageneric rank (OW=Old World SA=South America MEX=Mexican NA=North America)	Literature Reference†
A. Granular	*L. albus*	5.12a	OW Sect. Albi	HH, PH, F
B. Papillose	*L. angustifolius*	Not shown	OW Sect. Angustifoli	HH, F, SM
	L. micranthus	5.12b	OW Sect. Micranthi	HH, AB
C. Fascicular	*L. anatolicus*	-	OW Sect. Scabrispermae	AB
	L. pilosous	-	OW Sect. Scabrispermae	HH, AB
	L. palaestinus	-	OW Sect. Scabrispermae	HH
	L. cosentinii	5.12c	OW Sect. Scabrispermae	HH
	L. atlanticus	-	OW Sect. Scabrispermae	PH
	L. digitatus	-	OW Sect. Scabrispermae	PH
D. Simple-foveolate	*L. angustifolius*	-	OW Sect. Angustifoli	AB
	L. luteus	5.13a	OW Sect. Lutei	HH
	L. hispanicus	5.13c	OW Sect. Lutei	PH
	L. cumulicola	5.13e	Sect. Boreo-Unifolius	
	L. villosus	-	Sect. Boreo-Unifolius	M
	L. havardii	5.13d	Sect. Orientalis Ser. Boreo-Orientalis	B
	L. subcarnosus	-	Sect. Orientalis Ser. Boreo-Orientalis	B, PH
	L. texensis	5.13b	Sect. Orientalis Ser. Boreo-Orientalis	B
	L. albescens	-	Sect. Orientalis Ser. Austro-Unifolius	F
	L. bandelierae	5.12d	Sect. Orientalis Ser. Austro-Orientalis	
	L. bracteolaris	Not shown	Sect. Orientalis Ser. Austro-Orientalis	M
	L. gibertianus	Not shown	Sect. Orientalis Ser. Austro-Orientalis	F
	L. sp (Wood s.n.)	Not shown	Sect. Orientalis Ser. Austro-Orientalis	
	L. arenarius	-	Sect. Orientalis Ser. Austro-Unifolius	M
	L. comptus	-	Sect. Orientalis Ser. Austro-Unifolius	M
	L. crotalarioides	-	Sect. Orientalis Ser. Austro-Unifolius	M
	L. decurrans	-	Sect. Orientalis Ser. Austro-Unifolius	M
	L. guaranticus	5.13f	Sect. Orientalis Ser. Austro-Unifolius	M
	L. lanatus	Not shown	Sect. Orientalis Ser. Austro-Unifolius	M
	L. linearis	Not shown	Sect. Orientalis Ser. Austro-Unifolius	
	L. parvifolius	-	Sect. Orientalis Ser. Austro-Unifolius	M
	L. sellowianus	-	Sect. Orientalis Ser. Austro-Unifolius	M
	L. velutinus	-	Sect. Orientalis Ser. Austro-Unifolius	M

E. Rugulate	*L. albicaulis*	-	NA Sect. Occidentalis Ser. Occidentalis	HH
	L. albifrons	-	NA Sect. Occidentalis Ser. Occidentalis	M
	L. alveorum	-	NA Sect. Occidentalis Ser. Occidentalis	M
	L. arboreus	-	NA Sect. Occidentalis Ser. Occidentalis	HH
	L. arizonicus	5.13g	NA Sect. Occidentalis Ser. Occidentalis	
	L. bicolour	Not shown	NA Sect. Occidentalis Ser. Occidentalis	PH
	L. brevicaulis	Not shown	NA Sect. Occidentalis Ser. Occidentalis	
	L. covillei	-	NA Sect. Occidentalis Ser. Occidentalis	M
	L. elongatus	-	NA Sect. Occidentalis Ser. Occidentalis	M
	L. excubitus	-	NA Sect. Occidentalis Ser. Occidentalis	M
	L. exochus	-	NA Sect. Occidentalis Ser. Occidentalis	M
	L. hartwegii	-	NA Sect. Occidentalis Ser. Occidentalis	M
	L. lepidus	Not shown	NA Sect. Occidentalis Ser. Occidentalis	SM
	L. nanus	Not shown	NA Sect. Occidentalis Ser. Occidentalis	HH
	L. nootkatensis	-	NA Sect. Occidentalis Ser. Occidentalis	M
	L. subcarnosus	-	NA Sect. Occidentalis Ser. Occidentalis	M
	L. succulentus	-	NA Sect. Occidentalis Ser. Occidentalis	HH
	L. aschenbornii	-	MEX Sect. Occidentalis Ser.	HH
	L. elegans	-	Occidentalis	M
	L. montanus	Not shown	MEX Sect. Occidentalis Ser.	
	L. sp	Not shown	Occidentalis	
	(CEH2106)	Not shown	MEX Sect. Occidentalis Ser.	F
	L. sp	Not shown	Occidentalis	M
	(CEH2160)	Not shown	MEX Sect. Occidentalis Ser.	
	L. alopecuroides	-	Occidentalis	HH, PH, F
	L. bangii	-	MEX Sect. Occidentalis Ser.	M
	L. honoratus	-	Occidentalis	M
	L. humifusus	5.12e	SA Sect. Occidentalis Ser. Occidentalis	
	L. magdalensis	-	SA Sect. Occidentalis Ser. Occidentalis	M
	L. mutabilis	5.13h	SA Sect. Occidentalis Ser. Occidentalis	
	L. paniculatus	-	SA Sect. Occidentalis Ser. Occidentalis	
	L. piurensis	-	SA Sect. Occidentalis Ser. Occidentalis	M
	L. praestabilis	-	SA Sect. Occidentalis Ser. Occidentalis	M
	L. prostratus	-	SA Sect. Occidentalis Ser. Occidentalis	M
	L. pubescens	Not shown	SA Sect. Occidentalis Ser. Occidentalis	M
	L. ramosissimus	Not shown	SA Sect. Occidentalis Ser. Occidentalis	
	L. sarmentosus	Not shown	SA Sect. Occidentalis Ser. Occidentalis	M
	L. semperflorens	Not shown	SA Sect. Occidentalis Ser. Occidentalis	
	L. smithianus	5.14h	SA Sect. Occidentalis Ser. Occidentalis	
	L. tarapacensis	Not shown	SA Sect. Occidentalis Ser. Occidentalis	
	L. verbasciformis	Not shown	SA Sect. Occidentalis Ser. Occidentalis	
	L. weberbaueri		SA Sect. Occidentalis Ser. Occidentalis	
	L. lindleyanus		SA Sect. Occidentalis Ser. Occidentalis	
			SA Sect. Occidentalis Ser. Occidentalis	
			SA Sect. Occidentalis Ser. Occidentalis	
F. Rugulate & Protuberant	*L. microcarpus*	5.14a	Sect. Occidentalis Ser.Platycarpos	
	L. odoratus	5.12f, 5.14b	Sect. Occidentalis Ser.Platycarpos	
G. Substriate simple-foveolate	*L. hirsutissimus*	5.14c	NA Sect. Occidentalis Ser. Occidentalis	
	L. truncatus	5.12g, 5.14d	NA Sect. Occidentalis Ser. Occidentalis	
H. Ocellate	*L. sp.* (CEH1644)	5.14f	NA Sect. Occidentalis Ser. Occidentalis	
	L. sp. (CEH2099)	5.12h, 5.14e	NA Sect. Occidentalis Ser. Occidentalis	

AB= Ainouche, A., R. J. Bayer, and M.-T. Misset. (2004). Molecular phylogeny, diversification and character evolution in *Lupinus* (Fabaceae) with special attention to Mediterranean and African lupines. *Plant Systematics and Evolution* **246,** 211-222.

B= Bragg, L. H. (1983). Seed coats of some *Lupinus* species. *Scanning Electron Microscopy* **IV**, 1739-1745.

F= Fuentes, E., P. Perisse, and A. M. Planchuelo. (1993). Seed morphology and lipids analysis of cultivated and wild species of *Lupinus*. *7th International Lupin Conference*, Edora, Portugal. 185-190.

HH= Heyn, C. C., and L. Herrnstadt. (1977). Seed coat structure of Old World *Lupinus* species. *Botanica Nordic* 130: 427-435.

M= Monteiro, R. (1987) Seed testa pattern of unifoliolate species of *Lupinus* L. (Leguminosae). *Salusvita*, **6**, 20-31.

PH= Plitmann, U., and C. C. Heyn. (1984). Old World *Lupinus*: taxonomy, evolutionary relationships and links with New World species. *3rd International Lupin Conference*. 55-66.

SM= Saint-Martin, M. (1986). Micromorphologie tégumentaire des graines de Papilionaceae. *Bulletin de la societé Botanique de France, lettres Botanique* **133**. 137-153.

Appendix 16 Chromosome numbers reported in *Lupinus*

Species	Infrageneric classification	Collector Name and Number	2n	Reference (Primary reference where known)
L. albicaulis	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. albicoccineus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Tuschnjakowa, 1935)
L. albus	Section Albi		30	Darlington and Wylie (1955) (Olsezewska, 1954)
L. albus	Section Albi		30	Gladstones (1980)
L. albus	Section Albi		40	Fedorov (1969) (Smet, 1914)
L. albus	Section Albi		40	Darlington and Wylie (1955)
L. albus	Section Albi		40	Gladstones (1980)
L. albus	Section Albi		48	Fedorov (1969) (Tuschnjakowa, 1935)
L. albus	Section Albi		48	Fedorov (1969)
L. albus	Section Albi		48	Fedorov (1969)
L. albus	Section Albi		48	Gladstones (1980)
L. albus	Section Albi		50	Fedorov (1969) (Tuschnjakowa, 1935)
L. albus	Section Albi		50	Fedorov (1969) (Olsezewska ,1954)
L. albus	Section Albi		50	Fedorov (1969) (Gilot ,1965)
L. albus	Section Albi		50	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. albus	Section Albi		50	Fedorov (1969) (Atabekova, 1959)
L. albus	Section Albi		50	Fedorov (1969) (Senn, 1938)
L. albus	Section Albi		50	Maisurian and Atabekova (1974)
L. albus	Section Albi		50	Darlington and Wylie (1955) (Malheiros, 1942)
L. albus	Section Albi		50	Gladstones (1998)
L. albus	Section Albi		50	Kazmierski (1988)
L. albus	Section Albi		100	Kazimierski (1988)
L. anaeanus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Diers, 1961)
L. angustifolius	Section Angustifoli		40	Fedorov (1969) (Gilot, 1965)
L. angustifolius	Section Angustifoli		40	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. angustifolius	Section Angustifoli		40	Fedorov (1969) (Atabekova, 1959)
L. angustifolius	Section Angustifoli		40	Fedorov (1969) (Tuschnjakowa, 1935)
L. angustifolius	Section Angustifoli		40	Fedorov (1969)
L. angustifolius	Section Angustifoli		40	Maisurian and Atabekova (1974) (Winge, 1925)
L. angustifolius	Section Angustifoli		40	Darlington and Wylie (1955) (Malheiros, 1942)
L. angustifolius	Section Angustifoli		40	Gladstones (1998)
L. angustifolius	Section Angustifoli		40	Kazmierski (1988)
L. angustifolius	Section Angustifoli		48	Darlington and Wylie (1955) (Kawakami, 1930)
L. angustifolius	Section Angustifoli		48	Gladstones (1980)
L. arboreus	Section Occidentalis Seris Occidentalis		40	Fedorov (1969)
L. arboreus	Section Occidentalis Seris Occidentalis		40	Darlington and Wylie (1955)
L. arboreus	Section Occidentalis Seris Occidentalis		48	Maisurian and Atabekova (1974)
L. arboreus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964)
L. arboreus	Section Occidentalis Seris Occidentalis		48	Dunn and Gillett (1966)
L. arboreus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. arboreus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gilot, 1965)
L. arboreus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gustafasson and Gadd, 1969)
L. arbustus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. arbustus	Section Occidentalis Seris Occidentalis		96	Phillips (1957)

L. arcticus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Holmen, 1962)
L. arcticus	Section Occidentalis Seris Occidentalis		48	Dunn and Gillett (1966)
L. argenteus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Ledingham, 1957)
L. argenteus	Section Occidentalis Seris Occidentalis		48	Index to Plant Chromosome numbers (IPCN)
L. argenteus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. argenteus subsp. argenteus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. argenteus subsp. parviflorus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. aridus	Section Occidentalis Seris Occidentalis		24	Dunn (1984)
L. aridus	Section Occidentalis Seris Occidentalis		24	Reveal (1973)
L. aridus	Section Occidentalis Seris Occidentalis		48	Cox (1972)
L. aridus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gilot, 1965)
L. articus	Section Occidentalis Seris Occidentalis		48	IPCN
L. articus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. arvensis	Section Occidentalis Seris Occidentalis	RJE 221	48	Schifino-Wittmann unpublished.
L. atlanticus	Section Scabrispermae		38	Gladstones (1998)
L. atlanticus	Section Scabrispermae		38	Kazmierski (1988)
L. ballianus	Section Occidentalis Seris Occidentalis	CEH 1989, 2038	48	Conterato and Schifino-Wittmann (2006)
L. bandelierae	Section Orientalis Series Austro-orientalis	CEH 2301, 2318, 2321	36	Conterato and Schifino-Wittmann (2006)
L. barceri	Section Occidentalis Seris Occidentalis		48	Maisurian and Atabekova (1974)
L. barkei	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. barkei	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. barkei	Section Occidentalis Seris Occidentalis		50	Fedorov (1969)
L. barkei	Section Occidentalis Seris Occidentalis		50	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. bicolor subsp microphyllus	Section Occidentalis Seris Occidentalis		48	Dunn and Gillett (1966)
L. bicolor subsp umbellatus	Section Occidentalis Seris Occidentalis		48	Dunn and Gillett (1966)
L. bracteolaris	Section Orientalis Series Austro-orientalis	S.T.S. Miotto 1637, S.T.S. Miotto & Pinheiro 1797, 1819	32	Macieal and Schifino-Wittmann (2002)
L. bracteolaris	Section Orientalis Series Austro-orientalis	UCN 128464	34	Conterato and Schifino-Wittmann (2006)
L. bracteolaris	Section Orientalis Series Austro-orientalis	S.T.S. Miotto 1684, S.T.S. Miotto & Pinheiro 1810, I. Conterato s.n. (ICN 130222)	34	Macieal and Schifino-Wittmann (2002)
L. caballoanus	Section Occidentalis Seris Occidentalis		48	IPCN
L. caespitosus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Heiser, 1963)
L. caudatus subsp. argophyllus	Section Occidentalis Seris Occidentalis		96	Phillips (1957)
L. caudatus subsp. caudatus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. chlorolepis	Section Occidentalis Seris Occidentalis	CEH 2261	48	Schifino-Wittmann unpublished.
L. chrysanthus	Section Occidentalis Seris Occidentalis	CEH 2264	48	Schifino-Wittmann unpublished.
L. cosentinii	Section Scabrispermae		32	Darlington and Wylie (1955) (Malheiros, 1942)

L. cosentinii	Section Scabrispermae		32	Gladstones (1998)
L. cosentinii	Section Scabrispermae		32	Kazmierski (1988)
L. crotalaroides	Section Orientalis Series Austro Unifolius	CEH 2482	36	Conterato and Schifino-Wittmann (2006)
L. cumulicola	Section Boreo-unifolius	CEH 1984, 1985	52	Conterato and Schifino-Wittmann (2006)
L. cytisoides	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gilot, 1965)
L. densiflorus	Section Occidentalis Seris Platycarpos		48	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. densiflorus	Section Occidentalis Seris Platycarpos		48	Dunn and Gillett (1966)
L. densiflorus	Section Occidentalis Seris Platycarpos		48	Fedorov (1969) (Atabekova, 1959)
L. digitatus	Section Scabrispermae		36	Gladstones (1998)
L. digitatus	Section Scabrispermae		36	Kazmierski (1988)
L. digitatus	Section Scabrispermae		40	Fedorov (1969) (Atabekova, 1959)
L. digitatus	Section Scabrispermae		40	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. digitatus	Section Scabrispermae		40	Gladstones (1980)
L. digitatus	Section Scabrispermae		42	Fedorov (1969) (Hackbarth and Troll, 1959)
L. digitatus	Section Scabrispermae		42	Gladstones (1980)
L. digitatus	Section Scabrispermae		52	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. digitatus	Section Scabrispermae		52	Gladstones (1980)
L. 'dillonii'	Section Occidentalis Seris Occidentalis	CEH 2002	48	Conterato and Schifino-Wittmann (2006)
L. douglasii	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. douglasii	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. elegans	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955)
L. elegans	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Tuschujakowa, 1935)
L. elegans	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Senn, 1938)
L. elegans	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. elegans	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gilot, 1965)
L. elegans	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. gibertianus	Section Orientalis Series Austro-orientalis	S.T.S. Miotto & Pinheiro 1835	36	Macieal and Schifino-Wittmann (2002)
L. guaranticus	Section Orientalis Series Austro Unifolius	S.T.S. Miotto & M.T.P. Santos 1841, I. Conterato s.n. (ICN 129985), I. Conterato s.n. (ICN 130223), S.T.S. Miotto 2107, 2109, 2111, 2125	36	Conterato and Schifino-Wittmann (2006)
L. hartwegii	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Giloy, 1965)
L. hartwegii	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. hartwegii	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. hartwegii	Section Occidentalis Seris Occidentalis		48-50	Fedorov (1969) (Senn, 1938)
L. hartwegii	Section Occidentalis Seris Occidentalis		48-50	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. hilarianus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gilot, 1965)
L. hirsutus	Section Scabrispermae		24	Darlington and Wylie (1955)
L. hirsutus* v. *micranthus	Section Scabrispermae		50?	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. hispanicus	Section Lutei		52	Gladstones (1998)
L. hispanicus	Section Lutei		52	Kazmierski (1988)

L. huaronensis	Section Occidentalis Seris Occidentalis	CEH 2241	48	Conterato and Schifino-Wittmann (2006)
L. humicola	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. humicola	Section Occidentalis Seris Occidentalis		96	Phillips (1957)
L. hybridus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Tuschujakowa, 1935)
L. lanatus	Section Orientalis Series Unifolius	S.T.S. Miotto 2110, 2119, Valls *et al.* 9525	36	Conterato and Schifino-Wittmann (2006)
L. lanatus	Section Orientalis Series Austro Unifolius	S.T.S. Miotto 1688, 1704, 1714, S.T.S. Miotto & M. Pinheiro 1796, 1806, 1837, 1839.	36	Macieal and Schifino-Wittmann (2002)
L. laxiflorus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. laxiflorus var laxiflorus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. laxiflorus var laxiflorus	Section Occidentalis Seris Occidentalis		96	Phillips (1957)
L. laxiflorus var pseudoparviflorus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. lepidus subsp. lepidus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. lepidus subsp. lyallii	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. leptophyllus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Turner and Fearing, 1960)
L. leucophyllus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. leucophyllus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. leucophyllus	Section Occidentalis Seris Occidentalis		96	Phillips (1957)
L. lindlianus	Section Occidentalis Seris Occidentalis	RJE 141	48	Conterato and Schifino-Wittmann (2006)
L. linearis	Section Orientalis Series Austro-orientalis	S.T.S. Miotto & Pinheiro 1798, 1804	32	Macieal and Schifino-Wittmann (2002)
L. linearis	Section Orientalis Series Austro-orientalis	S.T.S. Miotto & Pinheiro 1813	34	Macieal and Schifino-Wittmann (2002)
L. linifolius	Section Angustifoli		40	Maisurian and Atabekova (1974) (Atabekova, 1959)
L. littoralis	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. luteus	Section Lutei		46	Darlington and Wylie (1955)
L. luteus	Section Lutei		46	Gladstones (1980)
L. luteus	Section Lutei		48	Darlington and Wylie (1955) (Kawakami, 1930)
L. luteus	Section Lutei		48	Gladstones (1980)
L. luteus	Section Lutei		50	Gladstones (1980)
L. luteus	Section Lutei		50	Fedorov (1969) (Atabekova, 1959)
L. luteus	Section Lutei		52	Maisurian and Atabekova (1974) (Atabekova, 1959)
L. luteus	Section Lutei		52	Darlington and Wylie (1955) (Malheiros, 1942)
L. luteus	Section Lutei		52	Fedorov (1969) (Tuschujakowa, 1935)
L. luteus	Section Lutei		52	Fedorov (1969) (Savchenko, 1935)
L. luteus	Section Lutei		52	Fedorov (1969) (Gilot, 1965)
L. luteus	Section Lutei		52	Fedorov (1969) (Gustafasson and Gadd, 1965)

L. luteus	Section Lutei		52	Fedorov (1969) (Troll, 1963)
L. luteus	Section Lutei		52	Fedorov (1969) (Atabekova, 1959)
L. luteus	Section Lutei		52	Fedorov (1969) (Olsezewska, 1954)
L. luteus	Section Lutei		52	Gladstones (1998)
L. luteus	Section Lutei		52	Kazmierski (1988)
L. luteus	Section Lutei		104	Fedorov (1969) (Troll, 1963)
L. luteus	Section Lutei		104	Kazimierski (1988)
L. lyalli	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. magnistipulatus	Section Orientalis Series Austro Unifolius	S.T.S. Miotto 1633, S.T.S. Miotto & Pinheiro 1840, 1842, M. Pinheiro 122, 164	36	Macieal and Schifino-Wittmann (2002)
L. mexicanus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Beaman, 1962)
L. micranthus	Section Micranthi		48	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. micranthus	Section Micranthi		48	Gladstones (1980)
L. micranthus	Section Micranthi		50	Gladstones (1980)
L. micranthus	Section Micranthi		52	Gladstones (1998)
L. micranthus	Section Micranthi		52	Kazmierski (1988)
L. microphyllus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Diers, 1961)
L. microphyllus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Heiser, 1963)
L. microphyllus	Section Occidentalis Seris Occidentalis	CEH 2272	48	Conterato and Schifino-Wittmann (2006)
L. misticola	Section Occidentalis Seris Occidentalis	CEH 2250	48	Schifino-Wittmann unpubl.
L. montanus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Beaman, 1962)
L. multiflorus	Section Orientalis Series Austro-orientalis	S.T.S. Miotto 2122	36	Conterato and Schifino-Wittmann (2006)
L. multiflorus	Section Orientalis Series Austro-orientalis	S.T.S. Miotto & Pinheiro 1825, 1851, 1857	36	Macieal and Schifino-Wittmann (2002)
L. mutabilis	Section Occidentalis Seris Occidentalis		42	Gladstones (1980)
L. mutabilis	Section Occidentalis Seris Occidentalis		42	Fedorov (1969)
L. mutabilis	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Heiser, 1963)
L. mutabilis	Section Occidentalis Seris Occidentalis	CEH 1999, 2009, 2302, 2323, 2326	48	Conterato and Schifino-Wittmann (2006)
L. mutabilis	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955) (Malheiros, 1942)
L. mutabilis	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Winge, 1925)
L. mutabilis	Section Occidentalis Seris Occidentalis		48	Gladstones (1980)
L. mutabilis	Section Occidentalis Seris Occidentalis		48	Fedorov (1969)
L. mutabilis	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova,(1959)
L. mutabilis	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. mutabilis	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Tuschujakowa (1935))
L. nanus	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. nanus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. nanus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gilot, 1965)
L. nanus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. nanus* v. *latifolius	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Dunn, 1956)
L. nanus* v. *nanus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Dunn, 1956)
L. nootkatensis	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Bragdo, 1957)

L. nootkatensis	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955) (Maude, 940)
L. nootkatensis	Section Occidentalis Seris Occidentalis		48	Dunn and Gillett (1966)
L. nootkatensis	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. nubigenus	Section Occidentalis Seris Occidentalis	CEH 2239	48	Conterato and Schifino-Wittmann (2006)
L. oreganus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. ornatus	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. ornatus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. palaestinus	Section Scabrispermae		42	Gladstones (1998)
L. palaestinus	Section Scabrispermae		42	Kazmierski (1988)
L. paniculatus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Heiser, 1963)
L. paniculatus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Huynh, 1965)
L. paraguariensis	Section Orientalis Series Austro Unifolius	Valls *e.t al* 002861	36	Conterato and Schifino-Wittmann (2006)
L. paranensis	Section Orientalis Series Austro-orientalis	ICN 129986	36	Conterato and Schifino-Wittmann (2006)
L. perennis	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. perennis	Section Occidentalis Seris Occidentalis		96	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. perennis subsp. perennis	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. perennis subsp. plattensis	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. perennis subsp. latifolius	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. perennis subsp. latifolius	Section Occidentalis Seris Occidentalis		96	Phillips (1957)
L. pilosus	Section Scabrispermae		40	Darlington and Wylie (1955)
L. pilosus	Section Scabrispermae		42	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. pilosus	Section Scabrispermae		42	Gladstones (1998)
L. pilosus	Section Scabrispermae		42	Kazmierski (1988)
L. pilosus	Section Scabrispermae		50	Fedorov (1969)
L. pilosus	Section Scabrispermae		50	Fedorov (1969) (Atabekova, 1959)
L. pilosus	Section Scabrispermae		50	Gladstones (1980)
L. piurensis	Section Occidentalis Seris Occidentalis	CEH 1997, RJE113	48	Conterato and Schifino-Wittmann (2006)
L. platamodes	Section Occidentalis Seris Occidentalis		48	IPCN
L. polycarpus	Section Occidentalis Seris Occidentalis		48	Dunn and Gillett (1966)
L. polyphyllus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Bragdo, 1957)
L. polyphyllus	Section Occidentalis Seris Occidentalis		48	Maisurian and Atabekova (1974) (Tuschujakowa, 1935)
L. polyphyllus	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955) (Cooper (1936))
L. polyphyllus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969)
L. polyphyllus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. polyphyllus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Gustafasson and Gadd, 1965)
L. polyphyllus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Berger, 1958)
L. polyphyllus	Section Occidentalis Seris Occidentalis		48	IPCN
L. polyphyllus var. polyphyllus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)

L. polyphyllus var. prunophilus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. praestablis	Section Occidentalis Seris Occidentalis	RJE 142	48	Conterato and Schifino-Wittmann (2006)
L. princei	Section Scabrispermae		38	Gladstones (1998)
L. prostratus	Section Occidentalis Seris Occidentalis	CEH 2001	48	Conterato and Schifino-Wittmann (2006)
L. prostratus	Section Occidentalis Seris Occidentalis	CEH 2000	48	Schifino-Wittmann unpubl.
L. pubescens	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Heiser, 1963)
L. pubescens	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955)
L. pubescens	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. pubescens	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. pubescens	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Senn, 1938)
L. pubescens	Section Occidentalis Seris Occidentalis		50	Fedorov (1969) (Senn, 1938)
L. pulvinaris	Section Occidentalis Seris Occidentalis	CEH 2333	48	Schifino-Wittmann unpubl.
L. purosericeus	Section Occidentalis Seris Occidentalis	CEH 2248	48	Conterato and Schifino-Wittmann (2006)
L. pusillus	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Ledingham, 1957)
L. pusillus	Section Occidentalis Seris Occidentalis		48	Dunn and Gillett (1966)
L. pycnostachys	Section Occidentalis Seris Occidentalis	CEH 2444	48	Conterato and Schifino-Wittmann (2006)
L. ramossisimus	Section Occidentalis Seris Occidentalis	CEH 2037	48	Conterato and Schifino-Wittmann (2006)
L. reitzii	Section Orientalis Series Austro Unifolius	S.T.S. Miotto 2098, 2099, 2105, 2113, 2126, 2131	36	Conterato and Schifino-Wittmann (2006)
L. reitzii	Section Orientalis Series Austro Unifolius	S.T.S. Miotto & Pinheiro 1833, 1844, 1846, M. Pinheiro 310	36	Macieal and Schifino-Wittmann (2002)
L. reticulatus	Section Albi		40	Fedorov (1969) (Gilot, 1965)
L. rothmaleri	Section Scabrispermae		50	Fedorov (1969) (Gilot, 1965)
L. rothmaleri	Section Scabrispermae		52	Darlington and Wylie (1955) (Malheiros, 1942)
L. rothmaleri	Section Scabrispermae		52	Kazmierski (1988)
L. rubriflorus	Section Orientalis Series Austro-orientalis	S.T.S. Miotto 2102, 2104, 2114, 2117, 2123, 2139	36	Conterato and Schifino-Wittmann (2006)
L. rubriflorus	Section Orientalis Series Austro-orientalis	S.T.S. Miotto & Pinheiro 1838, I. Conterato s.n. (ICN 129995)	36	Macieal and Schifino-Wittmann (2002)
L. russellianus	Section Orientalis		36	Fedorov (1969) (Gilot, 1965)
L. saxosus	Section Occidentalis Seris Occidentalis		96	Fedorov (1969) (Atabekova, 1959)
L. saxosus	Section Occidentalis Seris Occidentalis		96	Phillips (1957)
L. semperflorens	Section Occidentalis Seris Occidentalis	CEH 2012, RJE 219	48	Conterato and Schifino-Wittmann (2006)
L. sericeus subsp. asotinensis	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. sericeus subsp. sabinii	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. sericeus subsp. sericeus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. smithianus	Section Occidentalis Seris Occidentalis	RJE 201	48	Schifino-Wittmann unpubl.
L. sp?	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Diers, 1961)
L. sp2	Section Occidentalis Seris Occidentalis	CEH 2004	48	Schifino-Wittmann unpublished.
L. spC	Section Occidentalis Seris Occidentalis	CEH 2045	48	Schifino-Wittmann unpubl.
L. spE	Section Occidentalis Seris Occidentalis	CEH 2218	48	Conterato and Schifino-Wittmann (2006)

L. spF	Section Occidentalis Seris Occidentalis	CEH 2247	48	Schifino-Wittmann unpubl.
L. spK	Section Occidentalis Seris Occidentalis	CEH 2276	48	Schifino-Wittmann unpubl.
L. spL	Section Occidentalis Seris Occidentalis	CEH 2296	48	Conterato and Schifino-Wittmann (2006)
L. spN	Section Occidentalis Seris Occidentalis	CEH 2332	48	Conterato and Schifino-Wittmann (2006)
L. spO	Section Occidentalis Seris Occidentalis	RJE 54	48	Schifino-Wittmann unpubl.
L. spR	Section Occidentalis Seris Occidentalis	RJE 102	48	Schifino-Wittmann unpubl.
L. spS	Section Occidentalis Seris Occidentalis	RJE 126	48	Schifino-Wittmann unpubl.
L. spV	Section Occidentalis Seris Occidentalis	RJE 221	48	Schifino-Wittmann unpubl.
L. subacaulis	Section Occidentalis Seris Occidentalis	CEH 2319	48	Schifino-Wittmann unpubl.
L. subcarnosus	Section Orientalis Series Boreo-orientalis		34	Fedorov (1969) (Atabekova, 1959)
L. subcarnosus	Section Orientalis Series Boreo-orientalis		36	Darlington and Wylie (1955) (Tuschujakowa, 1935)
L. subcarnosus	Section Orientalis Series Boreo-orientalis		36	Fedorov (1969)
L. subcarnosus	Section Orientalis Series Boreo-orientalis		36	Fedorov (1969) (Atabekova, 1959)
L. subcarnosus	Section Orientalis Series Boreo-orientalis		36	Turner (1957)
L. subcarnosus	Section Orientalis Series Boreo-orientalis		48	Fedorov (1969) (Gilot, 1965)
L. subcarnosus	Section Orientalis Series Boreo-orientalis		48	Darlington and Wylie (1955)
L. succulentus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Tuschujakowa, 1935)
L. suksdorfi	Section Occidentalis Seris Occidentalis		96	Fedorov (1969) (Atabekova, 1959)
L. suksdorfi	Section Occidentalis Seris Occidentalis		96	Phillips (1957)
L. sulphureus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Atabekova, 1959)
L. sulphureus	Section Occidentalis Seris Occidentalis		96	Fedorov (1969) (Atabekova,1959)
L. sulphureus subsp. kincaidii	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. sulphureus subsp. subsaccatus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. sulphureus subsp. subsaccatus	Section Occidentalis Seris Occidentalis		96	Phillips (1957)
L. sulphureus subsp. sulphureus	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. sulphureus subsp. whithamii	Section Occidentalis Seris Occidentalis		48	Phillips (1957)
L. superbus* v. *elongata	Section Occidentalis Seris Occidentalis		48	Cave (1956 - 1964) (Snow, 1959)
L. tassilicus	Section Orientalis		36	Darlington and Wylie (1955) (Eichorn, 1949)
L. texensis	Section Orientalis Series Boreo-orientalis		34	Fedorov (1969) (Atabekova, 1959)
L. texensis	Section Orientalis Series Boreo-orientalis		36	Turner (1957)
L. tomentosus	Section Occidentalis Seris Occidentalis	CEH 2325	48	Conterato and Schifino-Wittmann (2006)
L. tominensis	Section Occidentalis Seris Occidentalis	CEH 2294	48	Schifino-Wittmann unpubl.
L. uleanus	Section Orientalis Series Austro Unifolius	S.T.S. Miotto & Pinheiro 1848	36	Macieal and Schifino-Wittmann (2002)
L. vallicola	Section Occidentalis Seris Occidentalis		48	Dunn and Gillett (1966)
L. varius	Section Occidentalis Seris Occidentalis		48	Darlington and Wylie (1955)
L. varius	Section Occidentalis Seris Occidentalis		48	Fedorov (1969)
L. vavilovi	Section Albi		50	Maisurian and Atabekova (1974)

L. velutinus	Section Orientalis Series Austro Unifolius	CEH 2477, 2478	36	Conterato and Schifino-Wittmann (2006)
L. verustus	Section Occidentalis Seris Occidentalis		48	Fedorov (1969) (Tuschujakowa, 1935)
L. villosus	Section Boreo-unifolius	CEH1986	52	Conterato and Schifino-Wittmann (2006)
L. weberbaueri	Section Occidentalis Seris Occidentalis	CEH 2233	48	Schifino-Wittmann unpubl.

References

Atabekova, A.I. (1959) Polyploidie spontaneé du *Lupinus* (Tourn.) L. In IX International Botanical Congress, Vol. 2, pp. 11.

Beaman, J.H., Dejong, D.C.D., & Stoutamire, W.P. (1962) Chromosome numbers in the alpine and subalpine floras of Mexico and Guatemala. *American Journal of Botany*, **49**, 41-50.

Berger, C.A., Witkus, E.R., & McMahon, R.M. (1958) Cytotaxonomic studies ib the Leguminosae. *Bulletin of the Torrey Botanic Club*, **85**, 405-414.

Bragdö, M. (1957) Interspecific crosses in *Lupinus*. Cytology and inheritance of flower colour. *Hereditas*, **43**, 338-357.

Cave, M.S., ed. (1956 - 1964) *Index to plant chromosome numbers*. University of North Carolina Press, Chapel Hill.

Conterato, I. F., and M. T. Schifino-Wittmann. (2006). New chromosome numbers, meiotic behaviour and pollen ferility in American taxa of *Lupinus* (Leguminosae): contributions to taxonomic and evolutionary studies. *Botanical Journal of the Linnean Society* **150**, 229-240.

Cox, B.J. (1973) A chemosystematic comparision of the *Lupinus lepidus-L. caespitosus* complex. *Bulletin of*

References

Atabekova, A.I. (1959) Polyploidie spontaneé du *Lupinus* (Tourn.) L. In IX International Botanical Congress, Vol. 2, pp. 11.

Beaman, J.H., Dejong, D.C.D., & Stoutamire, W.P. (1962) Chromosome numbers in the alpine and subalpine floras of Mexico and Guatemala. *American Journal of Botany*, **49**, 41-50.

Berger, C.A., Witkus, E.R., & McMahon, R.M. (1958) Cytotaxonomic studies ib the Leguminosae. *Bulletin of the Torrey Botanic Club*, **85**, 405-414.

Bragdö, M. (1957) Interspecific crosses in *Lupinus*. Cytology and inheritance of flower colour. *Hereditas*, **43**, 338-357.

Cave, M.S., ed. (1956 - 1964) *Index to plant chromosome numbers*. University of North Carolina Press, Chapel Hill.

Conterato, I. F., and M. T. Schifino-Wittmann. (2006). New chromosome numbers, meiotic behaviour and pollen ferility in American taxa of *Lupinus* (Leguminosae): contributions to taxonomic and evolutionary studies. *Botanical Journal of the Linnean Society* **150**, 229-240.

Cox, B.J. (1973) A chemosystematic comparision of the *Lupinus lepidus-L. caespitosus* complex. *Bulletin of the Torrey Botanic Club*, **100**, 12-17.

Darlington, C.D. & Wylie, A.P. (1955) *Chromosome atlas of flowering plants*, 2nd edn. Allen & Unwin, London.

Diers, L. (1961) Der anteil an polyploiden in den vegetationsgüteln der westkordillere Perus. *Zeitschrift für Botanik*, **49**, 437-488.

Dunn, D.B. (1965) *Lupinus* notes. *Leaflet West. Bot.*, **10**.

Dunn, D. B., AND J. Gillett. (1966). The lupines of Canada and Alaska. *Canada Department of Agriculture Research Branch Monograph* **2**, 87.

Eichorn, A. (1949) A propos de la caryologie du *Lupinus tassilicus* Maire. *Rev. Cytol. et Biol. Veg.*, **11**, 333-350.

Fedorov, A.A., ed. (1969) *Chromosome Numbers of Flowering Plants.* IZdatelstvo, Leningrad.

Gilot, J. (1965) Contribution á l'étude cytotaxinomique des *Genistea* et des *Lotea*. *Cellule*, **65**, 317-347.

Gladstones, J. S. (1980). Recent developments in the understanding, improvement and use of *lupinus*. *In* R. J. Summerfield and A. H. Bunting [eds.], Advances in legume science, 603-612. Royal Botanic Gardens

Gladstones, J.S., Atkins, C.A., & Hamblin, J., eds. (1998) *Lupins as crop plants: biology, production and utilization*, pp 465. CABI, Wallingford.

Gustafsson, Å. & Gadd, I. (1965) Mutations and crop improvements II. The genus *Lupinus* (*Leguminosae.*). *Hereditas*, **53**, 12-39.

Hackbarth, J. & Troll, H.-J. (1959) Lupinen als Körnerleguminosen und Futterpflanzen. *Handb. Pflanzenzücht*, **4**, 1-51.

Heiser, C.B. (1963) Numeración cromosomica de plantas ecutorianas. *Ciencia y Nat.*, **6**, 2-6.

Holmen, K. (1962) Chromosome studies in some artic Alaskan *Leguminosae*. *Botaniska Notiser*, **115**, 87-92.

Huynh, K.L. (1965) Contribution á l'étude caryologique ey embryologique des Phanérogames du Pérou. *Mem. Soc. Helvetique Sci. Nat.*, **85**, 1-178.

Index to Plant Chromosome numbers (IPCN), Missouri Botanical Garden

Kawakami, I. (1930) Chromosome numbers in *Leguminosae*. *Botanical Magazine (Tokoyo)*, **44**, 319-328.

Kazimierski, T. (1988). An attempt to present lupin evolution of the old world /materials and impressions/. *5th International Lupin Conference*, Poland. 103-109.

Ledingham, G.F. (1957) Chromosome numbers of some saskatchewan *Leguminosae* with particular reference to *Astragalus* and *Oxytropis*. *Canadian Journal of Botany*, **35**, 657-666.

Maciel, H. S., and M. T. Schifino-Wittmann. (2002). First chromosome number determinations in south-eastern South American species of *Lupinus* L. (Leguminosae). *Botanical Journal of the Linnean Society* **139**, 395-400.

Maisurian, N.A. & Atabekova, A.I. (1974) *Lupin*, Moscow.

Malheiros, N. (1942) Elementos para o estudo cietológico do género *Lupinus*. *Agron. Lusitana*, **4**, 231-236.

Maude, P.F. (1940) Chromosome numbers in some British planys. *New Phytologist*, **39**, 17-32.

Olsezewska, M.J. (1954) Obserwacje nad euchromocentrami i jaderkiem u *Lupinus albus* L. i *Lupinus luteus* L. *Acta Soc. Bot. Polon.*, **23**, 699-725.

Phillips, L. L. (1957). Chromosome numbers in *Lupinus*. *Madroño* **14**, 30-36.

Reveal, J.L. & Stryer, E.L. (1973) Miscellaneaus chromosomes counts of western American plants. *Great Basin Naturalist*, **33**, 19-25.

Senn, H.A. (1938) Chromosome number relationships in the Leguminosae. *Bibliogr. Genetica*, **12**, 175-345.
Smet, E.d. (1914) Chromosomes, prochromosomes et nucléole dans quelques Dicotylées. *Cellule*, **29**, 335-377.
Snow, R. (1959) Chromosome numbers of California plants, with notes on some cases of cytological interest. *Madroño*, **15**, 81-89.
Troll, H.J., Yagoda, G., & Kunze, A. (1963) Polyploide *Lupinus luteus*. *Züchter*, **33**, 184-190.
Turner, B. L. (1957). The chromosomal distributional relationships of *Lupinus texensis* and *L. subcarnosus*. *Madroño* **14**, 13-16.
Turner, B.L. & Fearing, O.S. (1960) Chromosome numbers in the *Leguminosae* III. Species of the sothwestern United States and Mexico. *American Journal of Botany*, **47**, 603-608.
Tuschnjakowa, M. (1935) Über die Chromosomen ewiniger *Lupinus*-Arten. *Züchter*, **7**, 169-174.
Winge, Ö. (1925) Contributions to the knowledge of chromosome numbers in plants. *Cellule*, **35**, 303-324.

Appendix 17 Morphological matrix

Species	Collector and No.	Cotyledon type	Common Chromosome No.	Polymorphic Chromosome No.	Metaphyll type	First Eophyll type	Seed coat surface pattern
Genista anglica	CEH2339	?	2n = 42	2n = 42	Digitate	?	Simple-foveolate
Genista pilosa	CEH2340	?	2n = 24	2n = 24	Digitate	?	Simple-foveolate
Retama monosperma	CEH2182	?	2n = 48	2n = 48	Digitate	?	Simple-foveolate
L. aff. revolutus	CEH2015	?	?	?	Digitate	?	?
L. albescens	CEH2509	Petiolar	?	?	Digitate	Non-digitate	Simple-foveolate
L. albus	CEH1970	Petiolar	2n = 50	2n = 30, 40, 48, 50	Digitate	Digitate	Granular
L. angustifolius	CEH1972	Petiolar	2n = 40	2n = 40, 48	Digitate	Digitate	Papillose, Simple-foveolate
L. arboreus	CEHs.n.	?	2n = 48	2n = 48	Digitate	?	Rugulate
L. arboreus	G. Kenicer 120	?	2n = 48	2n = 48	Digitate	?	Rugulate
L. arboreus	R. Spellenberg 13488	?	2n = 48	2n = 48	Digitate	?	Rugulate
L. argenteus var. argenteus	R. Spellenberg 13448	?	2n = 48	2n = 48	Digitate	?	Rugulate
L. arizonicus	USDA 577285	?	?	?	Digitate	Digitate	Rugulate
L. arvensis	RJE188	?	?	?	Digitate	?	?
L. mantaroensis	CEH2000	Petiolar	2n = 48	2n = 48	Digitate	Digitate	?
L. ballianus	CEH 038	Petiolar	2n = 48	2n = 48	Digitate	Digitate	?
L. bandelierae	CEH 310	Petiolar	2n =36	2n =36	Digitate	Digitate	Simple-foveolate
L. bandelierae	CEH 318	Petiolar	2n =36	2n =36	Digitate	Digitate	Simple-foveolate
L. bandelierae	CEH2301	Petiolar	2n =36	2n =36	Digitate	Digitate	Simple-foveolate
L. bangii	CEH2444	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. bicolor	USDA 284713	?	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. bracteolaris	S.T.S. Miotto and M. Pinheiro 1819	Petiolar	2n = 32	1,2	Digitate	Digitate	Simple-foveolate
L. bracteolaris	S.T.S. Miotto and M. Pinheiro 1810	Petiolar	2n = 34	1,2	Digitate	Digitate	Simple-foveolate
L. brevicaulis	Baker 10768	?	?	?	Digitate	Digitate	Rugulate
L. brevicaulis	C.D. Bailey 242	?	?	?	Digitate	Digitate	Rugulate
L. ellsworthianus	RJE	Petiolar	2n = 48	2n = 48	Digitate	?	?

L. chachas	CEH2027	?	?	?	Digitate	?	?
L. chamissonis	G. Kenicer 121	?	?	?	Digitate	?	?
L. chrysanthus	CEH2264	Petiolar	2n = 48	2n = 48	Digitate	Digitate	?
L. chrysanthus	CEH2390	Petiolar	2n = 48	2n = 48	Digitate	Digitate	?
L. cosentinii	CEH1967	Petiolar	2n = 32	2n = 32	Digitate	Digitate	Fascicular
L. crotalarioides	CEH2481	?	2n =36	2n =36	Non-digitate	?	?
L. crotalarioides	CEH2482	?	2n =36	2n =36	Non-digitate	?	?
L. cumulicola	CEH1983	Petiolar	2n = 52	2n = 52	Non-digitate	Non-digitate	Simple-foveolate
L. cumulicola	CEH1989	Petiolar	2n = 52	2n = 52	Non-digitate	Non-digitate	Simple-foveolate
L. densiflorus	C. Drummond 0331A02	Connate	2n = 48	2n = 48	Digitate	Digitate	Rugulate and Protuberant
L. digitatus	Genbank AY338948	?	2n = 32	2n = 36, 40, 42, 52	Digitate	Digitate	Fascicular
L. gibertiannus	S.T.S. Miotto and M. Pinheiro 1835	Petiolar	2n =36	2n =36	Digitate	Digitate	Simple-foveolate
L. guaraniticus	CEH2488	Petiolar	2n =36	2n =36	Non-digitate		Simple-foveolate
L. guaraniticus	G. Kenicer s.n.	Petiolar	2n =36	2n =36	Non-digitate	?	Simple-foveolate
L. havardii	USDA 910433	Petiolar	2n =36	2n =36	Digitate	Digitate	Simple-foveolate
L. havardii	USDA 930123	Petiolar	2n =36	2n =36	Digitate	Digitate	Simple-foveolate
L. hirsutissimus	S & S Seeds Ltd s.n.	Petiolar	?	?	Digitate	Digitate	Substriate simple-foveolate
L. hirsutissimus	USDA 920120	Petiolar	?	?	Digitate	Digitate	Substriate simple-foveolate
L. hispanicus	J. Pannell 1	?	2n = 52	2n = 52	Digitate	Digitate	Simple-foveolate
L. huaronensis	CEH2241	?	2n = 48	2n = 48	Digitate	?	?
L. purosericeus	CEH2248	?	2n = 48	2n = 48	Digitate	?	?
L. lanatus	S.T.S. Miotto and M. Pinheiro 1806	Petiolar	2n =36	2n =36	Digitate	Non-digitate	Simple-foveolate
L. lanatus	S.T.S. Miotto and M. Pinheiro 1837	Petiolar	2n =36	2n =36	Digitate	Non-digitate	Simple-foveolate
L. latifolius	Baker 10938	?	?	?	Digitate	Digitate	Rugulate
L. lepidus	C. D. Bailey 254	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. lindleyanus	RJE141	?	2n = 48	2n = 48	Digitate	?	?
L. linearis	S.T.S. Miotto and M. Pinheiro 1813	Petiolar	2n = 34	2n = 32, 34	Digitate	Digitate	Simple-foveolate

L. linearis	S.T.S. Miotto and M. Pinheiro 1798	Petiolar	2n = 32	2n = 32, 34	Digitate	Digitate	Simple-foveolate
L. luteus	CEH1973	Petiolar	2n = 52	2n = 48, 50, 52	Digitate	Digitate	Simple-foveolate
L. magnistipulatus	M. Pinheiro 164	Petiolar	2n =36	2n =36	Digitate	Non-digitate	?
L. magnistipulatus	S.T.S. Miotto and M. Pinheiro 1840	Petiolar	2n =36	2n =36	Digitate	Non-digitate	?
L. micranthus	CEH1982	?	2n = 52	2n = 48, 50, 52	Digitate	Digitate	Papillose
L. microcarpus	M. Gardener 1513	Connate	?	?	Digitate	?	Rugulate and Protuberant
L. microphyllus	CEH2238	Petiolar	?	?	Digitate	Digitate	?
L. microphyllus	CEH2272	Petiolar	2n = 48	2n = 48	Digitate	Digitate	?
L. misticola	RJE144	?	?	?	Digitate	?	?
L. mollendoensis	Weigend 97/703	?	?	?	Digitate	?	?
L. montanus	CEH2174	?	?	?	Digitate	?	Rugulate
L. multiflorus	S.T.S. Miotto and M. Pinheiro 1825	Petiolar	2n =36	2n =36	Digitate	Non-digitate	?
L. mutabilis	CEH2291	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. mutabilis	CEH2009	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. neomexicanus	R. Spellenberg 13422	?	?	?	Digitate	?	Rugulate
L. nubigenus	RJE200	?	2n = 48	2n = 48	Digitate	?	?
L. odoratus	C. Drummond 0311A01	Connate	?	?	Digitate	Digitate	Rugulate and Protuberant
L. paraguariensis	CEH2495	?	2n =36	2n =36	Non-digitate, Digitate	Non-digitate	?
L. paraguariensis	CEH2496	?	2n =36	2n =36	Non-digitate, Digitate	Non-digitate	?
L. paranensis	S.T.S. Miotto and M. Pinheiro ind2_18	Petiolar	2n =36	2n =36	Digitate	?	?
L. parvifolius	CEH2480	?	?	?	Non-digitate, Digitate	?	?
L. piurensis	CEH1997	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. piurensis	RJE110	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate

L. praestabilis	CEH2353	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. prostratus	CEH2239	Petiolar	2n = 48	2n = 48	Digitate	?	?
L. pubescens	RJE176	?	?	?	Digitate	?	?
L. pubescens	CEH2037	?	2n = 48	2n = 48	Digitate	Digitate	?
L. pulvinaris	CEH2273	Petiolar	?	?	Digitate	Digitate	?
L. ramossisimus	CEH2010	Petiolar	?	?	Digitate	Digitate	?
L. reitzii	S.T.S. Miotto and M. Pinheiro 1846	Petiolar	2n =36	2n =36	Digitate	Non-digitate	?
L. revolutus	RJE174	?	?	?	Digitate	?	?
L. rubriflorus	S.T.S. Miotto and M. Pinheiro 1838	Petiolar	2n =36	2n =36	Digitate	Non-digitate	?
L. semperflorens	CEH2012	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. semperflorens	CEH2014	?	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. semperflorens	RJE219	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. semperflorens	RJE220	Petiolar	2n = 48	2n = 48	Digitate	?	Rugulate
L. sierra-blancae	R. Spellenberg 13409	?	?	?	Digitate	?	Rugulate
L. tominensis	CEH2294	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate
L. tominensis	CEH2370	?	?	?	Digitate	?	?
L. tomentosus	CEH2322	Petiolar	?	?	Digitate	Digitate	?
L. sp1	CEH2002	Petiolar	?	?	Digitate	Digitate	?
L. sp1	CEH2005	Petiolar	?	?	Digitate	Digitate	?
L. sp2	CEH2004	Petiolar	2n = 48	2n = 48	Digitate	Digitate	?
L. sp7	RJE59	?	?	?	Digitate	?	?
L. sp8	RJE168	Petiolar	?	?	Digitate	?	?
L. sp9	RJE206	?	?	?	Digitate	?	?
L. sp	RJE217	?	2n = 48	2n = 48	Digitate	?	?
L. sp4	CEH2107	?	?	?	Digitate	?	?
L. sp5	CEH2117	?	?	?	Digitate	?	?

L. sp4	CEH2118	?	?	?	Digitate	?	?
L. sp6	CEH2160	?	?	?	Digitate	?	Rugulate
L. subacaulis	CEH2319	Petiolar	2n = 48	2n = 48	Digitate	Digitate	?
L. subsessilis	CEH2479	Petiolar	?	?	Non-digitate	?	?
L. tarapacensis	M. Gardener 220A	?	?	?	Digitate	?	Rugulate
L. texensis	-	Petiolar	2n =36	2n =36	Digitate	Digitate	Simple-foveolate
L. texensis	S & S Seeds Ltd s.n.	Petiolar	2n =36	2n =36	Digitate	Digitate	Simple-foveolate
L. truncatus	S & S Seeds Ltd s.n.	Petiolar	?	?	Digitate	Digitate	Substriate simple-foveolate
L. uleanus	S.T.S. Miotto and M. Pinheiro 1848	Petiolar	2n =36	2n =36	Digitate	Non-digitate	?
L. velutinus	CEH2478	Petiolar	2n =36	2n =36	Non-digitate	?	?
L. velutinus	CEH2477	Petiolar	2n =36	2n =36	Non-digitate	?	?
L. villosus	CEH1986	Petiolar	2n = 52	2n = 52	Non-digitate	Non-digitate	Simple-foveolate
L. villosus	CEH1987	Petiolar	2n = 52	2n = 52	Non-digitate	Non-digitate	Simple-foveolate
L. weberbaueri	CEH2233	Petiolar	2n = 48	2n = 48	Digitate	Digitate	Rugulate

Appendix 18 Infrageneric classification of *Lupinus*

Author	Scope	Classification		No. species	Species
Agardh (1835)	Whole genus	Foliis digitatis	Annuals		
			Tribe A. Lutei	4	*L. densiflorus, L. luteus, L. menziesii, L. microcarpus*
			Tribe B. Pilosi	7	*L. aureus, L. bracteolaris, L. consinnus, L. hirsutissimus, L. hirsutus, L. pilosus, L. varius*
			Tribe C. Albi	4	*L. albus, L. lindleyanus, L. mutabilis, L. termis*
			Tribe D. Angustifolii	4	*L. angusifolius, L. leptophyllus, L. linifolius, L. nanus*
			Tribe E. Micranthi	4	*L. bicolor, L. gracilis, L. micranthus, L. pusillus*
			Perennials		
			Tribe F. Polyphylli	7	*L. cytisoides, L. elegans, L. grandifolius, L. latifolius, L. lepidus, L. minimus, L. polyphyllus*
			Tribe G. Nootkatenses	9	*L. affinis, L. alopecuroides, L. hornmanni, L. marshallianus, L. mexicanus, L. montanus, L. nootkatensis, L. nubigenus, L. stipulatus*
			Tribe H. Arborei	3	*L. arboreus, L. glabratus, L. rivularis*
			Tribe I. Sericei	14	*L. argyraeus, L. aridus, L. chamissonis, L. laxiflorus, L. leucophyllus, L. perennis, L. plumosus, L. sabinianus, L. sericeus, L. sulphurous*
			Tribe J. Paniculati	11	*L. albifrons, L. decumbens, L. douglasii, L. flexuosus, L. littoralis, L. macrocarpus, L. multiflorus, L. paniculatus, L. rupestris, L. sarmentosus, L. tomentosus*
			Tribe K. Microphylli	4	*L. knuthii, L. microphyllus, L. prostratus, L. smithianus*
		Follis integris		4	*L. cochinchinensis, L. diffusus, L. integrifolius, L. villosus*
		Species Inquirende		7	*L. africanus, L. bimaculatus, L. incanus, L. milleri, L. pulchellus, L. variegatus, L. versicolor*
Bentham (1859)	Brazil	Simplicifolieae		9	*L. arenarius, L. crotalarioides, L. decurrens, L. laevigatus, L. ovalifolius, L. parvifolius, L. subsessilis, L. vaginans, L. velutinus*
		Digitatae		6	*L. bracteolaris, L. comptus, L. lanatus, L. linearis, L. hilarianus, L. multiflorus*

Watson (1873)	Extra-tropical North America	Section Lupinus	Perennials	38	*L. albicaulis, L. albifrons, L. andersoni, L. arboreus, L. argenteus, L. aridus, L. articus, L. barbiger, L. breweri, L. burkei, L. caespitosus, L. calcaratus, L. confertus, L. diffusus, L. douglasii, L. flexuosus, L. holoserieus, L. kingii, L. laxiflorus, L. lepidus, L. leucophyllus, L. littoralis, L. lyallii, L. meionanthus, L. minimus, L. nootkatensus, L. nuttallii, L. ornatus, L. palmeri, L. parviflorus, L. perennis, L. polyphyllus, L. rivularis, L. sabinii, L. sericeus, L. sitgreavii, L. sulphurous, L. villosus, L. wyethii*
			Annuals	11	*L. affinis, L. concinnus, L. gracilis, L. hirsutissimus, L. leptophyllus, L. micranthus, L. nanus, L. sparsiflorus, L.stiveri, L. subcarnosus, L. truncatus*
		Section Platycarpos		5	*L. brevicaulis, L. bridgesii, L. densiflorus, L.microcarpus, L. pusillus*
		Section Lupinellus		1	*L. uncialis*
Taubert (1894)	Whole genus	Digitatae gerontogeae		12 (5 listed)	*L. albus, L. angustifolius, L. hirsutus, L. luteus, L. termis*
		Digitatae neogeae	Eulupinus (Wats.)		
			perennials	60	inc. *L. arboreus, L. argenteus, L. douglasii, L. holoserieus, L. leucophyllus, L. mutabilis, L. polyphyllus*
			annuals		inc. *L. hilarianus, L. nanus, L. sparsiflorus, L. subcarnosus*
			Platycarpos Wats.	2	*L. microcarpus, L. pusillus*
			Lupinellus Wats.	1	*L. uncialis*
		Simplicifolieae		12 (7 listed)	*L. coriaceus, L. crotalaroides, L. decurrens, L. diffusus, L. subsessilis, L. vaginans, L. villosus*
Piper & Robinson (1906)	North America, State of Washington	Subgen. I. Platycarpos Wats.		2	*L. microcarpus, L. pusillus*
		Subgen. II Lupinus	Annuals		
			1. Micranthi	2	*L. bicolor, L. micranthus*
			Perennials		
			2. Sericei	14	*L. albicaulis, L. alpicola, L. aridus, L. canescens, L. lepidus, L. leucophyllus, L. littoralis, L. lyallii, L. mimus, L. ornatus, L. piperi, L. sericeus, L. subsericeus, L. sukdorfi*
			3. Saxosi	4	*L. saxosus, L. subalpinus, L. volcanicus, L. wyethii*
			4. Rivularis	3	*L. burkei, L. polyphyllus, L. rivularis*
			5. Sulphurei	2	*L. sabinii, L. sulphurous*
			6. Calcarati	1	*L. laxiflorus*

Ascherson & Graebner (1907)	Whole genus	A. Subgenus Eulupinus			*Everything else.*
		B. Subgenus Platycarpos			*L. densiflorus et al*
Rydberg (1917)	North America, Rocky Mountains and adjacent plains		I. Plattenses	1	*L. plattensis*
			II. Parviflori	14	*L. alpestris, L. argenteus, L. floribundus, L. laxispicus, L. leptostachyus, L. leucanthus, L. maculates, L. parviflorus, L. pseudoparviflorus, L. rubricaulis, L. scheuberae, L.spathulatus, L. stenophyllus, L. tenellus*
			III.Calcarati	1	*L. calcaratus*
			IV. Argophylli	4	*L. aduncus, L. argentinus, L. laxiflorus, L. lupinus*
			V. Sulphurei	3	*L. flavescens, L. lucidulus, L. sulphurous*
			VI. Albicaules	1	*L. albicaulis*
			VII. Foliosi	8	*L. adscendens, L. argenteus, L. foliosus, L. ingrates, L. laxus, L. macounii, L. monticola, L. pulcherrimus*
			VIII. Ornati	8	*L. flexuosus, L. greenei, L. hellerae, L. humicola, L. leucopsis, L. ornatus, L. rydbergii, L. tenuispicus*
			IX. Cytisoides	6	*L. ampus, L. burkei, L. cytisoides, L. marianus, L. nootkatensis, L. wyethii*
			X. Polyphylli	2	*L. glabratus, L. polyphyllus*
			XI. Ammophili	2	*L. ammophilus, L. glabratus*
			XII. Cyanei	7	*L. apronosus, L. bakeri, L. comatus, L. cyaneus, L. dichrous, L. flavicaulis, L. sericeus*
			XIII. Leucophylli	7	*L. barbiger, L. jonesii, L. leucophyllus, L. macrostachys, L. plumosus, L. ramosus, L. retrorsus*
			XIV. Lepidi	5	*L. aridus, L. depressus, L. erectus, L. lepidus, L. minimus*
			XV. Candicantes	3	*L. candicans, L. evermannii, L. psoraliodes*
			XVI. Monticolae	2	*L. monticola, L. roseolus*
			XVII. Caespitosi	3	*L. caespitosus, L. cusickii, L. watsonii*
			XVIII. Concinni	1	*L. micensis*
			XIX. Pusilli	5	*L. brevicaulis, L. kingii, L. pusillus, L. rubens, L. scaposus*
Smith (1944)	North America, Pacific States		A. Microcarpi	5	*L. densiflorus, L. horizontalis, L. luteolus, L. microcarpus, L. subvexus*
			B. Pusilli	5	*L. brevicaulis, L. odoratus, L. pusillus, L. rubens, L. shockleyi*
			C. Concinni	1	*L. concinnus*
			D. Stiversiani	1	*L. stiversii*
			E. Succulenti	1	*L. succulentus*

			F. Sparsiflori	6	*L. benthamii, L. citrinus, L. deflexus, L. hirsutissimus, L. sparsiflorus, L. truncatus*
			G. Micranthi	4	*L. bicolor, L. pachylobus, L. nanus, L. micranthus*
			H. Caespitosi	6	*L. aridus, L. caespitosus, L. hypolasius, L. lepidus, L. lyallii, L. pratensis*
			I. Calcarati	13	*L. biddlei, L. caudatus, L. corymbosus, L. holosericeus, L. inyoensis, L. leucopsis, L. meionanthus, L. mucronulatus, L. nevadensis, L. onustus, L. oreganos, L. sulphureus, L. tenellus*
			J. Breweriani	1	*L. breweri*
			K. Leucophylli	2	*L. leucophyllus, L. peirsonii*
			L. Sericei	15	*L. abramsii, L. albifrons, L. alpicola, L. breweri var., L. cervinus, L. chamissonis, L. excubitus, L. grayi, L. lapidicola, L. ludovicianus, L. ornatus, L. sabinii, L. sericatus, L. sericeus, L. suksdorfii*
			M. Arborei	8	*L. arboreus, L. corvilleri, L. gracilentus, L. littoralis, L. longifolius, L. rivularis, L. tidestromii, L. variicolor*
			N. Albicaules	7	*L. adsurgens, L. albicaulis, L. andersonii, L. croceus, L. elatus, L. formosus, L. fulcratus*
			O. Latifolii	1	*L. latifolius*
			P. Polyphylli	3	*L. burkei, L. polyphyllus, L. superbus*
			Q. Saxosi	3	*L. arcticus, L.minimus, L. saxosus, L. volcanicus*
			R. Magnifici	1	*L. magnificus*
Smith (1946)	North America exc. Mexico	Subgenus Lupinus	A5 Concinnus complex		inc. *L. concinnus*
			A6 Stiversii complex	1	*L. stiversii*
			A7 Succulentus complex	1	*L. succulentus*
			A8 Sparsiflorus complex		inc. *L. sparsiflorus*
			A9 Micranthus-Nanus complex		inc. *L. micranthus, L. nanus*
			A10 Subcarnosus complex		inc. *L. subcarnosus*
			P1 Caespitosus-Lepidus complex		inc. *L. caespitosus, L. lepidus*
			P2 Breweri complex		inc. *L. breweri*
			P3 Calcaratus complex		inc. *L. calcaratus*
			P3a		inc. *L. depressus*

			P3b		inc. *L. aduncus, L. helleri*
			P3c		inc. *L. argenteus, L. parviflorus*
			P3d		inc. *L. ingratus L. laxispicatus*
			P3e		inc. *L. caudatus, L. sulphureus*
			P3f		inc. *L. laxiflorus*
			P3g		inc. *L. leucophyllus*
			P4 Sericeus-Albrifons complex		inc. *L. albifrons, L. sericeus*
			P5 Arboreus complex		inc. *L. arboreus*
			P6 Albicaulis complex		inc. *L. albicaulis*
			P7 Latifolius		inc. *L. latifolius*
			P8 Articus-Wyethii complex		inc. *L. articus, L. wyethii*
			P9 Polyphyllus complex		inc. *L. polyphyllus*
			P10 Magnificus complex		inc. *L. magnificus*
			P11 Perennis complex		inc. *L. perennis*
			P12 Villosus complex		inc. *L. villosus*
			P13 Kiskensis complex		inc. *L. kiskensis*
		Subgenus Platycarpos	A1	1	*L. uncialis*
			A2 Microcarpus complex	13	*L. arenicola, L. bridgesii, L. densiflorus, L. glareosus, L. horizontalis, L. lacteus, L. luetolus, L. malacophyllus, L. menziesii, L. microcarpus, L. palustris, L. ruber, L. subvexus*
			A3	1	*L. malycophyllus*
			A4 Pusillus complex	13	*L. argillaceus, L. brevicaulis, L. capitatus, L. dispersus, L. flavoculatus, L. intermontanus, L. kingii, L. odoratus, L. pusillus, L. rubens, L. scaposus, L. shockleyi, L. sileri*
Smith (1945) New World	Brazil	Annuals	Group 1	1	*L. comptus*
			Group 2	2	*L. czermakii, L. bracteolaris*
			Group 3	3	*L. lanatus, L. linearis, L. mutliflorus,*
			Group 4	1	*L. luteus*
			Group 5	2	*L.hilarianus, L. reineckianus*
			Group 6	6	*L. bandelierae, L. dusenianus, L. paranensis, L. regnellianus, L. russellianus, L. uleanus*
		Perennials	Group 7	1	*L. paraguariensis*

			Group 8	17	*L. aliattenuatus, L. arenarius, L. attenuatus, L. coriaceus, L. crotalaroides, L. decurrens, L. guaraniticus, L. glaziouanus, L. insignis, L. laevigatus, L. lutzelburgianus, L. nitidissimus, L. ovalifolius ,L. parvifolius, L. provensalanus, L. schwackeanus, L. sellowianus, L. subsessiles, L. succisaefolius, L. velutinus*
Smith (1951)	Peru		A Microcarpoids	1	*L. microcarpus*
			B Pulvinaroids	1	*L. pulvinaris*
			C Concoids	1	*L. conicus*
			D Asagrayanoids	1	*L. asa-grayanus*
			E Ananeanoids	1	*L. ananeanus*
			F Nubigenoids	16	*L. ccorilazensis, L. chumbivilcensis, L. chrysanthus, L. cuzcensis, L. cymboides, L. egens, L. espinarensis, L. fieldii, L. huanonensis, L. kalenbornorum, L. malacotrichus, L. pinguis, L. oquendoanus, L. sericeolodix, L. weberaueri, L. yauleyensis*
			G Matucanicoids	2	*L. aridulus, L. matucanicus,* later added *L. acopalus, L. albicolor, L. austrobivolor, L. huariacus, L. lorenzensis, L. mantaroensis, L. otuzcoensis*
			H Prostratoids	10	*L. allargyreius, L. chavanillensis, L. chongos-bajous, L. dorae, L. falsoprostratus, L. pickeringii, L. prostratus, L. pucapucensis, L. subumbellatus, L. velillensis*
			I Microphylloids	9	*L. biiclinatus, L. carazensis, L. cesaranus, L. chlorlepis, L. gilbertianus, L. peruvianus,* later added *L. chachas, L. hautcarasensis, L. yarushensis*
			J Lindleyanoids	26	*L. alinanus, L. albert-smithii, L. arequipensis, L. cesar-vargasi, L. diasemus, L. disjunctus, L. exochus, L. francis-whittieri, L. juninesis, L. lindleyanus, L. lorenzensis, L. mollendoensis, L. nehmadae, L. neocotus, L. puyupatensis, L. romasanus, L. sandiensis, L. syriggedes, L. urubambensis, L. vilcabambensis, L. vilcabambensis, L. volubilis, L. yanahuancensis, L. Ynesiae,* later added *L. carhuamayus, L. Carlos-Ochoai*
			K Montanoids	3	*L.montanus, L. praestabilis, L. proculaustrinus*
			L Misticoloids	1	*L. misticola*
			M Mutabiloids	4	*L. mutabilis,* later added *L. eanophyllus, L. falsomutabilis, L. tayacajensis*

			N Saxatiloids	10	*L. carpapaticus, L. convencionensis, L. hinkleyorum, L. lechleranus, L. munzianus, L. paruroensis, L. praetermissus, L. saxatilis, L. storkianus, L. sufferugineus*
			O Tomentosoids	41	*L. ampaiensis, L. antensis, L. arizelus, L. bombycinocarpus, L. brachypremonon, L. cachupatensis, L. calcensis, L. cochopatensis, L. eriodcladus, L. gayanus, L. grauensis, L. hererranus, L. holwayorum, L. hortonianus, L. inuitatus, L. james-westii, L. jelksianus, L. ledigianus, L. notabilis, L. paucartambensis, L. perblandus, L. pisacensis, L. platytenus, L. punensis, L. staffordiae, L. subtomentosus, L. soukupianus, L. tarapacensis, L. tetracercophorus, L. tometosus, L. toratensis, L. ulbrichianus, L. urcoensis, L. vargasianus, L. william-lobbi,* later added *L. amboensis, L. austrosericeus, L. barbatilabius, L. colcabambensis, L. pachitensis, L. surcoensis*
			P Paniculatoids	29	*L. adinoathus, L. alimanens, L. ambaticus, L. appositus, L. asymbepus, L. ballinaus, L. celsimontanus, L. condesiflorus, L. cookianus, L. edysomatus, L. elsworthianus, L. hamaticalys, L. hornmanni, L. huancayoensis, L. macbrideanus, L. martinetianus, L. mathewsianus, L. paniculatus, L. praealtus, L. quellomaus, L. ramosussumus, L. richardianus, L. semiprostratus, L. wilkesianus, L. xanthophyllus,* later added *L. ochoanus, L. piurensis, L. semperflorens, L. tarmaensis*
			Q Visoensoids	1	*L. visoensis*
Maisurian and Atabekova (1974)			Tribe L. lutei	4	*L. hispanicus, L. luteus, L. palaestinus, L. rothmaleri*
			Tribe L. pilosi	3	*L. digitatus, L. hirstus, L. pilosus*
			Tribe L. albi	4	*L. albus, L. graecus, L. termis, L. vavilovi*
			Tribe L. angustifolii	2	*L. angustifolius, L. linifolius*
			Tribe L. nani	8	*L. benthami, L. gracilis, L. leptophyllus, L. lindleyanus, L. linearis, L. micranthus, L. nanus, L. truncatus*
			Tribe L. hirsutissimi	10	*L. abramsii, L. bracteolaris, L. brittoni, L. concinnus, L. elegans, L. hartwegii, L. hirsutissimus, L. pubescens, L. sparsiflorus, L. subcarnosus*
			Tribe L. mutabili	3	*L. hybridus, L. mutabilis, L. ornatus*
			Tribe L. barkeri	4	*L. aureus, L. barkeri, L. stiversii, L. succulentus*
			Tribe L. microcarpi	3	*L. densiflorus, L. microcarpus, L. pusillus*
			Tribe L. polyphylli	6	*L. cytisoides, L. grandifolius, L. latifolius, L. lepidus, L. minimus, L. polyphyllus*

			Tribe L. nootkatensii	10	*L. affinis, L. alopecuroides, L. hornmanni, L. leucophyllus, L. marschallianus, L. mexicanus, L. montanus, L. nootkatensis, L. nubigenus, L. stipulatus*
			Tribe L. sericei	8	*L. albicaulis, L. argenteus, L. aridus, L. laxiflorus, L. perennis, L. sabinianus, L. sericeus, L. sulphureus*
			Tribe L. formosi	4	*L. chamissonis, L. formusus, L. leucopsis, L. plumosus*
			Tribe L. paniculati	11	*L. albifrons, L. decumbens, L. douglasii, L. flexuosus, L. littoralis, L. macrocarpus, L. multiflorus, L. paniculatus, L. rupestris, L. sarmentosus, L. tomentosus*
			Tribe L. microphylli	3	*L. microphyllus, L. prostratus, L. smithianus*
			Tribe L. arborei	4	*L. arborescens, L. arboreus, L. glabratus, L. rivularis*
Plitmann and Heyn (1984)	Old World	Sect. Scabrispermae		7	*L. atlanticus, L. cosentinii, L. digitatus, L. palaestinus, L. pilosus, L. princei, L. somaliensis*
Gross (1984)		1. New World	1.1 South America		
			1.1.1 Atlantic		
			1.1.2 Simple-leaved		
			1.2 North American		
		2. Old World	2.1 Rough-seeded		
			2.2 Smooth-seeded		
Sholars (1996)	North America - perennials		1. Albifrons complex	13	*L. albifrons, L. alpicola, L. cervinus, L. chamissonis, L. excubitius, L. grayi, L. lapidicola, L. ludovicianus, L. ornatus, L. sabinii, L. sericatus, L. sericeus, L. suksdorfi*
			2. Albicaules complex	12	*L. adsurgens, L. albicaules, L. andersonii, L. angustiflorus, L. croeus, L. dalesae, L. elatus, L. elmeri, L. formosus, L. fulcratus, L. hyacinthinus, L. padre-crowleyi*
			3. Arboreus complex	6	*L. arboreus, L. littoralis, L. longifolius, L. rivularis, L. tidestromii, L. variicolor*
			4. Argenteus complex	17	*L. argenteus, L. biddlei, L. corymbosus, L. fulvomaculatus, L. hillii, L. holoserieus, L. leucopsis, L. meionanthus, L. mucronulatus, L. nevadensis?, L. onustus, L. oreganus, L. parviflorus, L. sulphureus, L. utahensis*
			5. Breweri complex	6	*L. breweri, L. constancii, L. duranii, L. lapidicola, L. obtusilobus, L. tegeticulatus*
			6. Polyphyllus complex	7	*L. articus, L. magnificus, L. minimus, L. nevadensis?, L. polyphyllus, L. saxosus, L. volcanicus*
			7. Lepidus complex	2	*L. lepidus, L. pratensis*
			8. Latifolius complex	1	*L. latifolius*
			9. Perennis complex	2	*L. perennis, L. plattensis*
			10. Unifoliatata complex	3	*L. villosus, L. diffusus, L. westianus*

Kurlovich (2002)	Whole genus	Subgenus Platycarpos		200-1000	*L. arboreus, L. rivularis, L. multiflorus, L. littoralis, L. paniculatus, L. chamissonis, L. polyphyllus, L. latifolius, L. minimus, L. leucophyllus, L. nootkatensis, L. affinis, L. perennis, L. argenteus, L. sulphureus, L. aridus, L. mutabilis, L. ornatus, L. elegans, L. hartwegii, L. subcarnosus, L. pubescens, L. nanus, L. bicolor, L. truncatus, L. polycarpos, L. succulentus, L. stiversii, L. microcarpos, L. pusillus, L. densiflorus +*
		Subgenus Lupinus		11	*L. albus, L. angustifolius, L. hispanicus, L. luteus, L. micranthus, L. anatolicus, L. atlanticus, L. cosentinii, L. digitatus, L. palaestinus, L. pilosus, L. princei*
Ainouche, Bayer and Misset (2004)	Old World		Sect. Albi	1	*L. albus*
			Sect. Angustifolii		*L. angustifolius*
			Sect. Lutei		*L. hispanicus, L. luteus*
			Sect. Micranthi		*L. micranthus*
			Sect. Scabrispermae		*L. anatolicus, L. atlanticus, L. cosentinii, L. digitatus, L. palaestinus, L. pilosus, L. princei*

Eastwood & Hughes (in press)	Whole genus	Old World	Section Albi	1	*L. albus*
			Section Angustifolii	1	*L. angustifolius*
			Section Lutei	2	*L. hispanicus, L. luteus*
			Section Micranthi	1	*L. micranthus*
			Section Scabrispermae	8	*L. anatolicus, L. atlanticus, L. cosentinii, L. digitatus, L. palaestinus, L. pilosus, L. princei, L. somaliensis*
		New World	Section Boreo-unifolius	ca. 4	*L. cumulicola, L. diffusus, L. villosus, L. westianus*
			Section Occidentalis	ca. 250	
			Series Platycarpos	ca. 12	*L. affinis?, L. brevicaulis, L. bridgesii, L. densiflorus, L. horizontalis, L. kingii, L. luteolus, L. microcarpus, L. odoratus, L. pusillus, L. rubens, L. scaposus, L. shockleyi, L. subvexus, L. uncialis*
			Series Occidentalis	ca. 238	
			Section Orientalis	ca. 32	*L. attenuatus, L. comptus, L. czermakii, L. dusenianus, L. reineckianus, L. regellianus, L. setifolius (placement of these taxa within series unknown)*
			Series Boreo-orientalis	ca. 4	*L. harvardi, L. leonensis?, L. subcarnosus, L. texensis,*
			Series Austro-unifolius	ca. 14	*L. albescens, L. arenarius, L. aureonitens, L. coriaceus, L. crotalaroides, L. decurrens, L. guaranticus, L. laevigatus, L. magnistipulatus, L. multiflorus, L. ovalifolius, L. paraguariensis, L. prouvensalanus, L. velutinus*
			Series Austro-orientalis	ca. 7	*L. bandelierae, L. bracteolaris, L. gibertianus, L. linearis, L. paranensis, L. rubiflorus, L. uleanus*

Colours represent the newly proposed infrageneric classification. Species names which are unrecognised are coloured black.
Species in the Andean radiation are underlined.

References

Agardh, J.G. (1835) *Synopsis Generis Lupini*, Berlin.

Ainouche, A., Bayer, R.J., & Misset, M.-T. (2004) Molecular phylogeny, diversification and character evolution in *Lupinus* (Fabaceae) with special attention to Mediterranean and African lupines. *Plant Systematics and Evolution*, **246**, 211-222.

Ascherson, P. & Graebner, P. (1907) *Sinopsis der Mitteleuropaischen flora*, Leibzig.

Bentham, G. (1859). Papilionaceae. In *Flora Brasiliensis* (ed C.F.P. Martius), Vol. 15(1), pp. 9-16, Munich.

Gross, R. (1984) First Reinhold von Sengbusch Memorial lecture: Lupins in the old and new world - A biological-cultural coevolution. In *4th Internation Lupin Conference*, Vol. 1, pp. 244-277.

Kurlovich, B. S. 2002. Lupins: Geography, classification, genetic resource and breeding. Intan, St Petersburg.

Maisurian, N.A. & Atabekova, A.I. (1974) *Lupin*, Moscow.

Piper, C.V. & Robinson, B.L. (1906). *Lupinus*. In *Flora of the state of Washington* (ed C.V. Piper), pp. 350-358.

Plitmann, U. & Heyn, C.C. (1984) Old World *Lupinus*: taxonomy, evolutionary relationships and links with New World species. In *3rd International Lupin Conference*, pp. 55-66.

Rydberg, P.A. (1917) *Flora of the Rocky Mountains and adjacent plains: Colorado, Utah, Wyoming, Idaho, Montana, Saskatchewan, Alberta, and neighboring parts of Nebraska, South Dakota, North Dakota, and British Columbia*, 1st ed edn.

Sholars, T.A. (1996) The systematics of perennial *Lupinus* in North America. In Towards the 21st Century. *Proceeding of the 8th International Lupins Conference* (ed G. Hill, D.), pp. 401-413.

Smith, C.P. (1944). 3 *Lupinus* L. In *Illustrated Flora of the Pacific States* (ed L. Abrams), Vol. 2, pp. 485-519. Stanford University Press, Stanford, California.

Smith, C.P. (1945). Paper 46, *Lupinus* in Brazil. In *Synopsis Lupinorum* (ed C.P. Smith), Vol. Signature thirty, pp. 481-500. Privately published.

Smith, C.P. (1946). *Lupinus* in North America - Basic Taxonomy. In *Species Lupinorum*, pp. 545-576.

Smith, C., Piper (1951). Peru Nineteen, Miscellany one, Paper one-hundred. In *Species Lupinorum* (ed C. Smith, Piper), pp. 712-718. privately published.

Taubert, A. (1894). *Lupinus*. In *Die Natürlichen Pflanzenfamilien* (eds H.G. Engler & K.A. Prantl), Vol. 3(3), pp. 231-232, Berlin.

Watson, S. (1873) Revision of the extra-tropical North American species of the genus *Lupinus*. *Proceedings of the American Academy of Arts and Science*, **8**, 517-548.