Lost crops of the Incas: origins of domestication of the Andean pulse crop 'tarwi', *Lupinus mutabilis*.[1]

Guy W. Atchison[2,7], Bruno Nevado[3], Ruth J. Eastwood[4], Natalia Contreras-Ortiz[5], Carlos Reynel[6], Santiago Madriñán[5], Dmitry A. Filatov[3], and Colin E. Hughes[2]

[2]Institute of Systematic Botany, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland;
[3]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK;
[4]Millennium Seed Bank, Royal Botanic Gardens, Ardingly, RH17 6TN, UK;
[5]Laboratorio de Botánica y Sistemática, Departamento de Ciencias Biológicas, Universidad de los Andes, Apartado Aéreo 4976, Bogotá, Colombia;
[6]Departamento de Manejo Forestal, Facultad de Ciencias Forestales, Universidad Nacional Agraria La Molina, Apartado 456, La Molina, Lima, Peru.

Running head: Origins of domestication of the Andean pulse crop *tarwi*

Premise of the study: The Andean highlands are a hotspot of domestication, yet our understanding of the origins of early Andean agriculture remains fragmentary. Key questions of where, when, how many times and from what progenitors many Andean crops were domesticated remain unanswered. The Andean crop lupin, tarwi, *Lupinus mutabilis*, is a regionally important pulse crop with exceptionally high seed protein and oil content and is the focus of modern breeding efforts, but its origins remain obscure.

Methods: A large genome-wide DNA polymorphism dataset was generated using nextRADseq to infer relationships among more than 200 accessions of Andean *Lupinus* species, including 24 accessions of *L. mutabilis* and close relatives. Phylogenetic and demographic analyses were used to identify the likely progenitor of tarwi, and elucidate the area and timing of domestication in combination with archaeological evidence.

[1] Manuscript received _____; revision accepted _____.
[7] Author for correspondence (e-mail: guy.atchison@systbot.uzh.ch)

Key results: We infer that tarwi was domesticated once in northern Peru, most likely in the Cajamarca region within, or adjacent to the extant distribution of *L. piurensis* which is the most likely wild progenitor. Demographic analyses suggest that tarwi split from *L. piurensis* around 2,600 BP and suffered a classical domestication bottleneck. The earliest unequivocal archaeological evidence of domesticated tarwi seeds is from the Mantaro Valley, central Peru c. 1,800 BP.

Conclusions: A single origin of tarwi from *L. piurensis* in northern Peru provides a robust working hypothesis for the domestication of this regionally important crop, and is one of the first clearcut examples of a crop originating in the highlands of northern Peru.

Key words: Andes; domestication; Leguminosae; *Lupinus mutabilis*; nextRAD; tarwi.

The Inca and pre-Inca peoples of the Andes developed agricultural systems involving a suite of domesticated pseudo-cereal, pulse, and tuber crops, as well as livestock, dating back to at least 8,000 BP. By the late pre-Hispanic period highland Andean agriculture came to be amongst the most sophisticated in the world, reliant on intensive technologies such as irrigation and built environments such as terracing and raised fields, involving large scale landscape modification, and underpinning advanced agricultural economies that supported large populations organized into highly complex societies (National Research Council, 1989; Hernández-Bermejo and Landeón, 1994; Piperno and Pearsall, 1998; Pearsall, 1992, 2008; Larson et al., 2014). Despite this sophistication and prominence of early Andean agriculture and the importance of the Andes as a centre of crop diversity, the geotemporal trajectories of early agricultural origins and development in this region remain very incompletely understood (Pearsall, 2008). The traditional view of highland Andean agricultural origins has focused on a core set of crops and livestock including the potato (*Solanum tuberosum* L.) (Hawkes, 1990; Spooner et al., 2005), quinoa (*Chenopodium quinoa* Willd.) (Wilson, 1990; Bruno and Whitehead, 2003; Bruno, 2006; Langlie et al., 2011; Walsh et al., 2015), llama (*Lama glama* L.) and alpaca (*Vicugna pacos* L.) (Wheeler, 1995; Mengoni Goñalons, 2008; Larson and Fuller, 2014), all thought to have been domesticated in a core region in the south-central Andes in southern Peru and northern Bolivia, potentially centred in the Lago Titicaca basin (e.g. Pearsall, 1992; Morris, 1999). However, archaeological data suggest a more geographically extensive and diffuse (Pearsall, 1992, 2003, 2008), and temporally protracted history spanning 10,000 to 3,500 BP (Pearsall, 2008; Larson et al., 2014). Furthermore, the discovery of early remains of a suite of crops including cotton, squash, peanut and quinoa in north coastal Peru (Dillehay et al., 2007) suggest that agricultural economies, cultivation and extensive translocation of some of these crops into northern Peru were already in place by 8,000-5,000 BP, in line with evidence suggesting ancient cultural connections between the highlands and the lowlands to the east of the Andes and on the Pacific coast (Pearsall, 1992). There is no consensus as to whether agricultural development in the highland Andes followed geographically diffuse vs more centric, or temporally condensed vs more protracted trajectories. In the Andes, this is further complicated by the close geographical proximity of the distinctive but closely interconnected and broadly contemporaneous agricultural systems found in the highlands, at mid elevations and in the lowlands, where different sets of crops were grown (Pearsall, 1992, 2008).

Many potent, and often more precise insights into the origins of domestication in terms of where, when, how many times and from what progenitors crops evolved have been made in recent years, providing evidence about how many times agriculture arose independently, where, when and how that happened, and the geotemporal trajectories of agricultural origins and development within and among regions (Smith, 2001; Fuller et al., 2012b; Larson et al., 2014; Larson and Fuller, 2014; Zeder, 2015). However, these insights have not yet been matched in the Andes. In this region there remains a lack of detailed evidence about the precise locations and timing of domestication of even some mainstream crops, such as the potato (Hawkes, 1990; van den Berg et al., 1996; Spooner et al., 2005), quinoa (Wilson, 1990; Bruno and Whitehead, 2003; Bruno, 2006; Langlie et al., 2011; Walsh et al., 2015), or tomato (*Solanum lycopersicum* L.) (Peralta and Spooner, 2006), let alone the large cohort of lesser-known crops, the so-called '*Lost Crops of the Incas*' (National Research Council, 1989), many of which remain poorly understood, or essentially unstudied. Much more precise and complete data about the origins of individual Andean crops are needed in order to understand the trajectory of agricultural origins and development in this region (Pearsall, 2008; Larson et al., 2014).

Understanding the origins of Andean crops is beset by particular challenges. For many Andean crop genera, including wild tuber-bearing *Solanum* sect. Petota Dumort (Hawkes, 1990; van den Berg et al., 1996; Spooner et al., 2005), *Chenopodium* L. (Walsh et al., 2015) and *Oxalis* L. (e.g. Emshwiller, 2002a), basic taxonomy is lacking or extremely confused raising doubts about what species there are and where they are distributed. Furthermore, in the high elevation Andes the tempo of plant evolution and species diversification has often been high (Madriñán et al., 2013; Hughes and Atchison, 2015), such that several genera harbouring crops present a background of fast and very recent species diversification, making it harder to obtain robust resolution of species relationships using conventional DNA sequence data. Finally, for some crops, such as the potato, quinoa and oca (*Oxalis tuberosa* Molina), additional complexities of hybridization and polyploidy are apparent (e.g. Walsh et al., 2015), with potentially reticulate crop origins, set against a possible backdrop of extensive spontaneous hybridisation following cultivation (Brush et al., 1981; van den Berg et al., 1996; Hughes et al., 2007). The Andean tuber crop oca exemplifies these challenges. Despite extensive field, herbarium and molecular work (Emshwiller, 2002a, 2002b; Emshwiller and Doyle, 1998, 2002; Emshwiller et al., 2009), the on-going discovery of new wild tuber-bearing species, lack of robustly supported resolution in gene trees and a complex history of polyploidy, mean that considerable uncertainty about the origins of oca remains.

Here we investigate the origins of domestication of one of these *Lost Crops of the Incas*, tarwi (in Quechua, tauri in Aymara and chocho in Spanish), the Andean crop lupin, *Lupinus mutabilis* Sweet, an important pulse crop in subsistence agricultural systems across the central Andes (Fig. 1) (Jacobsen and Mujica, 2008). Tarwi seeds have high protein (46-48%) and oil (16-23%) contents comparable to soya bean (*Glycine max* (L.) Merr.) (Hernández-Bermejo and León, 1994; Jacobsen and Mujica, 2008), making it an excellent alternative to animal protein, which could help to drive the much needed shift away from meat-based diets (Wellesley et al., 2015), and has been an important component of highland crop rotations in the Andes contributing to soil fertility via high phosphorus uptake and nitrogen fixation. Tarwi provides another good example of the lack of understanding of Andean crop origins. Despite being a regionally important food crop grown and marketed throughout the highlands of Bolivia, Peru and Ecuador (Figs. 1 and 2; FAOSTAT, 2015), and the focus of modern crop breeding in Australia and Europe (Hardy et al., 1998; Caligari et al., 2000; Clements et al., 2008a, 2008b; Jacobsen and Mujica, 2008; Baer, 2011), key questions regarding the origin of tarwi remain unanswered (Pearsall, 2008).

Lupinus epitomises the dilemmas and challenges surrounding the understanding of the origins of Andean crops: the taxonomy of Andean *Lupinus* remains in a state of considerable chaos resulting from prolific over-description of species by Smith (1938-1953), with an estimated 85-100 distinct species, but more than 500 species names, and the last 'complete' taxonomic account was that of Agardh (1835). The monophyly of the Andean taxa, with the exception of *L. microphyllus* from Chile, has been established (Ainouche and Bayer, 1999; Hughes and Eastwood, 2006). However, species diversification has occurred recently and extremely rapidly making it challenging to obtain robustly supported hypotheses of species relationships using conventional DNA sequence datasets or morphological data (Hughes and Eastwood, 2006; Drummond et al., 2012; Hughes and Atchison, 2015). Indeed, the net species diversification rate estimated for the Andean *Lupinus* radiation is amongst the fastest documented for plants (Drummond et al., 2012). This recency and rapidity of diversification means that all previous phylogenies of *Lupinus* lack resolution among the Andean species. These issues have so far hindered progress towards understanding where, when, how many times and from what progenitors *L. mutabilis* was domesticated (Eastwood and Hughes, 2008). To combat these difficulties, here we have assembled new genome-wide DNA sequence data using the recently developed nextRAD technology to obtain, for the first time, robust resolution among these very recently diverged species, and to assess the demographic history of domestication.

Previous hypotheses of domestication of *L. mutabilis*, focusing on either a hybrid origin involving two North American species, *L. douglasii* and *L. ornatus* (Kazimierski and Nowacki, 1961), or from apparently unrelated Andean species (Blanco, 1986), such as *L. praestabilis* (Tapia and Vargas, 1982), were dismissed by Eastwood and Hughes (2008) as lacking support. Available phylogenetic evidence clearly shows that *L. mutabilis* is nested within a large clade of Andean species (Hughes and Eastwood, 2006; Drummond et al., 2012), ruling out involvement of non-Andean species in the origin of domesticated *L. mutabilis*, and demonstrating unequivocally that lupins were domesticated independently in the New and Old Worlds (Eastwood and Hughes, 2008). There is also no evidence that polyploidy played a role in the domestication of *L. mutabilis*, with almost all the western New World *Lupinus* species, and all documented Andean species, including *L. mutabilis*, having a uniform chromosome number, 2n=48 (Conterato and Schifino-Wittmann, 2006). *Lupinus mutabilis* is only known in cultivation. Despite sporadic reports to the contrary (Tapia and Vargas, 1982; Blanco, 1986), no apparently wild populations have so far been discovered during extensive fieldwork by the authors throughout the Andes. Based on morphological and preliminary genetic data, Eastwood and Hughes (2008) suggested four species as possible close relatives of *L. mutabilis*: *L. ellsworthianus* C.P.Sm., *L. piurensis* C.P.Sm., *L. praestabilis* C.P.Sm. and *L. semperflorens* Hartweg ex Benth. Of these, *L. piurensis* was hypothesized as the most likely progenitor due to its close morphological similarity to *L. mutabilis*: these two species share notably glaucous stems (especially the young shoots), similar overall leaf morphologies, notably thin leaflets with glabrous adaxial surfaces, large flowers on relatively lax erect inflorescences held above the foliage, and closely matching flower colour patterns, the main differences between the two being in pod size, dehiscence, seed size and seed coat thickness (Fig. 3). Tarwi possesses a suite of pod and fruit traits (Figs. 1 and 3), including non-shattering (indehiscent) pods, large seeds, permeable seed coats and reduced seed dormancy, pale seed colour, as well as a shift from perennial to near-annual habit, a set of traits typical of the classical legume domestication syndrome (Hancock, 2012; Weeden, 2007). However, in general, seed size remains the only reliable indicator of legume crop domestication used to assess archaeobotanical remains. In contrast to early Old World and modern lupin domesticates, seed alkaloid levels remain high in domesticated *L. mutabilis*, and intensive processing is needed to de-bitter seeds prior to consumption.

Domestication can also provide important evolutionary insights into the tempo and trajectories of phenotypic and genetic change, the underlying genetic basis of morphological traits, and the links between

154 the two, both within the domesticate (Fuller et al., 2014) and the domesticator (Mathieson et al., 2015).
155 Lupins provide an especially interesting group for studying domestication because unrelated wild species
156 were domesticated independently within the last 3,000 years in both the Andes (*L. mutabilis*) and the
157 Mediterranean (*Lupinus albus* L. and *Lupinus luteus* L.), potentially providing an example of convergent
158 congeneric crop evolution (Cowling et al., 1998). Additionally, a third Mediterranean species, *Lupinus*
159 *angustifolius* L., which was domesticated much later within the twentieth century, has gone on to become an
160 important modern agricultural crop (Gladstones et al., 1998), and is the focus of much recent and on-going
161 research, including genome sequencing efforts (Yang et al., 2013; Kamphuis et al., 2014). Multiple parallel
162 domestication events provide exciting opportunities to compare phenotypic and genetic changes associated
163 with domestication among congeners in different locations and cultural contexts, once their progenitors have
164 been ascertained (Schmutz et al., 2014). All domesticated lupins exhibit a similar syndrome of
165 morphological changes associated with domestication prompting interesting questions as to whether the
166 independent domestications in the New and Old World are convergent, i.e. originating from very different
167 origins but following identical genetic pathways to reach the same phenotype, or represent parallelism,
168 whereby similar, but not identical genetic pathways were involved (Fuller et al., 2014).
169 Finally, a better understanding of the origins of crops provides the basis for assessment, collection, use
170 and conservation of crop genetic resources (Hoisington et al., 1999; Jarvis et al., 2008). Such data are
171 urgently needed for Andean crops where erosion of valuable crop diversity is apparent both of varieties of
172 mainstream crops and lesser-known crops. This is especially the case for the large suite of *Lost Crops of the*
173 *Incas*, for which on-farm, *circa situm* genetic conservation has been particularly important (Hernández-
174 Bermejo and León, 1994; Brush, 1999). Given the interest in crop improvement of *L. mutabilis,* identifying
175 its close cross-compatible wild relatives is vital as these may allow transfer of important traits. Many
176 regionally important but lesser-known Andean crops are being displaced and marginalized by introduced
177 crops (e.g. faba beans, *Vicia faba* L., displacing tarwi, or wheat, *Triticum aestivum* L., displacing quinoa)
178 (National Research Council, 1989; Hernández-Bermejo and León, 1994), with tarwi as one of the crops most
179 affected by introduced European crops (Hernández-Bermejo and León, 1994). At the same time many of
180 these crops, including tarwi, show great potential for maintaining the resilience of indigenous Andean
181 agricultural systems, and in providing excellent nutritional benefits. To unlock the potential of tarwi and

other Andean crops we need to understand the past processes of human manipulation and evolution of domestication traits (Zeder, 2015).

In this study we address questions about where, when, how many times and from what progenitor tarwi was domesticated and assess the implications of our results for understanding the origins and development of early agriculture in the Andes. To do this, we have generated a large genome-wide DNA sequence dataset of over 100,000 loci for more than 200 accessions of Andean *Lupinus*, including 24 accessions representing *L. mutabilis* and its putative close relatives using novel nextRAD sequencing technology (Emerson et al., 2015; Russello et al., 2015). We also review and summarize the available archaeological data for tarwi, and integrate and synthesize these complementary sources of genetic and archaeobotanical evidence.

192 MATERIALS AND METHODS

193

194 ***Sampling –*** For genetic sequencing and phylogenetic inference we generated nextRAD sequence data
195 for 212 accessions of Andean *Lupinus*, representing 63 species. Where possible multiple accessions of
196 species were sampled representing potential genetic diversity within species across their geographical and
197 morphological ranges. This sampling includes 74% of known taxonomic entities of Andean *Lupinus* and
198 spans the entire geographic range of the clade from Colombia to Argentina, with the exception of a small
199 cohort of species endemic to the Venezuelan Andes. This Venezuelan gap is not considered important as *L.*
200 *mutabilis* is very rarely found in Venezuela and there is no evidence that it could have been domesticated
201 there. Material for DNA extraction was sampled from silica-dried leaf material, fresh tissue from living
202 plants raised from seed and, in some cases, leaf material sampled from herbarium vouchers (Appendix S1)
203 (see Supplemental Data with the online version of this article).

204 ***NextRADseq preparation and Sequencing –*** Total genomic DNA was extracted from sampled
205 material using either a CTAB DNA isolation protocol (modified from Doyle and Doyle, 1987) or a Qiagen
206 DNeasy kit (Qiagen, Hombrechtikon, Switzerland) according to manufacturer's guidelines. A Quibit
207 Fluorometer (ThermoFisher, Dietikon, Switzerland) was used to assess DNA quantity and gel
208 electrophoresis was used to measure quality and purity. Library preparation and sequencing of nextRAD
209 markers from genomic DNAs were performed by SNPsaurus (SNPsaurus LLC, Oregon, USA).

210 To amplify genomic loci consistently between samples the nextRAD method (Emerson et al., 2015;
211 Russello et al., 2015) uses selective PCR primers. Genomic DNA was first fragmented with Nextera reagent
212 (Illumina Inc., San Diego, California, USA), which also ligates short adapter sequences to the ends of the
213 fragments. Variable amounts of genomic DNA were used as input into the fragmentation reaction to adjust
214 for quality of DNA, with more input DNA for more degraded extracts. Fragmented DNA was then
215 amplified, with one of the primers matching the adapter and extending seven nucleotides into the genomic
216 DNA with the selective sequence (TGCAGAG). Thus, only fragments starting with a sequence that can be
217 hybridized by the selective sequence of the primer will be efficiently amplified. The resulting fragments are
218 fixed at the selective end, and have random lengths depending on the initial Nextera fragmentation. Because
219 of this, amplified DNA from a particular locus is present at many different sizes and careful size selection of

220 the library is not usually needed. Libraries were sequenced on an Illumina NextSeq 500 at the University of
221 Oregon sequencing centre to generate 150 bp single end reads.
222
223 ***RADseq data assembly*** **–** Different approaches for the assembly of the RADseq data were used for the
224 phylogenetic analysis and the population demographic analysis (the latter described below), reflecting the
225 very different phylogenetic breadths of these two analyses, spanning many species across the whole Andean
226 clade for the phylogeny, but just two sister species for the demographic analysis. For phylogenetic analysis,
227 de novo assembly using PyRAD (Eaton, 2014) based on an overall similarity criterion after alignment was
228 used. This allows incorporation of indel variation which is expected when more distantly related taxa are
229 included, and recovers more shared loci across disparate taxa than alternative approaches using Stacks
230 (Eaton, 2014; Pante et al., 2015; Ree and Hipp, 2015), which are needed to obtain robust phylogenetic
231 resolution, in line with recent phylogenetic analyses using RADseq data (Eaton and Ree, 2013; Takahashi et
232 al., 2014; Cavendar-Bares et al., 2015; Massatti et al., 2016). Prior to assembly, raw reads were processed
233 through Trimmomatic v0.33 (Bolger et al., 2014) to remove bases at the ends of reads with a quality score
234 less than 15. Reads were assembled *de-novo* into loci using pyRAD v.3.0.5 (Eaton, 2014). We ran
235 preliminary tests (results not shown) in pyRAD in order to determine the best parameters to use for our
236 dataset. Quality filtering of reads in pyRAD at step 2 converted bases with a quality score of less than 20 into
237 Ns and reads with more than 4 Ns were discarded. Within-sample clustering (step 3) was carried out at 0.85
238 which is the most suitable threshold for the phylogenetic distances within our dataset (see also Takahashi et
239 al., 2014). Consensus sequences were created for each cluster using error-rate and heterozygosity estimates
240 calculated from the base counts at each site across all clusters. Clusters with a minimum read coverage of
241 less than five were discarded. Consensus loci were clustered between samples at the same threshold as used
242 in step 3. All other parameters were set at default values.
243 To infer phylogenetic relationships across the Andean clade we generated a concatenated matrix of
244 loci for all 212 accessions (hereafter 212acc) using step 7 in pyRAD with a minimum sample coverage (the
245 minimum number of samples which must have data for a given locus to be included in the final matrix)
246 setting of 5, in line with other studies (Takahashi et al., 2014). In addition, we applied the maximum number
247 of SNPs filter in step 7 to include only loci with 60 or fewer SNPs a somewhat stricter cutoff than the default
248 PyRAD setting of 100 SNPs per locus which has been widely used (Eaton and Ree, 2013; Takahashi et al.,

2014; Cavendar-Bares et al., 2015; Massatti et al., 2016). Both left and right overhangs were trimmed on final loci.

We generated a second concatenated matrix containing 24 accessions (hereafter 24acc) focusing on a robustly supported subclade containing *L. mutabilis*, its putative progenitor, *L. piurensus*, and a few other closely related taxa, found in the initial ExaML analysis of the 212acc matrix. For this second 24acc matrix, the minimum sample coverage filter was set at 10. To investigate genetic structure between, and within, *L. mutabilis* and *L. piurensis*, we generated an unlinked SNP matrix (one SNP sampled at random per locus) (hereafter 20accUS) for all 20 samples of these two taxa with a minimum sample coverage setting of 10 to minimise the level of missing data across individuals. Analyses were carried out on the Science Cloud computational facility of the University of Zurich.

Phylogeny **–** Due to the computational challenges posed by the large size of the 212acc matrix (7,389,726 base pairs) we used ExaML v.3 (Kozlov et al., 2015), following a modified version of the approach used by Jarvis et al. (2014) and Misof et al. (2014) to generate an initial species phylogeny for the Andean clade as a whole. We generated 100 randomized stepwise addition parsimony trees under standard RAxML v.8 (Stamatakis, 2014) to use as input starting trees for a fast preliminary run of ExaML. From this we selected the ten best scoring trees and used the original parsimony starting trees as input for ten full ExaML runs. We considered the tree with the best overall log-likelihood score of all ten tree inferences to be the ML tree. We generated 150 non-parametric bootstrap replicates in ExaML as described in the ExaML manual. We obtained estimates of ML bootstrap support by computing a consensus tree from the bootstrap replicates using the respective standard RAxML options, and by drawing bootstrap values onto the best ML tree.

To infer the most robust possible hypothesis of relationships within and between accessions of *L. mutabilis* and *L. piurensis* we analyzed the 24acc matrix using RAxML to perform a bootstrap analysis and best-scoring ML tree search in a single analysis run. All trees were visualised in FigTree (Rambaut, 2009) with midpoint rooting used for the 212acc tree, and the *L. semperflorens* /*L. sp nov* (Celendín large flowers) subclade as an outgroup for rooting the 24acc tree. Midpoint rooting is justified as preliminary analyses of an even larger RAD data matrix including samples from across the whole western New World *Lupinus* clade show the monophyly of the Andes, with a set of Mexican lineages as sister to the Andean clade (results not shown), as found previously by Drummond et al. (2012).

Genetic Structure **–** To investigate patterns of genetic structure within and between *L. mutabilis* and *L. piurensis* we used Structure v2.3.4 (Pritchard et al., 2000). We used the 20accUS matrix with 20 individuals and 4,141 SNPs with no a priori grouping assignments. Ten replicates at each K value between 1 and 5 were run with 50,000 generations of 'burn-in' and 100,000 generations of sampling. Results were assessed using Structure Harvester (Earl and vonHoldt, 2012).

Demographic history of domestication **–** For demographic inference of the population history of *L. mutabilis* and *L. piurensis,* to estimate the approximate timing of domestication, and the parameters of the genetic bottleneck associated with domestication of *L. mutabilis,* we used isolation with migration models implemented in the *dadi* package (Gutenkunst et al., 2009). This method compares the observed Site-Frequency-Spectrum (SFS) to that expected given a specific population split demographic model in order to estimate model parameters and assess the likelihood of the observed data given the model. As this method relies on the analysis of SFS it is sensitive to genotype calling errors. Furthermore, using a polarised SFS (where the ancestral allele at each site is known) significantly increases the power of this method to distinguish alternative demographic scenarios. To deal with these issues, we employed a different bioinformatics approach from that used in the phylogenetic analysis to obtain SNP data for this analysis. We mapped the RADseq sequence reads for 20 *L. piurensis* / *L. mutabilis* accessions to the published scaffolds of the draft *L. angustifolius* genome (Yang et al., 2013; Assembly GCA_000338175 available from GenBank) using bwa (Li and Durbin, 2010) with the mem algorithm and default values. We excluded reads with mapping quality below 20, and used the package *Stacks* (Catchen et al., 2011, 2013) to extract only RAD loci present in both species and having at most 2 SNPs. As our analyses assume that polymorphisms in the dataset are unlinked, a single random SNP from each of these loci, with the ancestral state of polymorphisms inferred from comparisons with the genome of *L. angustifolius*, was used for subsequent analysis in *dadi*. Following the recommendations in the *dadi* manual, the sample sizes for each population were 'projected down' to include 10 alleles per population to account for missing data while maximizing the number of SNPs analysed. Visualisation of the two-dimensional site frequency spectra (2D-SFS) was also done with the *dadi* package.

The analyses were conducted with a relatively parameter-rich model, *IM*, that allows for an unequal population size split followed by exponential growth in each species and differential migration rates in both directions. The model has six free parameters: the relative size of the two populations after they split (*s*),

time of the species split (T_s), extant modern population sizes of species 1 and 2 (N_1 and N_2), and effective migration rates in two directions ($M_{1<=2}$ and $M_{2<=1}$). Fixing specific values for some of these parameters allows testing demographic hypotheses using likelihood ratio tests. For example, by fixing migration rates at zero, we tested whether gene flow between species following their split improved model fit to the data. To ensure the analysis reached the global maximum we performed 50 runs with perturbed starting parameters (*perturb_params* function in *dadi*) and very wide search bounds for all parameters, followed by 50 runs with perturbed starting parameters and narrower ranges. We analysed the likelihood profiles for all parameters and selected parameters from the highest likelihood run as the optimal parameter set. To evaluate the robustness of the parameter estimates, we obtained confidence intervals using the Fisher Information Matrix approach (*FIM_uncert* function in *dadi*). This approach does not account for linkage between SNPs, but given that we use a single SNP per locus and relatively sparse genome-wide sampling of SNPs, it should provide reliable confidence intervals at much lower computational cost than parametric bootstrapping. The observed and modelled 2D-SFSs were plotted using the plot_2d_comp_multinom command in the *dadi* package. Comparison of models with and without migration was performed using non-parametric bootstraps and adjusted LRT (Coffman et al., 2015).

Morphological trait data – Measurements of weight, length, width and depth of a sample of 25 seeds per accession were taken. Additional seed weight data were retrieved from the Royal Botanic Gardens Kew Seed Information Database (SID) (2016) and incorporated into the dataset used in this study.

Archaeological evidence – Archaeological records of *Lupinus* remains from across the Andean region were assembled and assessed from the literature (Appendix 1).

RESULTS

NextRAD data assembly **–** The average number of sequence reads per sample was 2,830,091 with a total of ~600 Million bp of sequence generated. The total number of loci recovered in the 212acc and 24acc matrices after filtering was 140,608 and 24,942 respectively. The numbers of parsimony informative sites (pis) and variable sites (vs) in the 212acc matrix were 723,042 and 2,350,809 and in the 24acc matrix there were 56,492 (pis) and 213,561 (vs). The raw sequences are archived in the NCBI sequence read archive (Appendix S1).

Phylogeny **–** The full 212acc phylogeny including all 63 Andean species (Figs. 4A and Appendix S2) comprises two well-supported (100% BS) and geographically structured clades. While there is some geographical overlap between these two clades, the northern Andean clade mainly contains accessions from Ecuador and Colombia, and the south-central Andean clade mainly accessions from Peru and Bolivia. These two clades are also resolved in wider phylogenies using even larger RADseq datasets spanning the whole western New World clade which confirm the monophyly of the Andes (results not shown). All *L. mutabilis* and *L. piurensis* accessions are placed within the south-central Andean clade and together these two species form a robustly supported clade with 99% bootstrap support sister to a clade that contains all accessions of *L. semperflorens* and one accession of *L. sp nov* (here denoted as '*Celendín large flowers*'). Within the *L. mutabilis* + *L. piurensis* subclade, all accessions of *L. mutabilis* form a well-supported (99% BS) clade which is nested within *L. piurensis* (Appendix S2).

The relationships recovered within the 24acc phylogeny (Fig. 4B) mirror those found in the larger phylogeny. Again *L. mutabilis* and *L. piurensis* form a well-supported clade (100% BS) and all accessions of *L. mutabilis* form a well-supported clade (100% BS). In both phylogenies accession CEH1998 of *L. piurensis* from San Juan in the Department of Cajamarca, northern Peru, is weakly supported as sister to the *L. mutabilis* clade.

Within *L. piurensis* there are two well-supported (100% BS) pairs of accessions, and one well-supported (99% BS) subclade. Within the *L. mutabilis* clade the internal support is variable with occasional moderate support.

Genetic Structure **–** Bayesian clustering of samples revealed that the best number of clusters, based on ΔK criterion (Evanno et al., 2005), was K = 4 (Fig. 4C). To assess the genetic relationships and levels of

admixture between *L. piurensis* and *L. mutabilis* plots of K=2 and K=3 are also presented (Fig. 4C). At K=2 there is a clear split into two groups corresponding to the two species, although two accessions of *L. piurensis* (RJE110 and CEH1998), both from San Juan, Cajamarca, Peru, show partial admixture with 20 to 25% of their genetic material being assigned to the *L. mutabilis* group. At K=3 two accessions of *L. piurensis* (RJE122 and RJE134), both from Piura in the extreme north of Peru, group together and show no admixture. All accessions of *L. mutabilis* show 75% affinity or greater to a single genetic group. The remaining accessions of *L. piurensis* show varying levels of admixture. At K=4 the structure is similar to K=3 with the same two accessions of *L. piurensis* (RJE122 and RJE134) forming a pair, each with no admixture and the structure for *L. mutabilis* is also repeated. The pattern within *L. piurensis* increases in complexity with three accessions showing admixture into a fourth group.

Demographic history of domestication **–** Our stringent filtering of RAD loci and SNPs resulted in 1,610 SNPs present in a sample that was 'projected down' to 10 alleles from each species that were used for demographic analysis in *dadi*. The parameter rich *dadi* model used here (*IM*) included six free parameters and allowed for population size change after divergence, as well as differential bi-directional migration rates (Fig. 5A). This model fitted the data well, except for a slight excess of low- and high-frequency polymorphisms in the observed frequency spectra for both species, compared to that predicted by the model (Fig. 5B). Using likelihood ratio tests on the nested models, the analyses revealed that a simpler species split model with no migration after splitting fits the data equally well (Table 1). The ML estimates of parameters of this model (Table 1) show that *L. mutabilis* was domesticated from a small subset of the ancestral population (5% of ancestral population size, N_a), and subsequently experienced a 20-fold expansion in population size such that current population size roughly matches the ancestral size. Conversely, population size in the wild relative *L. piurensis* remained relatively constant, with an increase of only about 30% relative to the ancestral population size. The estimated species split/domestication time ($T = 0.239 \pm 0.031$) is expressed in units of $2N_a$ generations, where N_a is ancestral population size. To convert this to years one needs to assume plausible values of N_a and the number of generations per year. The estimated average genome-wide nucleotide diversity per site for *L. piurensis* across all loci reported by *stacks* with a maximum of 2 SNPs and with or without variation, was 0.000549, which translates into an effective population size of ca. 5,490, assuming a generic 2.5×10^{-8} mutation rate per generation per site. Given this population size, the domestication occurred (0.239 * 2 * 5.500 = 2,624) ca. 2,600 (sd = 340) generations ago. If we assume one

385 generation per year, which is a reasonable assumption given that *L. piurensis*, although a perennial, flowers
386 and fruits within the first year in cultivation and in the wild, then domestication occurred 2,600 BP.

387 ***Seed Weights*** **–** The graph of seed weights (Fig. 6) shows *L. mutabilis* to have much heavier seeds (4-
388 8 g per 25 seeds) than any of the other measured Andean lupin species (all < 2 g per 25 seeds) (Appendix
389 S3). The next heaviest seeds belong to *L. semperflorens*. The seed weight of *L. piurensis* seeds is higher than
390 the majority of wild Andean species measured (and further sampling of seed size variation across the range
391 of *L. piurensis* might reveal even larger seeds), most of which have very small seeds, suggesting that *L.*
392 *piurensis* could have been among a small set of wild species potentially pre-disposed to domestication.

393 ***Archaeological Record*** **–** Seeds of *Lupinus* are generally infrequent and elusive in the archaeological
394 record, but have been sporadically recovered from a small number of archaeological sites scattered across
395 central Peru and at two sites in northern Ecuador (Fig. 2; Appendix S4). Most of these, and all the early
396 seeds, are reported to be of wild lupins, with the first unambiguous and convincing evidence of large and
397 clearly domesticated *L. mutabilis* seeds dating to 1,800 BP from the Mantaro Valley in central Peru. This
398 provides a minimum age estimate for when large-seeded domesticated tarwi was cultivated and suggests that
399 domestication occurred sometime between 1,800 BP and 3,450 BP (Fig. 2). In addition to the archaeological
400 records there are a number of historical references documenting cultivation of *L. mutabilis* in the form of
401 requests to tax lupin crops following the Spanish occupation (references in Antúnez de Mayolo, 1982; Gross,
402 1982), prompting Gross (1982) to suggest that an estimated 10,000 hectares of *L. mutabilis* was being grown
403 in Peru ca. 450 BP.

404 DISCUSSION

405

406 The transition from hunter gathering to farming brought about via domestication of crops and
407 livestock was a pivotal turning point in human history with far-reaching impacts and consequences for the
408 emergence of civilisations and societal development (Diamond 2002; Fuller 2010; Larson et al., 2014). This
409 was very much the case in the Andes, where early agriculture led to the emergence of sophisticated pre-Inca
410 and Inca civilisations extending at their peak from Argentina to Colombia. However, in contrast to several
411 other regions of early agriculture, understanding the origins of domestication in terms of where, when, how
412 many times and from what progenitors Andean crops evolved has lagged behind and remains fragmentary,
413 limiting our understanding of agricultural origins and development in this region.

414 Here we present evidence suggesting that the regionally important Andean pulse crop, tarwi, *L.*
415 *mutabilis*, was domesticated, not in the putative south-central Andean core area of early agriculture, but
416 instead much further north in the highlands of northern Peru, most likely in the department of Cajamarca.
417 We show that all sampled accessions of *L. mutabilis* form a monophyletic group that is nested within *L.*
418 *piurensis*, a wild lupin species endemic to the western flanks of the Andes from central Cajamarca north
419 through Piura to the extreme south of Ecuador in Loja (Figs. 2 and 4A and Appendix S2), between 1650 and
420 3300 m elevation. These results strongly support the hypothesis of Eastwood and Hughes (2008) based on
421 morphology, that *L. piurensis* is the most likely living descendant of the wild progenitor of tarwi, and that
422 tarwi was domesticated once, within or adjacent to the extant distribution of *L. piurensis*. With reasonably
423 dense taxon sampling including 75% of Andean *Lupinus* species, the robust support for the *L. mutabilis* and
424 *L. piurensis* relationship (Fig. 4), plus evidence of strong morphological similarities between these two taxa
425 (Fig. 3) based on extensive field and herbarium work spanning Andean *Lupinus* as a whole (Hughes et al. in
426 prep.), taken together provide compelling evidence that *L. piurensis* is the likely progenitor of tarwi. While
427 the domestication gap in seed size between the largest-seeded wild species and domesticated *L. mutabilis*
428 seeds is large (Fig. 6), seeds of *L. piurensis* are among the largest of the Andean wild lupin species, a
429 characteristic that could have pre-disposed *L. piurensis* to domestication, as suggested for Fertile Crescent
430 cereals, but less prevalently for Fertile Crescent pulses (Preece et al., 2015).

431 The placement of the *L. piurensis* accession CEH1998 from San Juan, Cajamarca as sister to the *L.*
432 *mutabilis* clade, albeit with weak support (Fig. 4B), and evidence of genetic admixture between the two

433 species in that accession (Fig. 4C), alongside the more distant relationship of the two accessions of *L.*
434 *piurensis* sampled from further north in Piura (RJE122 and RJE134) which show no admixture with *L.*
435 *mutabilis* samples at K=2 (Fig. 4C) and form a separate genetic cluster at K=3 and K=4, strongly point to
436 Cajamarca as the most likely origin of *L. mutabilis*. In addition, the phylogeographic relationships within *L.*
437 *mutabilis*, although not all robustly supported, suggest that the first-branching lineages of *L. mutabilis* are
438 from central-north Peru (Cajamarca, Ancash, Huánuco), with later nested expansions south to Bolivia and
439 north to Ecuador and Colombia, also pointing to a north-central Peruvian origin of tarwi, with later spread to
440 the north and south. Of course, the sheer geographical extent, remoteness, and physiographic heterogeneity
441 of the Andes, and the very sparse and fragmentary state of botanical collections in some parts of the north-
442 central Andes (Young et al., 2002), mean that it is entirely possible, perhaps even likely, that novel variants
443 related to tarwi and/or additional close wild relatives of *L. mutabilis* that could alter this picture, still remain
444 to be discovered, as has been the case for many other prominent New World crops in recent decades (Iltis,
445 1998; Emshwiller et al., 2009; Taba et al., 2011). Pending such discoveries, a single origin of tarwi from *L.*
446 *piurensis* in northern Peru provides a robust working hypothesis.

447 The sparse occurrences of *Lupinus* seeds in the archaeological record are in part a reflection of the
448 lack of archaeological sites in the Andean highlands (Pearsall, 1992), and the fact that lupins have non-
449 diagnostic starch grains and so their dietary contribution cannot be studied in that way (Piperno and
450 Dillehay, 2008), but it is almost certainly also indicative of its later domestication. Early records are
451 infrequent and comprise small quantities of small seeds comparable in size to extant wild species, which are
452 ubiquitous throughout the high elevation Andes. To what extent these small wild lupin seeds in the
453 archaeological record represent gathered wild food resources or accidental incursions is not clear. We have
454 made sporadic observations of local small-scale cultivation of several other Andean species of *Lupinus*,
455 including *L. semperflorens* (which is notable as the species with the largest seeds of any wild Andean lupin –
456 Fig. 6), and an as yet unidentified annual or weakly biennial narrowly-restricted endemic species locally
457 called '*chuguri*' which is highly valued, sown and encouraged as a green manure crop in fallow fields. Both
458 of these entities are found in extensive populations in different parts of the Cajamarca valley (close to the
459 hypothesized origin of tarwi), but it is not clear whether they were cultivated and used as food plants on any
460 scale in the past. This apparent confluence of lupin cultivation, involving several wild species and the
461 putative origin of tarwi in and around Cajamarca, may be further evidence for the importance of this region

462 for early lupin use in the Andes. The earliest unequivocal archaeological evidence of domesticated tarwi
463 seeds c. 1,800 BP from the Mantaro Valley, Junin in central Peru, lies less than 500 km south of Cajamarca,
464 still well to the north of the putative core south-central Andean region where other crops are suggested to
465 have been first domesticated, and in line with genetic data pointing to Cajamarca as the putative area of first
466 domestication. It is also notable that no *Lupinus* seed remains have apparently been recovered from
467 archaeological sites further south in and around the Lago Titicaca basin, perhaps suggesting that tarwi was
468 only introduced to southern Peru and Bolivia later, congruent with the derived placements of Bolivian
469 accessions of *L. mutabilis* in the phylogeny (Fig. 4).

470 These results are very much in line with the history of domestication of tarwi as inferred from the
471 demographic analyses. These analyses suggest that tarwi split from its progenitor *L. piurensis* around 2,600
472 BP, was derived from an initial small subset of the ancestral population, and suffered a classical
473 domestication bottleneck with subsequent rapid population expansion as tarwi became widely cultivated
474 across the Andes. The close coincidence in estimates of the timing of domestication derived from the
475 archaeological record (predating 1,800 BP) and the demographic analysis (2,600 BP) is striking. The first
476 evidence of domesticated tarwi also significantly post-dates the estimated timing of domestication of most
477 other Andean crop and livestock domesticates (Pearsall, 2008; Larson et al., 2014). This later origin, well to
478 the north, suggests that tarwi was most likely a later secondary domesticate, added into the existing original
479 Andean crop and livestock assemblage as agriculture expanded northwards. The demographic analysis found
480 no evidence for gene flow between *L. piurensis* and *L. mutabilis* after domestication which may be the result
481 of conscious selection for desirable traits or unconsciously, or simply due to its cultivation predominantly
482 outwith the relatively narrow geographic range of *L. piurensis* (Hancock, 2012). As far as we can ascertain,
483 this is one of the first clearcut examples of a crop originating in the highlands of northern Peru. It has been
484 suggested that lima bean (*Phaseolus lunatus* L.) and coca (*Erythroxylum coca* Lam.) could have had their
485 origins in the north-central Andes (Pearsall, 2008), with lima bean potentially domesticated on the west side
486 of the Andes of Ecuador and northern Peru, although hard evidence remains sparse and imprecise (Gepts,
487 1996), and coca postulated to have been first domesticated on the moist eastern mid-elevation slopes of the
488 Andes (Plowman, 1984). Furthermore, although the boundaries between mid-elevation and highland
489 agricultural systems are far from clearcut and were interconnected from early times, both lima bean and coca

490 are essentially mid-elevation, rather than true highland crops like tarwi, which is found up to 4,000m
491 elevation (Fig. 1).

492 Our results lend weight to the idea of a diffuse and temporally extended regional mosaic of
493 agricultural development and crop domestication in the Andean highlands, where processes of domestication
494 were broadly parallel, but not simultaneous for all crops, nor spatially concentrated into one core area
495 (Pearsall, 1992, 2008). This is very much in line with general recent thinking breaking away from centric,
496 'core area' hypotheses and interpretations in favour of more complex, protracted, multi-faceted regional
497 explanations and scenarios for the development of early agriculture (Fuller, 2010; Fuller et al., 2012a, 2012b;
498 Langlie et al., 2014).

499 It is notable that early domestication of lupins in the Mediterranean followed a similar trajectory to *L.*
500 *mutabilis* in the Andes, in that *L. albus* was apparently cultivated and domesticated only well after
501 agriculture, based on core Fertile Crescent crops (Zeder, 2008), spread westwards from the Middle East into
502 the Mediterranean (Deflorin et al. unpubl. data). Thus the independent domestications of wild lupins in the
503 Mediterranean and the Andes, both apparently represent late-stage secondary domestications, added to
504 existing crop assemblages several millennia after the establishment of agriculture, in areas distant from the
505 core areas of first agriculture and initial domestication in these regions, rather like *Phaseolus* L. beans in
506 Mesoamerica, which were apparently domesticated long after maize-based agriculture first started, and to the
507 west of the earliest maize and squash domestication (Smith, 2001). In this sense the independent
508 domestication of lupins in the Old and New Worlds apparently followed similar trajectories. They also
509 involved similar phenotypic changes to non-shattering pods, permeable seed coats and large seeds. However,
510 important differences seem to be that domestication of *L. mutabilis* involved a more than two-fold increase
511 in seed size and weight (Fig. 6), while the gap in seed size between wild and domesticated lupins in the
512 Mediterranean is apparently less pronounced, and the persistence of high alkaloid levels in domesticated *L.*
513 *mutabilis* contrasting with sweet, low-alkaloid Old World lupin domesticates.

514 ***RADseq data in domestication studies*** – This is one of the first studies of the origins of crop
515 domestication to use RADseq data, and specifically the novel nextRAD approach (Emerson et al., 2015;
516 Russello et al., 2015). RADseq data are potentially ideal for investigating domestication, because sequences
517 are obtained for many 1000s of loci distributed across the genome. Previous efforts to address the core
518 questions of where, when, how many times and from what progenitor tarwi was domesticated (Eastwood and

519 Hughes, 2008), were hampered by lack of phylogenetic resolution due to a lack of variation in the genetic
520 loci sequenced, even with 9kb of plastid and nuclear sequence data (Drummond et al., 2012). Our results
521 demonstrate the utility of RADseq data to generate large informative genome-wide DNA sequence datasets
522 that can resolve relationships among species which diverged as recently as a few thousand years ago, and
523 hence shed light on origins of domestication. However, it is notable that obtaining resolution among the
524 recently diverged species of Andean *Lupinus* depends on using a very large genome-wide dataset by
525 maximising the number of shared informative loci and SNPs. Similarly high data size thresholds in terms of
526 numbers of RADseq loci were also required to resolve other recent species radiations (Wagner et al., 2013;
527 Martin and Feinstein, 2014). Maximising the number of shared informative loci to obtain robust resolution
528 likely introduces SNP calling errors, consistent with the high proportion of singleton (uninformative) SNPs
529 and the inflated terminal branch lengths (Appendix S2), but such errors should not unduly affect tree
530 topology. Despite the computational challenges associated with the very large data matrix, we analyzed full
531 RAD sequences in a large concatenated alignment, rather than SNPs alone. This is desirable because it has
532 been shown that this improves accuracy of topologies and branch lengths (Leaché et al., 2015).
533
534
535 CONCLUSIONS
536
537 As described by Zeder (2015), understanding domestication is critical for discovery and development
538 of alternative crops, for on-going crop breeding and development, and for overall global food security. The
539 identification of *L. piurensis* as the most likely progenitor of *L. mutabilis* has important implications for
540 lupin breeders as a potential source of genetic variation, as well as providing important pointers for crop
541 genetic resource conservation. The early Hispanic tax records suggest that tarwi was grown on a much larger
542 scale in the past, with an estimated 10,000 hectares around 450 BP, a greater area than modern estimates
543 which suggest that there may have been as few as 5,200 hectares under tarwi cultivated in the 1980s (Tapia
544 et al., 1988). This is consistent with the idea that tarwi has declined and become marginalized with the
545 introduction of Old World crops (e.g. broad bean, *Vicia faba*). Our results suggest that the lupin genetic
546 resources of the Cajamarca region will be especially important to protect via a combination of *ex situ* and

547 *circa situm* farm-based conservation, if we are indeed going to realize the potential of this particular '*lost*
548 *crop of the Incas*'.
549

ACKNOWLEDGEMENTS

This work was supported by funds from the Swiss National Science Foundation (Grant 31003A_135522 to C.E.H.), and the Natural Environment Research Council UK (grant NE/K004352/1 to D.A.F.), as well as from the Claraz-Schenkung Foundation, Switzerland, Stanley Smith Horticultural Trust and the Royal Society, UK for fieldwork in the Andes. The authors thank the ScienceCloud service of S3IT, University of Zurich for computational support, Paul Etter and Eric Johnson of SNPsaurus, Oregon for guidance and help with RAD sequencing, the national authorities in Argentina, Bolivia, Colombia, Ecuador, and Peru for permission to collect plant material, all those who helped with and/or facilitated fieldwork in the Andes, including Tim Budden, Ruth Clark, David Neill, and especially Rob Forrest, Terry Pennington and Aniceto Daza Yamona in Peru, John Wood, Arely Palabral, Stefan Beck and Pablo Duchen in Boliva, and Mauricio Andres Rojas and Nicolai Nuerk in Colombia, Rayko Jonas and Markus Meierhofer for help with growing plants in Zurich, and Rosemary Wise for the botanical illustrations reproduced in Fig. 3.

564 Literature Cited

565

566 Agardh, J. G. 1835. Synopsis Generis Lupini. Carolus Fr., Berlin, Germany.

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

570 Antúnez de Mayolo, S. 1982. Tarwi in ancient Peru. In Proceedings of the1st International Lupin
571 Conference, 1-12, Lima/Cuzco, Peru. International Lupin Association.

572 Baer, E. 2011. Domestication of Andean lupin (*L. mutabilis*). *In* Lupin crops: an opportunity for today, a
573 promise for the future. Proceedings of the 13th International Lupin Conference, Poznan´, Poland, 6–
574 10 June 2011, 129–132. International Lupin Association.

575 Blanco, G. O. 1986. Wild stock of *Lupinus mutabilis* Sweet. *In* Proceedings of the 4th International Lupin
576 Conference, Vol. 1, 284, Geraldton, Western Australia. International Lupin Association.Bolger,
577 A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina
578 sequence data. *Bioinformatics* 30: 2114–2120.

579 Bruno, M. C. 2006. A morphological approach to documenting the domestication of *Chenopodium* in the
580 Andes. *In* M. A. Zeder, D. G. Bradley, E. Emshwiller, and B. D. Smith [eds.], Documenting
581 domestication: New genetic and archaeological paradigms, 32–45. University of California Press,
582 Berkeley, California, USA.

583 Bruno, M. C., and W. T. Whitehead. 2003. *Chenopodium* cultivation and Formative period agriculture at
584 Chiripa, Bolivia. *Latin American Antiquity* 14: 339–355.

585 Brush, S. B. 1999. Genes in the field: on–farm conservation of crop diversity. Lewis Publishers, London,
586 UK.

587 Brush, S. B., H. J. Carney, and Z. Hauman. 1981. Dynamics of Andean potato agriculture. *Economic Botany*
588 35: 70–88.

589 Caligari, P. D., P. Römer, M. A. Rahim, C. Huyghe, J. Neves–Martins, and E. J. Sawicka–Sienkiewicz.
590 2000. The potential of *Lupinus mutabilis* as a crop. *In* Linking research and marketing opportunities
591 for pulses in the 21st century, 569–573. Springer, Netherlands.

592 Catchen, J., A. Amores, P. Hohenlohe, W. Cresko, and J. Postlethwait. 2011. Stacks: building and
593 genotyping loci de novo from short–read sequences. *G3: Genes, Genomes, Genetics* 1: 171–182.

594 Catchen, J., P. Hohenlohe, S. Bassham, A. Amores, and W. Cresko. 2013. Stacks: an analysis tool set for
595 population genomics. *Molecular Ecology* 22: 3124–3140.

596 Cavender–Bares, J., A. González–Rodríguez, D. A. Eaton, A. A. Hipp, A. Beulke, and P. S. Manos. 2015.
597 Phylogeny and biogeography of the American live oaks (*Quercus* subsection Virentes): a genomic
598 and population genetics approach. *Molecular Ecology* 24: 3668–3687.

599 Clements, J. C., M. S. Sweetingham, L. Smith, G. Francis, G. Thomas, and S. Sipas, 2008a. Crop
600 improvement in *Lupinus mutabilis* for Australian agriculture – progress and Prospects. *In* Lupins for
601 health and wealth. Proceedings of the12th International Lupin conference, Freemantle, Western
602 Australia, 14–18 September 2008, 14–18. International Lupin Association.

603 Clements, J., L. Prilyuk, J. Quealy, and G. Francis. 2008b. Interspecific crossing among the New World
604 Lupin Species. *In* Lupins for health and wealth. Proceedings of the12th International Lupin
605 conference, Freemantle, Western Australia, 14–18 September 2008, 324–327. International Lupin
606 Association.

607 Coffman, A., P. Hsieh, S. Gravel, and R. Gutenkunst. 2015. Computationally efficient composite likelihood
608 statistics for demographic inference. *Molecular Biology and Evolution* 33: 591–593.

609 Conterato, I. F. and M. T. Schifino–Wittmann. 2006. New chromosome numbers, meiotic behaviour and
610 pollen fertility in American taxa of *Lupinus* (Leguminosae): contributions to taxonomic and
611 evolutionary studies. *Botanical Journal of the Linnean Society* 150: 229–240.

612 Cowling, W. A., B. J. Buirchell, and M. E. Tapia. 1998. Lupin. *Lupinus* L. Promoting the conservation and
613 use of underutilized and neglected crops. International Plant Genetic Resources Institute, IPGRI,
614 Rome, Italy.

615 Diamond, J. 2002. Evolution, consequences and future of plant and animal domestication. *Nature* 418: 700–
616 707.

617 Dillehay, T. D., J. Rossen, T. C. Andres, and D. E. Williams. 2007. Preceramic adoption of peanut, squash,
618 and cotton in northern Peru. *Science* 316: 1890–1893.

619 Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue.
620 *Phytochemical Bulletin*. 19: 11—15.

621 Drummond, C. S., R. J. Eastwood, S. T. S. Miotto and C. E. Hughes. 2012. Multiple continental radiations
622 and correlates of diversification in *Lupinus* (Leguminosae): Testing for key innovation with
623 incomplete taxon sampling. *Systematic Biology* 61: 443–460.

624 Earl, D. A., and B. M. von Holdt. 2012. STRUCTURE HARVESTER: a website and program for visualizing
625 STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:
626 359–361.

627 Eastwood, R. J., and C. E. Hughes. 2008. Origins of domestication of *Lupinus mutabilis* in the Andes. *In*
628 Lupins for health and wealth. Proceedings of the12th International Lupin conference, Freemantle,
629 Western Australia, 14–18 September 2008, 14–18. International Lupin Association.

630 Eaton, D. A., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*
631 30: 1–6.

632 Eaton, D. A., and R. H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: An example
633 from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* 62: 689–706.

634 Emerson, K. J., J. E. Conn, E. S. Bergo, M. A. Randel, and A. M. Sallum. 2015. Brazilian *Anopheles*
635 *darlingi* Root (Diptera: Culicidae) clusters by major biogeographical region. *PloS One* 10: e0130773.

636 Emshwiller, E. 2002a. Biogeography of the *Oxalis tuberosa* alliance. *The Botanical Review* 68: 128–152.

637 Emshwiller, E. 2002b. Ploidy levels among species in the *Oxalis tuberosa* alliance inferred by flow
638 cytometry. *Annals of Botany* 89: 741–753.

639 Emshwiller, E., and J. J. Doyle. 1998. Origins of domestication and polyploidy in oca (*Oxalis tuberosa*:
640 Oxalidaceae): nrDNA ITS data. *American Journal of Botany* 85: 975–985.

641 Emshwiller, E., and J. J. Doyle. 2002. Origins of domestication and polyploidy in oca (*Oxalis tuberosa*,
642 Oxalidaceae). 2. Chloroplast–expressed glutamine synthetase data. *American Journal of Botany* 89:
643 1042–1056.

644 Emshwiller, E., T. Thiem, A. Grau, V. Nina, and F. Terazas. 2009. Origins of domestication in oca (*Oxalis*
645 *tuberosa*; Oxalidaceae). 3. AFLP data of oca and four wild tuber–bearing taxa. *American Journal of*
646 *Botany* 96: 1839–1848.

647 Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the
648 software STRUCTURE: a simulation study. *Molecular Ecology* 14**:** 2611–2620.

649 Food and Agriculture Organization of the United Nations. 2015. FAOSTAT Database. Also at website
650 http://faostat3.fao.org/home/E

651 Fuller, D. Q. 2010. An emerging paradigm shift in the origins of agriculture. *General Anthropology* 17: 8–
652 12.

653 Fuller, D. Q., E. Asouti, and M. D. Purugganan. 2012a. Cultivation as slow evolutionary entanglement:
654 comparative data on rate and sequence of domestication. *Vegetation History and Archaeobotany* 21:
655 131–145.

656 Fuller, D. Q., G. Wilcox, and R. G. Allaby. 2012b. Early agricultural pathways: moving outside the 'core
657 area' hypothesis in Southwest Asia. *Journal of Experimental Botany* 63: 617–633.

658 Fuller, D. Q., T. Denham, M. Arroyo–Kalin, L. Lucas, C. J. Stevens, L. Qin, R. G. Allaby, and M. D.
659 Purugganan. 2014. Convergent evolution and parallelism in plant domestication revealed by an

660 expanding archaeological record. *Proceedings of the National Academy of Sciences, USA* 111: 6147–
661 6152.

662 Gepts, P. 1996. Origin and evolution of cultivated *Phaseolus* species. *In* B. Pickersgill and J. M. Lock [eds.],
663 Advances in legume systematics Part 8, Legumes of economic importance, 65–74. Royal Botanic
664 Gardens, Kew, UK.

665 Gladstones, J. S., C. A. Atkins, and J. Hamblin. 1998. Lupins as crop plants: biology, production and
666 utilization. CAB International, Wallingford, U.K.

667 Gross, R. 1982. El Cultivo y la Utiliziación del *Lupinus mutabilis* Sweet. FAO, Rome, Italy.

668 Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. Inferring the joint
669 demographic history of multiple populations from multidimensional SNP frequency data. *PLoS*
670 *Genetics* 5: e1000695.

671 Hancock, J. F. 2012. *Plant Evolution and the Origin of Crop Species*. CAB International, 3rd Ed.,
672 Wallingford, U.K. & Cambridge, U.S.A.

673 Hardy A., C. Huyghe, M. A. Rahim, P. Roemer, J. M. Neves–Martins, E. Sawicka–Sienkiewicz, and P. D.
674 Caligari. 1998. Effect of genotype and environment on architecture and flowering time of
675 indeterminate Andean lupins (*Lupinus mutabilis* Sweet). *Australian Journal of Agricultural Research*
676 49: 1241–1251.

677 Hastorf, C. A. 1993. Agriculture and the onset of political inequality before the Inka., 1st ed. Cambridge
678 University Press, Cambridge, UK.

679 Hawkes, J. G. 1990. The potato, evolution, biodiversity, and genetic resources. Smithsonian Institution Press,
680 Washington D.C., USA.

681 Hernández–Bermejo, J. E., and J. León. 1994. Neglected crops. 1492 from a different perspective. FAO,
682 Rome, Italy.

683 Hoisington, D., M. Khairallah, T. Reeves, J. M. Ribaut, B. Skovmand, S. Taba, and M. Warburton. 1999.
684 Plant genetic resources: what can they contribute toward increased crop productivity? *Proceedings of*
685 *the National Academy of Sciences, USA* 96: 5937–5943.

686 Hughes, C. E., and G. W. Atchison. 2015. The ubiquity of alpine plant radiations: from the Andes to the
687 Hengduan Mountains. *New Phytologist* 207: 275–282.

688 Hughes, C. E., and R. J. Eastwood. 2006. Island radiation on a continental scale: exceptional rates of plant
689 diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences, USA* 103:
690 10334–9.

691 Hughes, C. E., G. Rajanikanth, A. Robertson, D. L. Filer, S. A. Harris, and C. D. Bailey. 2007. Serendipitous
692 backyard hybridization and the origin of crops. *Proceedings of the National Academy of Sciences*, *USA*
693 104: 14389–14394.

694 Iltis, H. 1998. Serendipity in the exploration of biodiversity. What good are weedy tomatoes? *In* E. O.
695 Wilson [ed.] Biodiversity, 98–105. National Academy Press, Washington D.C., USA.

696 Jacobsen, S.-E., and A. Mujica. 2008. Geographical distribution of the Andean lupin (*Lupinus mutabilis*
697 Sweet). *Plant Genetic Resources Newsletter* 155: 1–8.

698 Jarvis, D. I., A. H. Brown, P. H. Cuong, L. Collado–Panduro, L. Latournerie–Moreno, S. Gyawali, T. Tanto,
699 M. Sawadogo, I. Mar, M. Sadiki, and N. T. N. Hue. 2008. A global perspective of the richness and
700 evenness of traditional crop–variety diversity maintained by farming communities. *Proceedings*
701 *National Academy of Sciences, USA* 105: 5326–5331.

702 Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. Ho, B. C. Faircloth, B. Nabholz, J. T.
703 Howard, and A. Suh. 2014. Whole–genome analyses resolve early branches in the tree of life of
704 modern birds. *Science* 346: 1320–1331.

705 Kamphuis L. G., J. K. Hane, M. N. Nelson, L. Gao, C. A. Atkins, and K. B. Singh. 2014. Transcriptome
706 sequencing of different narrow–leafed lupin tissue types provides a comprehensive uni–gene
707 assembly and extensive gene–based molecular markers. *Plant Biotechnology Journal* 13: 14–25.

Kazimierski, T., and E. Nowacki. 1961. Indigenous species of lupins regarded as the initial forms of the cultivated species *Lupinus albus* and *Lupinus mutabilis*. *Flora* 151: 202–209.

Kozlov, A., A. J. Aberer, and A. Stamatakis. 2015. ExaML Version 3: a tool for phylogenomic analyses on supercomputers. *Bioinformatics* 31: 2577–2579.

Langlie, B. S., C. A. Hastorf, M. C. Bruno, M. Bermann, R. M. Bonzani, and W. Castellón Condarco. 2011. Diversity in Andean *Chenopodium* domestication: Describing a new morphological type from La Barca, Bolivia 1300–1250 B.C. *Journal of Ethnobiology* 31: 72–88.

Langlie, B. S., N. G. Mueller, R. N. Spengler, and G. J. Fritz. 2014. Agricultural origins from the ground up: archaeological approaches to plant domestication. *American Journal of Botany* 101: 1601–1617.

Larson, G., and D. Q. Fuller. 2014. The evolution of animal domestication. *Annual Review of Ecology, Evolution, and Systematics* 45: 115–136.

Larson, G., D. R. Piperno, R. G. Allaby, M. D. Purugganan, L. Andersson, M. Arroyo–Kalin, L. Barton, C. C. Vigueria, T. Denham, K. Dobney, et al., 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences, USA* 111: 6139–6146.

Leaché, A. D., B. L. Banbury, J. Felsenstein, A. N. –M. De Oca, and A. Stamatakis. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology* 64: 1032–1047.

Li, H. and R. Durbin. 2010. Fast and accurate long–read alignment with Burrows–Wheeler Transform. *Bioinformatics* 26: 589–595.

Madriñán, S., A. J. Cortés, and J. E. Richardson. 2013. Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Frontiers in Genetics*, 4: 192.

Martin, C. H., and L. C. Feinstein. 2014. Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. *Molecular Ecology* 23: 1846–1862.

Massatti, R., A. A., Reznicek, and L. L. Knowles. 2016. Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in *Carex* sect. Racemosae. *American Journal of Botany* 103: 337–347.

734 Mathieson, I., I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K.
735 Stewardson, D. Fernandes, M. Novak, and K. Sirak. 2015. Genome–wide patterns of selection in 230
736 ancient Eurasians. *Nature* 528: 499–503.

737 Mengoni Goñalons, G. L. 2008. Camelids in ancient Andean societies: a review of the zooarchaeological
738 evidence. *Quaternary International* 185: 59–68.

739 Misof, B., S. Liu, K. Meusemann, R. S. Peters, A. Donath, C. Mayer, P. B. Frandsen, J. Ware, T. Flouri, R.
740 G. Beutel, and O. Niehuis. 2014. Phylogenomics resolves the timing and pattern of insect evolution.
741 *Science* 346: 763–767.

742 Morris, A. 1999. The agricultural base of the pre–Inca Andean civilizations. *The Geographical Journal* 165:
743 286–295.

744 National Research Council. 1989. Lost crops of the Incas – little–known plants of the Andes with promise
745 for worldwide cultivation. National Academy Press, Washington D.C., U.S.A.

746 Pante, E., J. Abdelkrim, A. Viricel, D. Gey, S. C. France, M. C. Boiselier, and S. Samadi. 2015. Use of RAD
747 sequencing for delimiting species. *Heredity* 114: 450–459.

748 Pearsall, D. M. 1980. Pachamachay ethnobotanical report: Plant utilization at a hunting base camp. *In* J. W.
749 Rick [ed.] Prehistoric hunters of the high Andes, 191-231. Academic Press, New York, USA.

750 Pearsall, D. M. 1988. Interpreting the meaning of macroremain abundance: the impact of source and context.
751 *In* C. A. Hastorf and V. S. Popper [eds.], Current paleoethnobotany: analytical methods and cultural
752 interpretations of archaeological plant remains, 97-118. The Chicago University Press, London, UK.

753 Pearsall, D. M. 1992. The origins of plant cultivation in South America. *In* C. W. Cowan and P. J. Watson
754 [eds.], The origins of agriculture. An international perspective, 173–205. Smithsonian Institution
755 Press, Washington D.C., USA.

756 Pearsall, D. M. 2003. Plant food resources of the Ecuadorian Formative: an overview and comparison to the
757 central Andes. *In* J. S. Raymond and R. L. Burger [eds.], Archaeology of Formative Ecuador, 213–
758 257. Dumbarton Oaks, Washington D.C., USA.

759 Pearsall, D. M. 2008. Plant domestication and the shift to agriculture in the Andes. *In* H. Sliverman and W.
760 H. Isbell [eds.], Handbook of South American archaeology, 105–120. Springer, New York, USA.

761 Peralta, I. E., and D. M. Spooner. 2006. History, origin and early cultivation of tomato (Solanaceae). *Genetic*
762 *Improvement of Solanaceous Crops* 2: 1–27.

763 Piperno, D., and D. Pearsall. 1998. The origins of agriculture in the lowland Neotropics. Academic Press,
764 New York.

765 Piperno, D. R., and T. D. Dillehay. 2008. Starch grains on human teeth reveal early broad crop diet in
766 northern Peru. *Proceedings of the National Academy of Sciences, USA* 105: 19622–19627.

767 Plowman, T. 1984. The origin, evolution, and diffusion of coca, *Erythroxylum spp*. in South and Central
768 America. *In* D. Stone [ed.], *Pre–Colombian Plant Migration,* 125–163. Peabody Museum, Harvard
769 University, Cambridge, MA, U.S.A.

770 Preece, C, A. Livarda, M. Wallace, G. Martin, M. Charles, P. –A. Christin, G. Jones, M. Rees, and C. P.
771 Osborne. 2015. Were Fertile Crescent crop progenitors higher yielding than other wild species that
772 were never domesticated? *New Phytologist* 207: 905–913.

773 Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus
774 genotype data. *Genetics* 155: 945–959.

775 Rambaut, A. 2009. FigTree v1. 3.1: Tree figure drawing tool. Website:
776 http://tree.bio.ed.ac.uk/software/figtree [accessed 23 March 2016]

777 Ree, R. H., and A. L. Hipp. 2015. Inferring phylogenetic history from restriction site associated DNA
778 (RADseq). *In* E. Hörandl and M. S. Appelhans [eds.], Next–generation sequencing in plant
779 systematics. Regnum Vegetabile Vol. 158., 181–204. Koeltz Scientific, Königstein, Germany.

780 Royal Botanic Gardens Kew. 2016. Seed Information Database (SID). Version 7.1. Website
781 http://data.kew.org/sid/ [accessed 01 April 2016]

782 Russello, M. A., M. D. Waterhouse, P. D. Etter, and E. A. Johnson. 2015. From promise to practice: pairing
783 non–invasive sampling with genomics in conservation. *PeerJ* 3: e1106.

784 Schmutz, J., P. E. McClean, S. Mamidi, G. A. Wu, S. B. Cannon, J. Grimwood, J. Jenkins, S. Shu, Q. Song,
785 C. Chavarro, et al. 2014. A reference genome for common bean and genome-wide analysis of dual
786 domestications. *Nature Genetics* 46: 707–713.

787 Smith, B. D. 2001. Documenting plant domestication: the consilience of biological and archaeological
788 approaches. *Proceedings of the National Academy of Sciences, USA* 98: 1324–1326.

789 Smith, C. P. 1938–1953. Species Lupinorum. Published privately, California, USA.

790 Spooner, D. M., K. McLean, G. Ramsay, R. Waugh, and G. J. Bryan. 2005. A single domestication for
791 potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the*
792 *National Academy of Sciences, USA* 102: 14694–9.

793 Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
794 phylogenies. *Bioinformatics* 30: 1312–1313

795 Taba, S., F. Santacruz–Ruvalcaba, S. Sood, J. B. Holland, S. Carvajal, and R. Barba–González. 2011. Three
796 new teosintes (*Zea spp.*, Poaceae) from Mexico. *American Journal of Botany* 98: 1537–1548.

797 Takahashi, T., N. Nagata, and T. Sota. 2014. Application of RAD–based phylogenetics to complex
798 relationships among variously related taxa in a species flock. *Molecular Phylogenetics and*
799 *Evolution* 80: 137–144.

800 Tapia, M. and C. Vargas. 1982. Wild lupines of the Andes of southern Peru. *In* Proceedings of the 1st
801 International Lupin Conference, pp. 23–32, Lima/Cuzco, Peru. International Lupin Association.

802 Tapia, M., A. Riva, and G. Hernàndez–Bravo. 1988. Cool season food legumes in South and Central
803 America. *In* R. J. Summerfield [ed.], World crops: cool season food legumes, 1051–1057. Springer,
804 Netherlands.

805 Towle, M. A. 1961. The ethnobotany of pre-Columbian Peru. Aldine Publishing Company, Chicago, USA.

van den Berg, R. G., N. Groendijk–Wilders, and J. P. Kardolus. 1996. The wild ancestors of cultivated potato: the brevicaule–complex. *Acta Botanica Neerlandica* 45: 157–171.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome–wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22: 787–798.

Walsh, B. M., D. Adhikary, P. J. Maughan, E. Emshwiller, and E. N. Jellen. 2015. *Chenopodium* polyploidy influences from Salt Overly Sensitive 1 (SOS1) genes. *American Journal of Botany* 102: 533–543.

Weeden, N. F. 2007. Genetic changes accompanying the domestication of *Pisum sativum*: Is there a common genetic basis to the 'domestication syndrome' for legumes? *Annals of Botany* 100: 1017–1025.

Wellesley, L., C. Happer, and A. Froggatt. 2015. Changing climate, changing diets: pathways to lower meat consumption. Royal Institute of International Affairs Report, Chatham House, London, UK.

Wheeler, J. C. 1995. Evolution and present situation of the South American Camelidae. *Biological Journal of the Linnean Society* 54: 271–295.

Wilson, H. D. 1990. Quinua and relatives (*Chenopodium*, sect. *Chenopodium,* subsect. *Cellulata*). *Economic Botany* 44: 92–110.

Yang H., Y. Tao, Z. Zheng, Q. Zhang, G. Zhou, M. W. Sweetingham, J. G. Howieson, C. Li. 2013. Draft genome sequence, and a sequence–defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. *PLoS One* 8: e64799.

Young, K. R., C. Ulloa, J. L. Luteyn, and S. Knapp. 2002. Plant evolution and endemism in Andean South America: an introduction. *The Botanical Review* 68: 4–21.

Zeder, M. A. 2008. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion and impact. *Proceedings of the National Academy of Sciences, USA* 105: 11597–11604.

Zeder, M. A. 2015. Core questions in domestication research. *Proceedings of the National Academy of Sciences, USA* 112: 3191–3198.

Table 1 – Parameter estimates for different demographic scenarios concerning the split between *L. mutabilis* and *L. piurensis*, and respective standard deviations (sd). IM – full IM model with migration in both directions following split. noM – IM model without migration after split. M12 – IM model with migration after split only from *L. piurensis* to *L. mutabilis*.

Model	s	sd (s)	N_1	sd (N_1)	N_2	sd (N_1)	T_s	sd (T_s)	$M_{1<=2}$	sd ($M_{1<=2}$)	$M_{2<=1}$	sd ($M_{2<=1}$)	LL	Adjusted LRT
IM	0.049	0.0062	1.042	0.152	1.359	0.197	0.249	0.02	0.088	0.050	7.60e-05	0.027	-382.055	NA
noM	0.053	0.0126	1.077	0.328	1.326	0.208	0.239	0.031	-	-	-	-	-383.700	2.3592
M12	0.049	0.0090	1.043	0.212	1.359	0.207	0.248	0.028	0.088	0.052	-	-	-382.056	0.0016

Parameters are: s – relative size of *L. mutabilis* population at time of split from ancestral species (*L. piurensis* relative population size at time of split is 1-s); N_1 and N_2 – current population size of *L. mutabilis* and *L. piurensis* respectively (relative to ancestral population size); T – time of population split (in units of 2 x ancestral population size); $M_{1<=2}$ and $M_{2<=1}$ – migration from *L. piurensis* into *L. mutabilis* and from *L. mutabilis* into *L. piurensis* respectively (in units of 2 x ancestral population size x m, where m is the proportion of the receiving population made of immigrants in each generation).

LL – loglikelihood of each model. Adjusted LRT is the adjusted Likelihood-Ratio-Test comparing each simpler model to the full IM model (test is non-significant in both cases, i.e. each of the simpler models fits the data as well as the more complex full IM model).

Figures

Fig. 1: Cultivation and food use of tarwi in the Andes. A & C, typical fields of tarwi (all the blue-green fields in C), here at 3550m altitude, close to Juntatuyo, Cochabamba, Bolivia; B, variation in seed coat pigmentation, including pale white seeds typical of domesticated tarwi, from cultivated plants on Isla del Sol, Lago Titicaca at 3840m altitude in Bolivia; D, pink flower colour mutant individuals occur sporadically in otherwise blue-flowered fields, at 3300m altitude, close to Independencia, Cochabamba, Bolivia; E & G, tarwi seed after de-bittering, for sale – E in Loja, Ecuador, G at Tiquina, Lago Titicaca, Bolivia; F, cultivated tarwi and *quinoa* on the outskirts of the pre-Inca ruins of Markahuamachuco, La Libertad, central Peru. All photos C. Hughes.

Fig. 2: A. Map showing the distributions of cultivated *L. mutabilis* and its wild progenitor *L. piurensis* based on herbarium specimen data, plus locations of archaeological sites where lupin seeds have been found; B timeline of archaeological remains of *Lupinus* in Peru and Ecuador. [1]Towle (1961); [2]Pearsall (1980); [3]Pearsall (1988); [4]Pearsall (2003); [5]Hastorf (1993); [6]Antúnez-de Mayolo (1982) (for more details of Archaeological records see Appendix S4).

Fig. 3: Morphology of domesticated tarwi, *L. mutabilis* (upper row) and its putative progenitor *L. piurensis* (lower row). A, habit; B, inflorescence, both from *Hughes 2220*, Celendín, Cajamarca, Peru; C, indehiscent pods from *Hughes 2326*, Isla del Sol, Lago Titicaca, Bolivia; D, drawing; E, habit from *Hughes 1997*, San Juan, Cajamarca, Peru; F, inflorescence and G, pods, both from *Hughes 3066*, Contumaza, Cajamarca, Peru; H, drawing. Photos C. Hughes, botanical illustrations drawn by Rosemary Wise.

Fig. 4: A. Phylogeny of Andean *Lupinus* showing major clades, including the *L. piurensis*/*L. mutabilis* subclade, based on RADseq data for 212 accessions (for a fully annotated phylogeny showing names of all terminals see Appendix S2); B. Phylogeny of the *L. piurensis* and *L. mutabilis* subclade. Numbers above nodes are bootstrap values. Geographic origins of accessions are shown next to tip labels with ISO three letter country codes and province/state; C. Genetic structure within *L. piurensis* and *L. mutabilis*.

Fig. 5: Demographic history of *L. mutabilis* and *L. piurensis*. A. The IM model used in the analyses presented here, showing all six parameters plus the ancestral population size (N_a) used to scale parameters. Parameters are: *s* – relative size of the *L. mutabilis* population at the time of splitting from the ancestral species (*L. piurensis* relative population size at time of split is 1-*s*); N_1 and N_2 – current population sizes of *L. mutabilis* and *L. piurensis* respectively (relative to N_a); *Ts* – time of population split (in units of 2 x N_a); $M_{1<=2}$ and $M_{2<=1}$ – migration rates from *L. piurensis* into *L. mutabilis* and from *L. mutabilis* into *L. piurensis* respectively (in units of 2 x N_a x *m*, where *m* is the proportion of the receiving population made of immigrants in each generation). The relative sizes of the parameters are roughly proportional to best estimates of IM model fitted in *dadi*. B. Comparison of observed (top left) and expected (top right) two-dimensional Site Frequency Spectra, and their respective residuals (bottom) after fitting the IM model without migration after splitting.

Fig. 6: Seed weights of 35 Andean *Lupinus* species, including putative close relatives of *L. mutabilis* and its most likely progenitor, *L. piurensis*, showing the dramatic weight increase associated with domestication of *L. mutabilis* (modified from Eastwood and Hughes, 2008).

Online Supplemental Materials

Appendix S1: Identities, localities and voucher data for sequenced accessions.

Appendix S2: ExaML phylogeny for Andean *Lupinus.*

Appendix S3: Raw data for seed measurements.

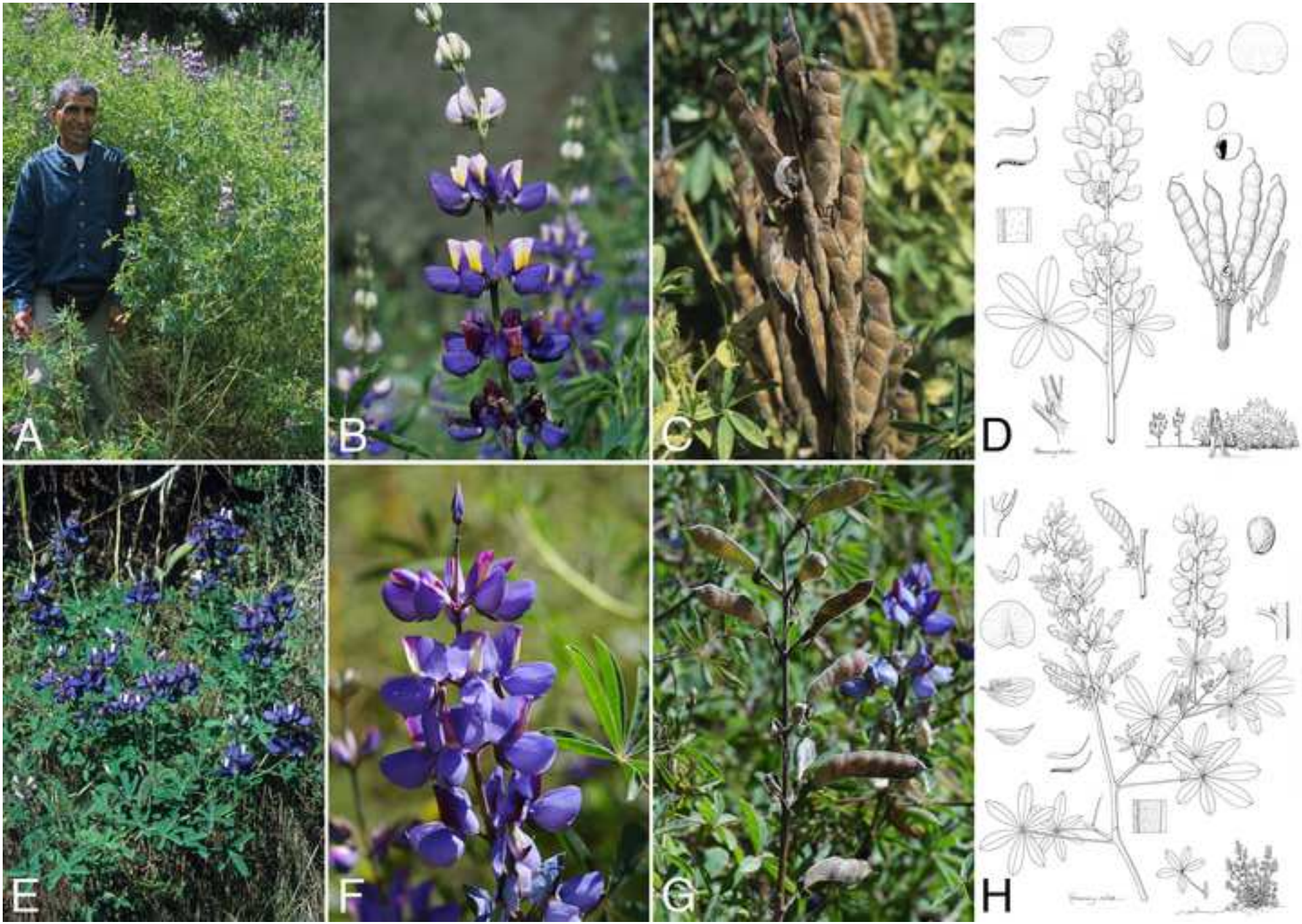
Appendix S4: Summary listing of Archaeological records containing *Lupinus* seeds in the Andean region.

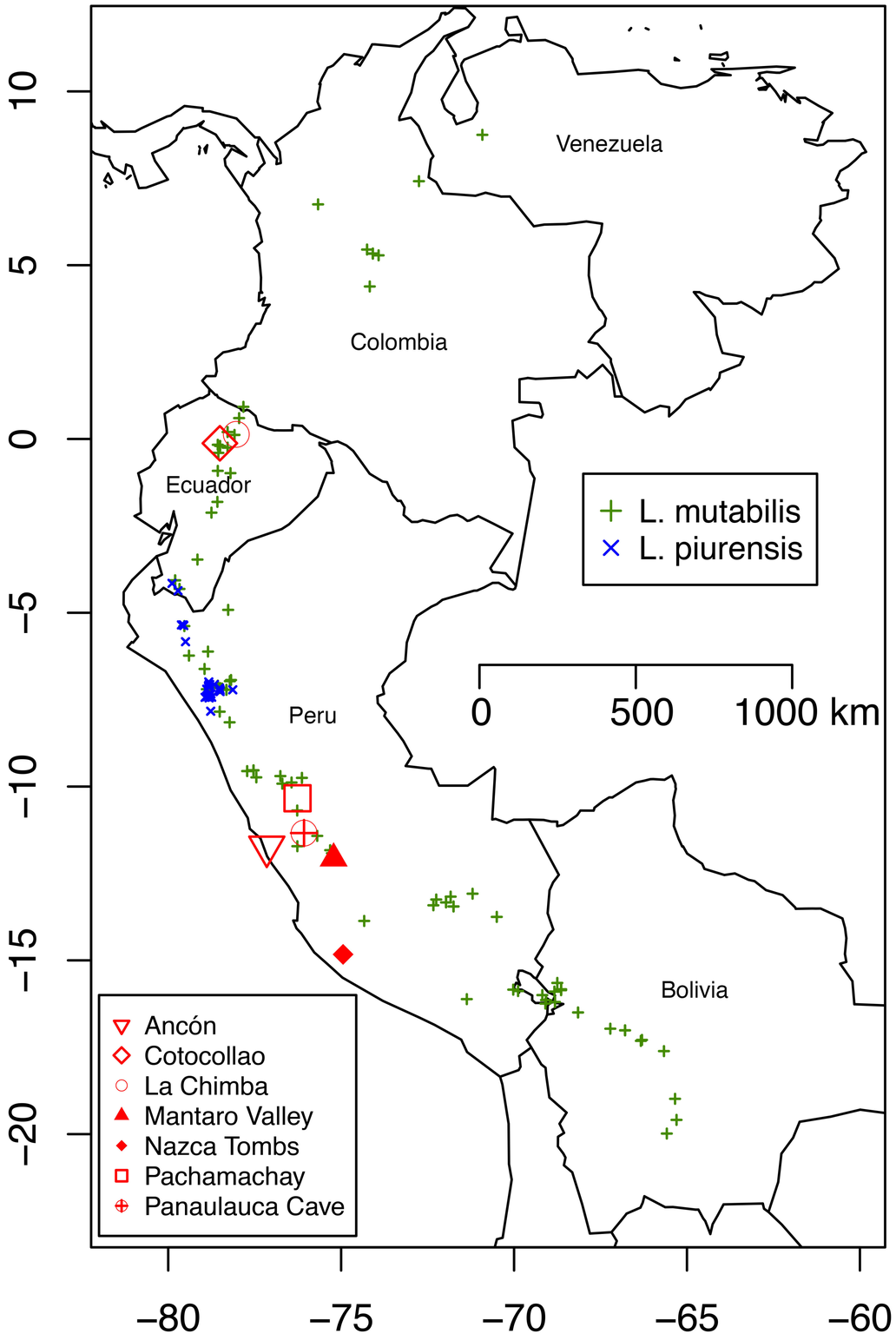
Figure

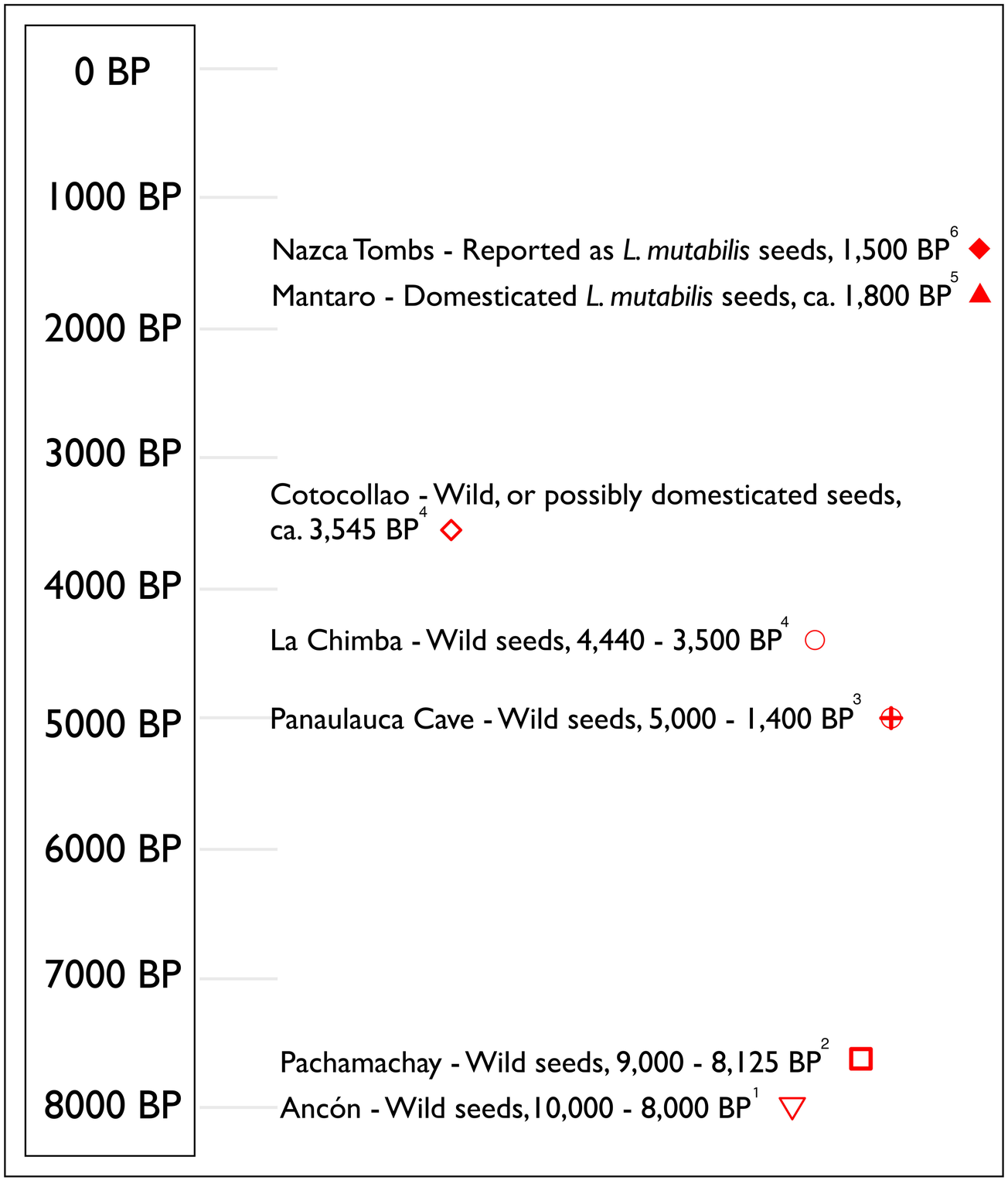
Click here to download Figure Fig_1_Cultivation and food use of tarwi in the Andes.tiff

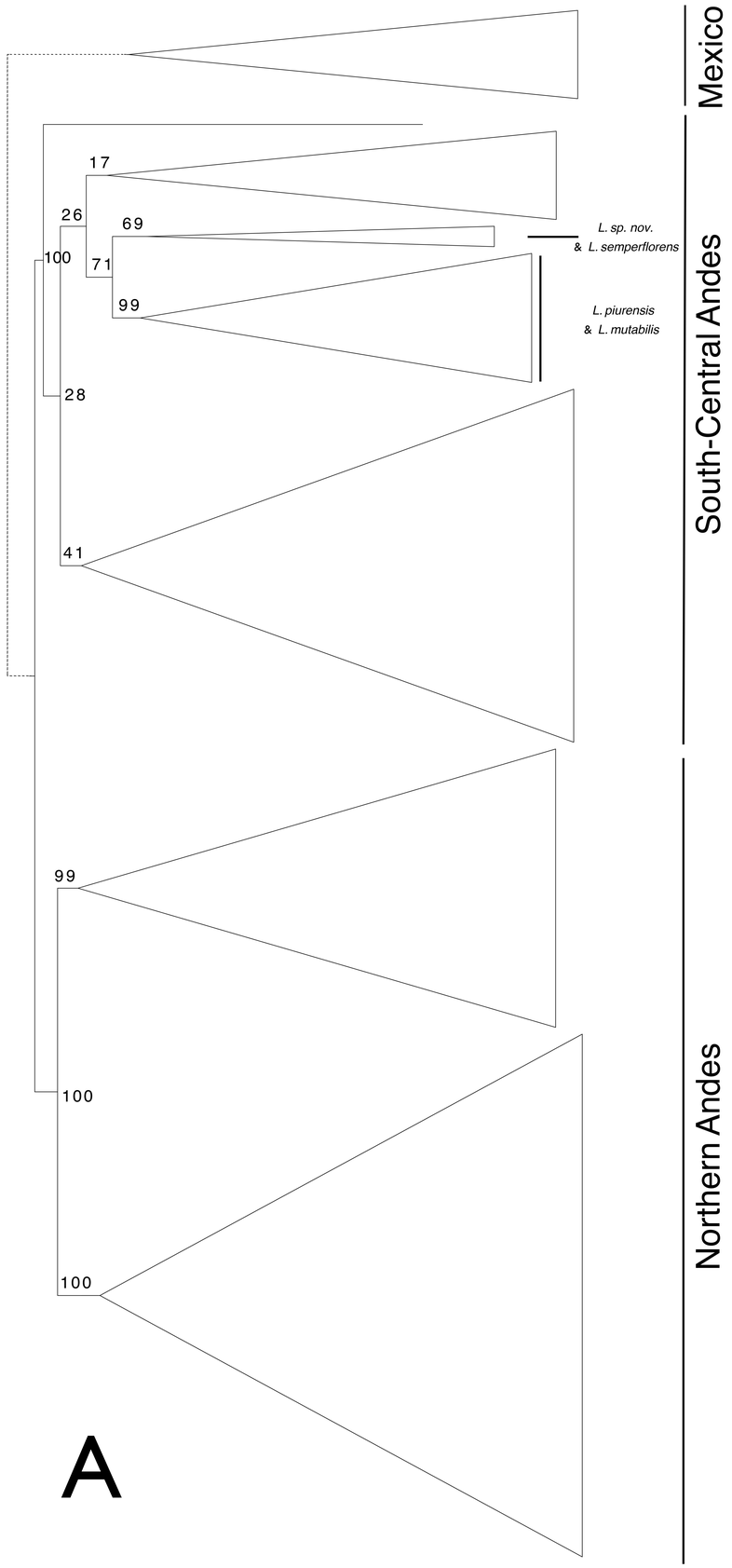
Figure

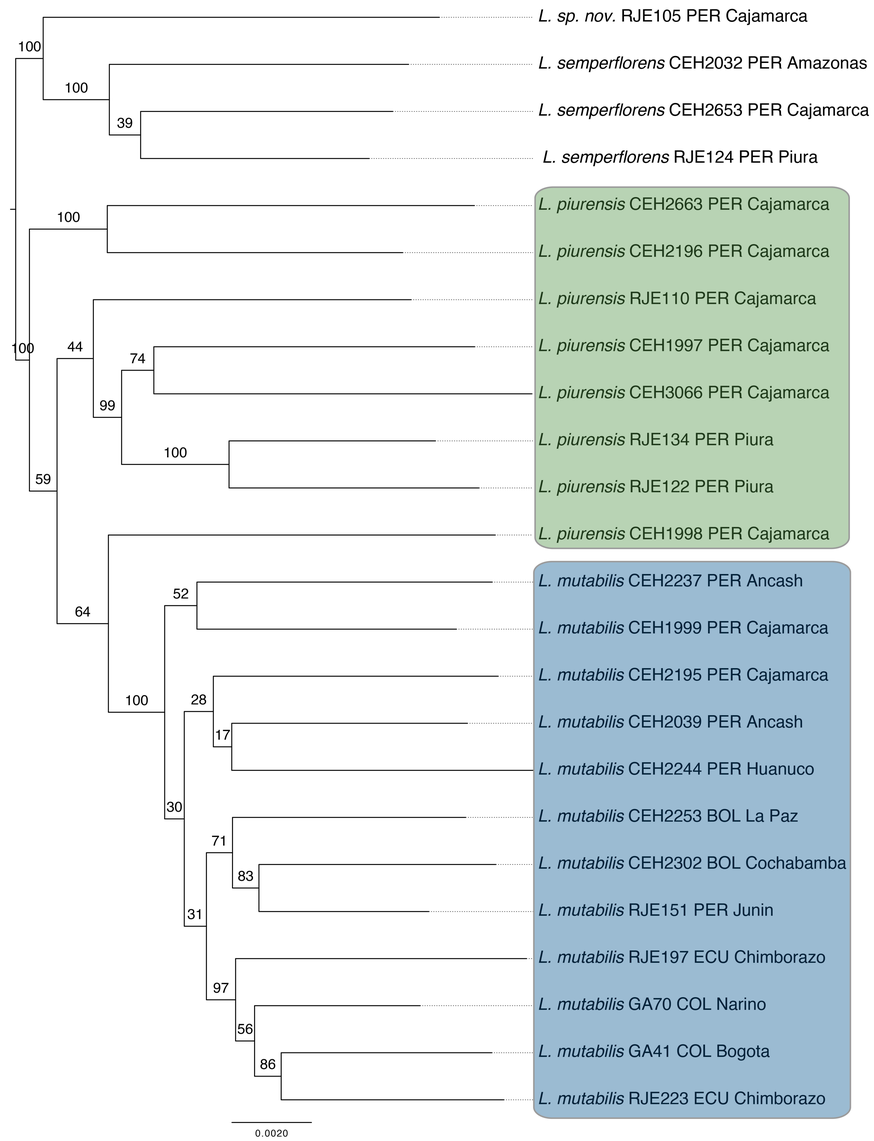
Click here to download Figure Fig_2_Morphology of domesticated tarwi and its putative progenitor.tiff

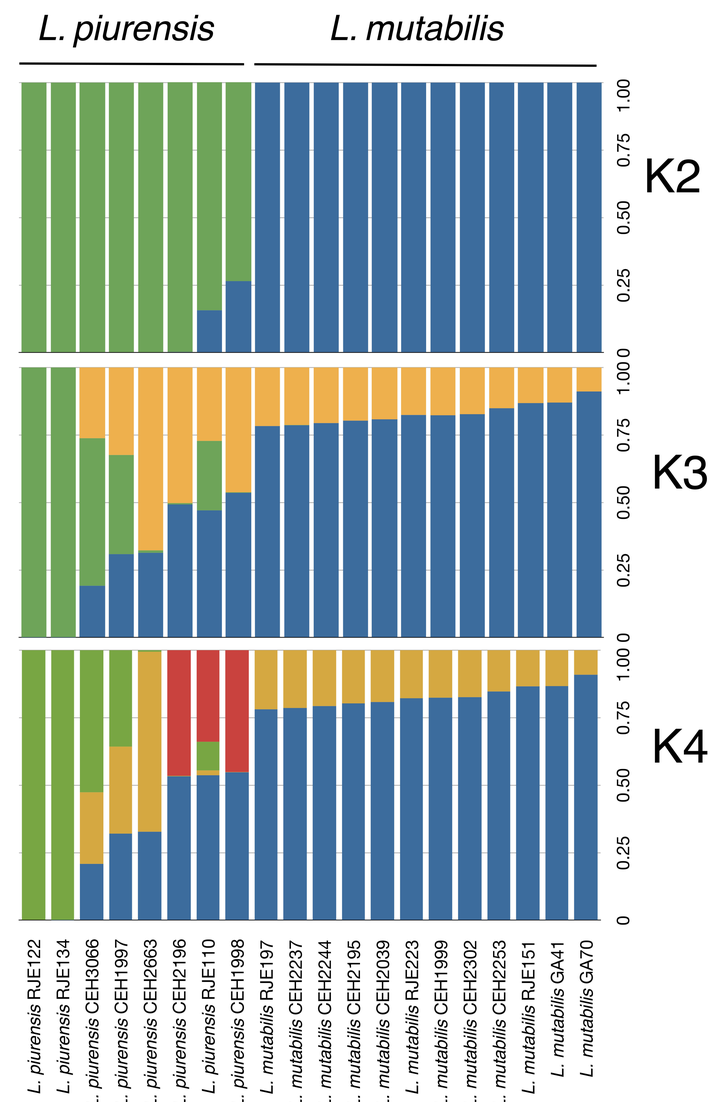




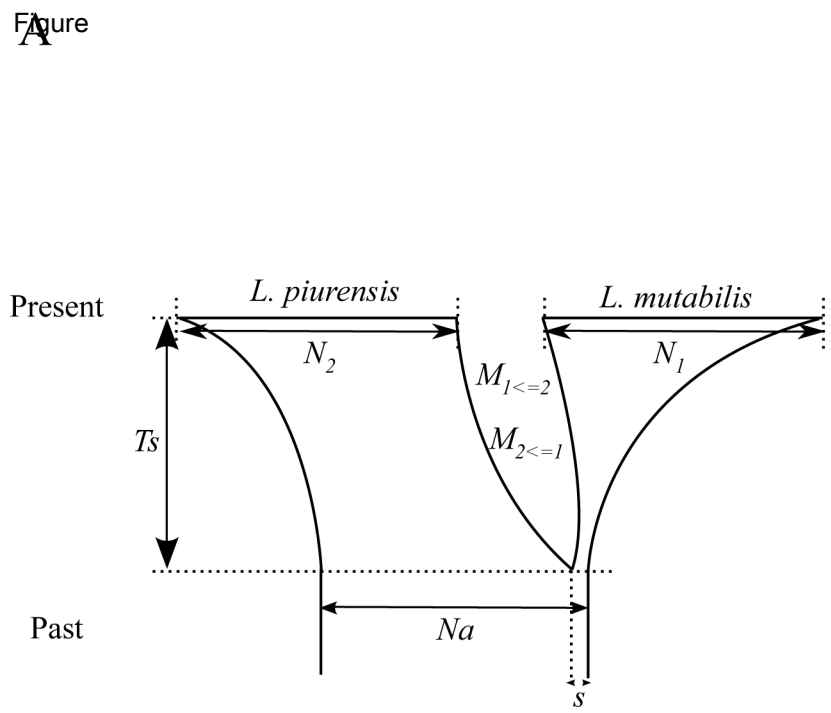


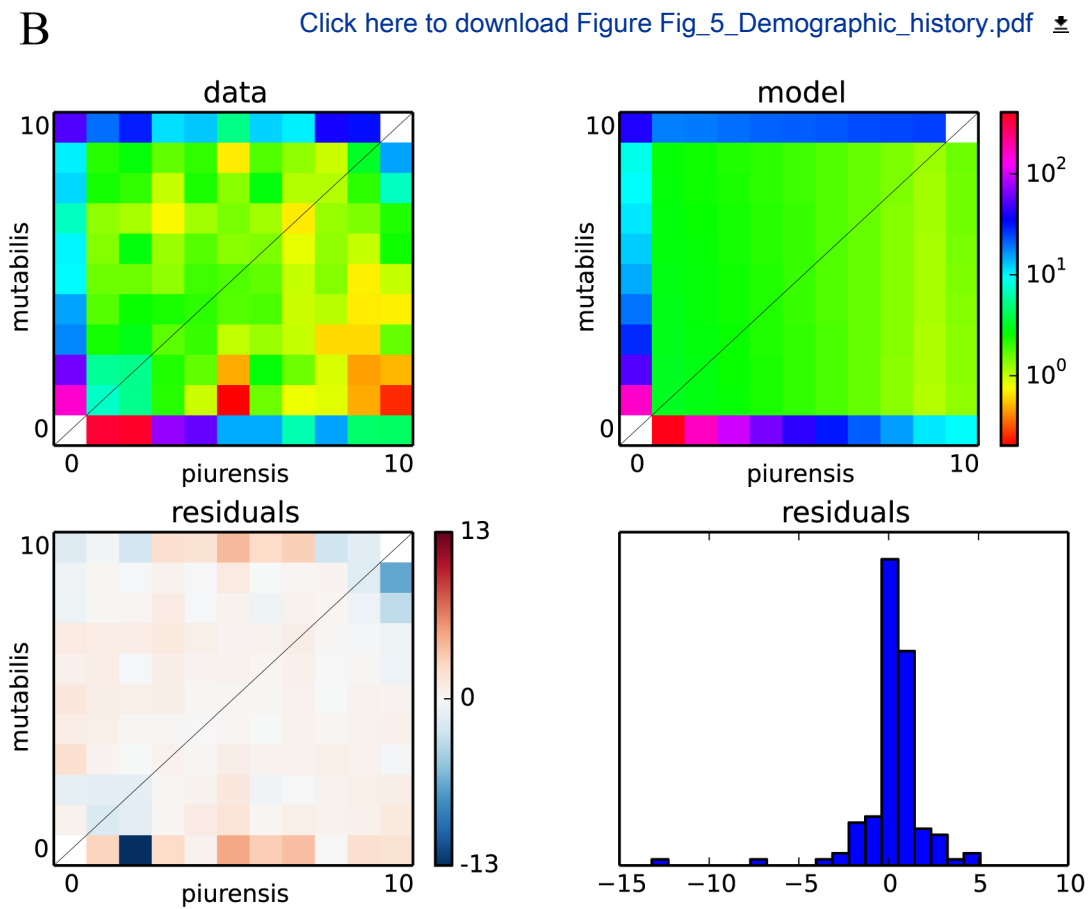






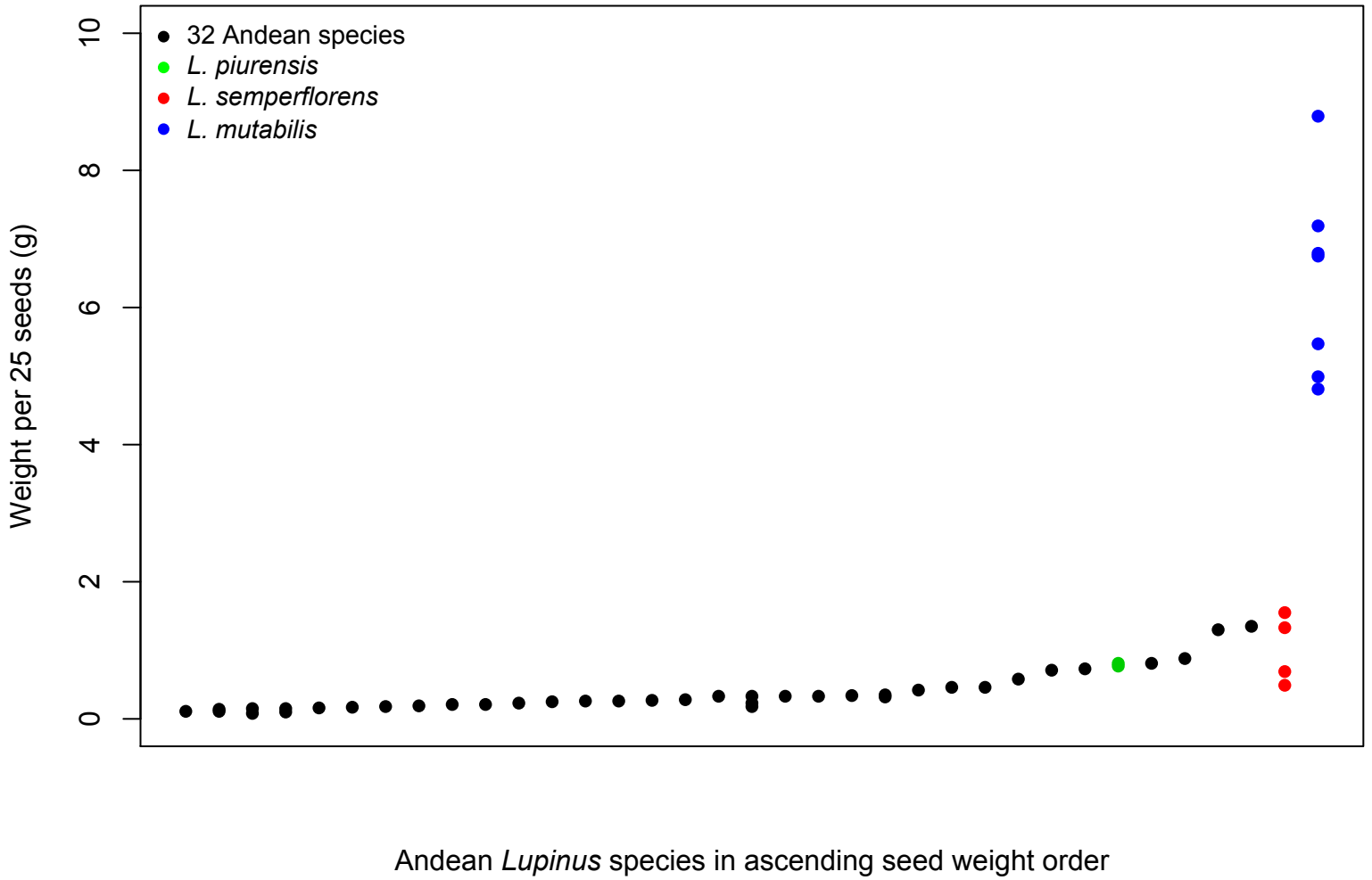
Figure


A
Present
Past
L. piurensis
L. mutabilis
N_2
N_1
$M_{1<=2}$
$M_{2<=1}$
Ts
Na
s


B
Click here to download Figure Fig_5_Demographic_history.pdf
data
model
residuals
residuals
mutabilis
piurensis
0
10
13
-13
10^2
10^1
10^0
−15
−10
−5
5

Figure

Click here to download Figure Fig_6_Seed_weights of 35 Andean Lupinus species.pdf



Click here to download Online Supplemental Appendix_S1 Identities localities and voucher data for sequenced accessions.xls

Taxon	ID	Latitude	Longitude	Country	State/Province	Nreads	SRA accession
Lupinus interruptus	AC11009	6.3791	-72.2806	Colombia	Arauca	#######	TBA
Lupinus smithianus Kunth.	BBK702	-0.0667	-78.3833	Ecuador	Cotopaxi	#######	TBA
Lupinus nubigenus Kunth	BBK704	-0.1710	-78.5980	Ecuador	Pichincha	342,660	TBA
Lupinus ballianus C.P.Sm.	CEH1989	######	-76.3167	Peru	Lima	#######	TBA
Lupinus huaronensis J.F.Macbr.	CEH1991	######	-76.2039	Peru	Lima	#######	TBA
Lupinus piurensis C.P.Sm.	CEH1997	-7.2578	-78.5114	Peru	Cajamarca	#######	TBA
Lupinus piurensis C.P.Sm.	CEH1998	-7.2578	-78.5114	Peru	Cajamarca	#######	TBA
Lupinus mutabilis Sweet	CEH1999	-7.1789	-78.4422	Peru	Cajamarca	#######	TBA
Lupinus mantaroensis C.P.Sm.	CEH2000	-7.2081	-78.4050	Peru	Cajamarca	#######	TBA
Lupinus dillonii	CEH2002	-7.1928	-78.3503	Peru	Cajamarca	#######	TBA
Lupinus sp. nov.	CEH2004	-7.0494	-78.2783	Peru	Cajamarca	#######	TBA
Lupinus solangorum C.P.Sm.	CEH2007	-7.0369	-78.2500	Peru	Cajamarca	#######	TBA
Lupinus solangorum C.P.Sm.	CEH2014	-7.0022	-78.5692	Peru	Cajamarca	#######	TBA
Lupinus humifusus Sessé & Moc. ex G.	CEH2015	-6.9058	-78.6081	Peru	Cajamarca	#######	TBA
Lupinus prostratus different	CEH2026	-6.7164	-77.8736	Peru	Amazonas	#######	TBA
Lupinus chachas Ochoa ex C.P.Sm.	CEH2027	-6.7128	-77.8456	Peru	Amazonas	#######	TBA
Lupinus semperflorens Hartw. ex Benth	CEH2032	-5.8086	-77.9525	Peru	Amazonas	#######	TBA
Lupinus mutabilis Sweet	CEH2039	-9.5556	-77.7175	Peru	Ancash	#######	TBA
Lupinus huaronensis J.F.Macbr.	CEH2048	-9.8681	-77.0819	Peru	Ancash	#######	TBA
Lupinus mutabilis Sweet	CEH2195	-7.1158	-78.8103	Peru	Cajamarca	#######	TBA
Lupinus piurensis C.P.Sm.	CEH2196	-7.1158	-78.8103	Peru	Cajamarca	#######	TBA
Lupinus tomentose Cajamarca	CEH2199	-7.0442	-78.2694	Peru	Cajamarca	#######	TBA
Lupinus weberbaueri Ulbr.	CEH2233	-9.1264	-77.5244	Peru	Ancash	#######	TBA
Lupinus mutabilis Sweet	CEH2237	-9.7358	-77.4522	Peru	Ancash	#######	TBA
Lupinus nubigenus Kunth	CEH2239	-9.1264	-77.5244	Peru	Ancash	#######	TBA
Lupinus huaronensis J.F.Macbr.	CEH2241	-9.8692	-77.1789	Peru	Ancash	#######	TBA
Lupinus mutabilis Sweet	CEH2244	-9.6986	-76.7569	Peru	Huanuco	#######	TBA
Lupinus ellsworthianus C.P.Sm.	CEH2246	-9.8919	-76.4628	Peru	Huanuco	#######	TBA
Lupinus purosericeus C.P.Sm.	CEH2248	######	-76.2167	Peru	Pasco	#######	TBA
Lupinus mutabilis Sweet	CEH2253	######	-68.6400	Bolivia	La Paz	#######	TBA
Lupinus soratensis Rusby	CEH2254	######	-68.6664	Bolivia	La Paz	#######	TBA
Lupinus otto-buchteinii C.P.Sm.	CEH2257	######	-68.6231	Bolivia	La Paz	#######	TBA
Lupinus breviscapus Ulbr.	CEH2258	######	-68.5922	Bolivia	La Paz	#######	TBA
Lupinus chrysanthus Ulbr.	CEH2264	######	-68.1200	Bolivia	La Paz	#######	TBA
Lupinus microphyllus Desr.	CEH2265	######	-68.1200	Bolivia	La Paz	#######	TBA

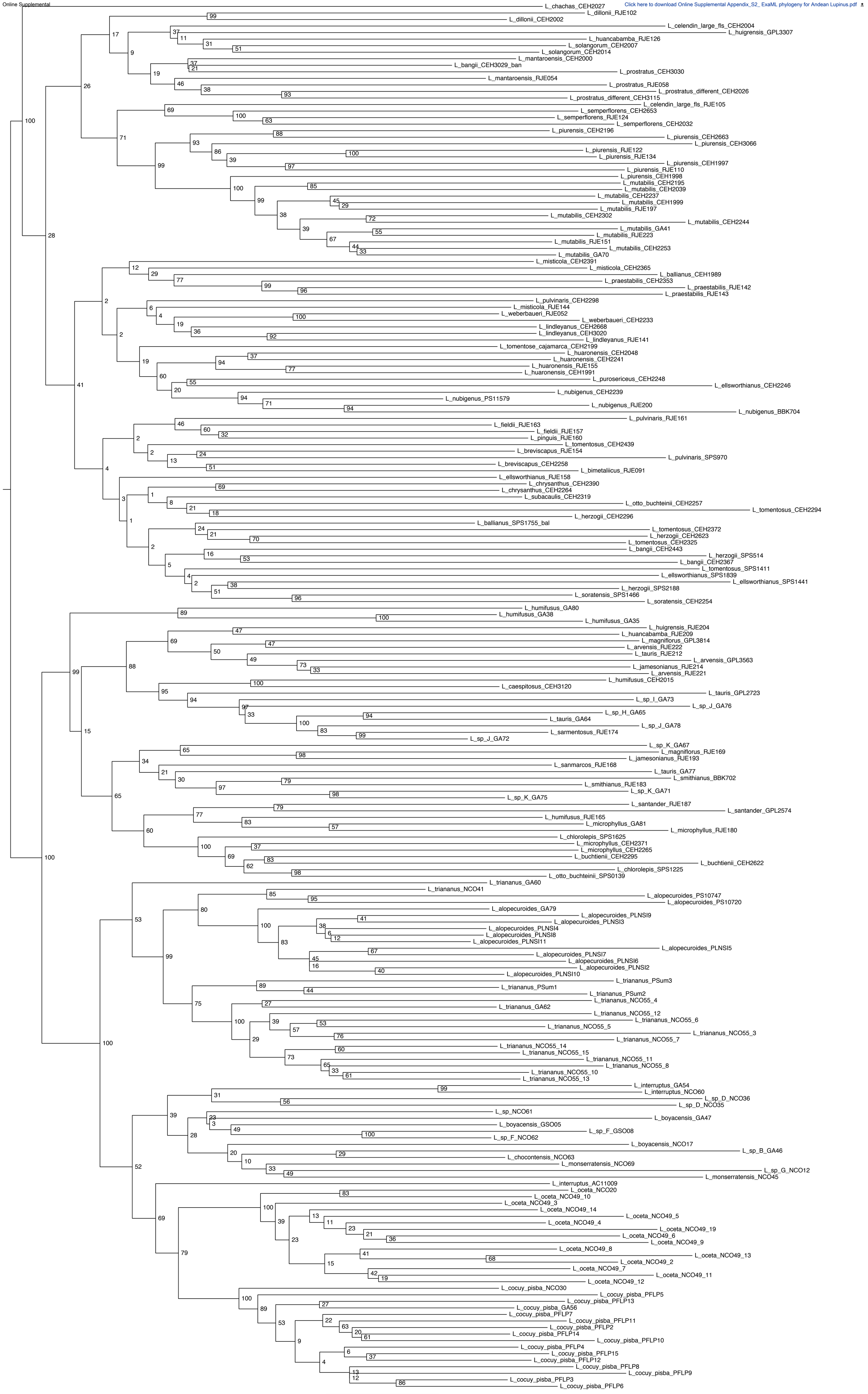
Lupinus tominensis Wedd.	CEH2294	######	-66.3283	Bolivia	Cochabamba	#######	TBA
Lupinus buchtienii Rusby	CEH2295	######	-66.3308	Bolivia	Cochabamba	#######	TBA
Lupinus herzogii	CEH2296	######	-66.3958	Bolivia	Cochabamba	#######	TBA
Lupinus pulvinaris	CEH2298	######	-66.3844	Bolivia	Cochabamba	#######	TBA
Lupinus mutabilis Sweet	CEH2302	######	-65.6661	Bolivia	Cochabamba	#######	TBA
Lupinus subacaulis	CEH2319	######	-65.7333	Bolivia	Potosi	#######	TBA
Lupinus tomentosus DC.	CEH2325	######	-69.1303	Bolivia	La Paz	#######	TBA
Lupinus praestabilis	CEH2353	######	-71.5072	Peru	Arequipa	#######	TBA
Lupinus misticola Ulbr.	CEH2365	######	-70.6592	Peru	Moquegua	#######	TBA
Lupinus bangii Rusby	CEH2367	######	-69.2950	Peru	Puno	#######	TBA
Lupinus microphyllus	CEH2371	######	-70.9589	Peru	Puno	#######	TBA
Lupinus tomentosus DC.	CEH2372	######	-71.0022	Peru	Cuzco	#######	TBA
Lupinus chrysanthus	CEH2390	######	-73.5736	Peru	Apurimac	#######	TBA
Lupinus misticola Ulbr.	CEH2391	######	-73.5781	Peru	Apurimac	#######	TBA
Lupinus tominensis	CEH2439	######	-68.0214	Bolivia	La Paz	#######	TBA
Lupinus bangii Rusby	CEH2443	######	-68.4847	Bolivia	La Paz	#######	TBA
Lupinus buchtienii Rusby	CEH2622	######	-64.9219	Bolivia	Tarija	#######	TBA
Lupinus herzogii Ulbr.	CEH2623	######	-64.9136	Bolivia	Tarija	#######	TBA
Lupinus semperflorens Hartw. ex Benth	CEH2653	-6.3186	-78.8400	Peru	Cajamarca	#######	TBA
Lupinus piurensis C.P.Sm.	CEH2663	-7.1231	-78.8175	Peru	Cajamarca	#######	TBA
Lupinus lindleyanus J.Agardh	CEH2668	-8.9203	-77.8683	Peru	Ancash	#######	TBA
Lupinus lindleyanus J.Agardh	CEH3020	######	-77.4919	Peru	Ancash	#######	TBA
Lupinus bangii	CEH3029	-9.969	-77.3597	Peru	Ancash	#######	TBA
Lupinus prostratus J.Agardh	CEH3030	-9.969	-77.3597	Peru	Ancash	#######	TBA
Lupinus piurensis C.P.Sm.	CEH3066	-7.3328	-78.8111	Peru	Cajamarca	#######	TBA
Lupinus prostratus different	CEH3115	-6.7289	-77.8844	Peru	Amazonas	#######	TBA
Lupinus caespitosus	CEH3120	-6.7333	-77.8939	Peru	Amazonas	#######	TBA
Lupinus humifusus Sessé & Moc. ex G.	GA35	4.9412	-75.3517	Colombia	Caldas	917,483	TBA
Lupinus humifusus Sessé & Moc. ex G.	GA38	4.9373	-75.3337	Colombia	Caldas	#######	TBA
Lupinus mutabilis Sweet	GA41	4.3873	-74.1720	Colombia	Bogota	#######	TBA
Lupinus sp. B	GA46	5.5517	-73.1015	Colombia	Boyaca	748,669	TBA
Lupinus boyacensis C.P.Sm.	GA47	5.5377	-73.0345	Colombia	Boyaca	#######	TBA
Lupinus interruptus	GA54	5.9918	-72.5700	Colombia	Boyaca	#######	TBA
Lupinus cocuy-pisba	GA56	5.9877	-72.5502	Colombia	Boyaca	#######	TBA
Lupinus triananus C.P.Sm.	GA60	4.6725	-73.7580	Colombia	Bogota	#######	TBA
Lupinus triananus C.P.Sm.	GA62	4.7085	-73.8045	Colombia	Bogota	#######	TBA
Lupinus tauris Benth.	GA64	2.3625	-76.3511	Colombia	Huila	#######	TBA
Lupinus sp. H	GA65	2.1707	-76.4645	Colombia	Cauca	#######	TBA

Lupinus sp. K	GA67	0.8443	-77.6860	Colombia	Narino	#######	TBA
Lupinus mutabilis Sweet	GA70	0.9250	-77.8258	Colombia	Narino	#######	TBA
Lupinus sp. K	GA71	0.9509	-77.8338	Colombia	Narino	#######	TBA
Lupinus sp. J	GA72	0.9509	-77.8338	Colombia	Narino	#######	TBA
Lupinus sp. I	GA73	0.9470	-77.8394	Colombia	Narino	#######	TBA
Lupinus sp. K	GA75	1.0911	-77.6737	Colombia	Narino	#######	TBA
Lupinus sp. J	GA76	1.0916	-77.7212	Colombia	Narino	#######	TBA
Lupinus tauris Benth.	GA77	1.0916	-77.7154	Colombia	Narino	#######	TBA
Lupinus sp. J	GA78	1.1912	-77.1478	Colombia	Narino	#######	TBA
Lupinus alopecuroides Desr.	GA79	4.8959	-75.3487	Colombia	Caldas	#######	TBA
Lupinus humifusus Sessé & Moc. ex G.	GA80	4.9362	-75.3275	Colombia	Caldas	#######	TBA
Lupinus microphyllus Desr.	GA81	4.9362	-75.3275	Colombia	Caldas	#######	TBA
Lupinus santander	GPL2574	-1.6333	-78.0833	Ecuador	Bolivar	#######	TBA
Lupinus tauris Benth.	GPL2723	-3.6833	-79.2833	Ecuador	Loja	#######	TBA
Lupinus huigrensis Rose ex C.P.Sm.	GPL3307	-4.0333	-79.0667	Ecuador	Loja	#######	TBA
Lupinus arvensis Benth.	GPL3563	-3.9833	-79.2000	Ecuador	Loja	892,424	TBA
Lupinus magniflorus C.P.Sm.	GPL3814	-3.5333	-79.2000	Ecuador	Loja	#######	TBA
Lupinus boyacensis C.P.Sm.	GSO05	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus sp. F	GSO08	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus sp. G	NCO12	5.6567	-73.4912	Colombia	Boyacá	#######	TBA
Lupinus boyacensis C.P.Sm.	NCO17	6.0167	-72.9333	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO20	5.6940	-72.8254	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	NCO30	6.3500	-72.3333	Colombia	Boyacá	#######	TBA
Lupinus sp. D	NCO35	6.4167	-72.3500	Colombia	Boyacá	#######	TBA
Lupinus sp. D	NCO36	6.4167	-72.3500	Colombia	Boyacá	#######	TBA
Lupinus triananus C.P.Sm.	NCO41	4.1800	-74.2279	Colombia	Bogotá D.C.	#######	TBA
Lupinus monserratensis	NCO45	5.1250	-74.0006	Colombia	Bogotá D.C.	#######	TBA
Lupinus interruptus	NCO60	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus sp. C	NCO61	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus sp. F	NCO62	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus chocontensis	NCO63	5.1524	-73.6748	Colombia	Cundinamarca	#######	TBA
Lupinus monserratensis	NCO69	4.5703	-74.0306	Colombia	Bogotá D.C.	#######	TBA
Lupinus oceta	NCO49_1	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_1	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_1	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_1	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_1	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_1	5.7061	-72.8125	Colombia	Boyacá	#######	TBA

Lupinus oceta	NCO49_2	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_3	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_4	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_5	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_6	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_7	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_8	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus oceta	NCO49_9	5.7061	-72.8125	Colombia	Boyacá	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_1	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_1	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_1	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_1	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_1	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_1	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_3	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_4	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_5	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_6	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_7	4.7135	-73.8168	Colombia	Cundinamarca	#######	TBA
Lupinus triananus C.P.Sm.	NCO55_8	4.7135	-73.8168	Colombia	Cundinamarca	766,748	TBA
Lupinus cocuy-pisba	PFLP10	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP11	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP12	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP13	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP14	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP15	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP2	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP3	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP4	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP5	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP6	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP7	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP8	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus cocuy-pisba	PFLP9	6.0233	-72.5494	Colombia	Boyacá	#######	TBA
Lupinus alopecuroides Desr.	PLNSI10	4.7976	-75.3872	Colombia	Risaralda	#######	TBA
Lupinus alopecuroides Desr.	PLNSI11	4.7976	-75.3872	Colombia	Risaralda	#######	TBA
Lupinus alopecuroides Desr.	PLNSI2	4.7976	-75.3872	Colombia	Risaralda	#######	TBA
Lupinus alopecuroides Desr.	PLNSI3	4.7976	-75.3872	Colombia	Risaralda	#######	TBA

Lupinus alopecuroides Desr.	PLNSI4	4.7976	-75.3872	Colombia	Risaralda	#######	TBA
Lupinus alopecuroides Desr.	PLNSI5	4.7976	-75.3872	Colombia	Risaralda	#######	TBA
Lupinus alopecuroides Desr.	PLNSI6	4.7976	-75.3872	Colombia	Risaralda	#######	TBA
Lupinus alopecuroides Desr.	PLNSI7	4.7976	-75.3872	Colombia	Risaralda	#######	TBA
Lupinus alopecuroides Desr.	PLNSI8	4.7976	-75.3872	Colombia	Risaralda	#######	TBA
Lupinus alopecuroides Desr.	PLNSI9	4.7976	-75.3872	Colombia	Risaralda	#######	TBA
Lupinus alopecuroides Desr.	PS10720	0.0553	-77.996	Ecuador	Pichincha	#######	TBA
Lupinus alopecuroides Desr.	PS10747	0.1199	-78.2594	Ecuador	Imbabura	#######	TBA
Lupinus nubigenus Kunth	PS11579	-1.082	-78.8421	Ecuador	Cotopaxi	971,141	TBA
Lupinus triananus C.P.Sm.	PSum1	4.2880	-74.2063	Colombia	Bogotá D.C.	#######	TBA
Lupinus triananus C.P.Sm.	PSum2	4.2880	-74.2063	Colombia	Bogotá D.C.	#######	TBA
Lupinus triananus C.P.Sm.	PSum3	4.2880	-74.2063	Colombia	Bogotá D.C.	#######	TBA
Lupinus weberbaueri Ulbr.	RJE052	-8.1425	-78.2739	Peru	La Libertad	#######	TBA
Lupinus mantaroensis C.P.Sm.	RJE054	-8.1467	-78.2033	Peru	La Libertad	#######	TBA
Lupinus prostratus J.Agardh	RJE058	-7.9169	-78.1603	Peru	La Libertad	#######	TBA
Lupinus bimetaliicus #1	RJE091	-7.8036	-77.7675	Peru	La Libertad	#######	TBA
Lupinus dillonii	RJE102	-7.3156	-78.1903	Peru	Cajamarca	#######	TBA
Lupinus sp. nov.	RJE105	-7.0333	-78.5500	Peru	Cajamarca	#######	TBA
Lupinus piurensis C.P.Sm.	RJE110	-7.2167	-78.5000	Peru	Cajamarca	#######	TBA
Lupinus piurensis C.P.Sm.	RJE122	-5.3500	-79.5667	Peru	Piura	#######	TBA
Lupinus semperflorens Hartw. ex Benth	RJE124	-5.3500	-79.5333	Peru	Piura	#######	TBA
Lupinus huancabamba	RJE126	-5.2000	-79.4667	Peru	Piura	#######	TBA
Lupinus piurensis C.P.Sm.	RJE134	-5.3500	-79.5667	Peru	Piura	#######	TBA
Lupinus lindleyanus J.Agardh	RJE141	######	-75.8250	Peru	Lima	#######	TBA
Lupinus praestabilis C.P.Sm.	RJE142	######	-75.8250	Peru	Lima	#######	TBA
Lupinus praestabilis C.P.Sm.	RJE143	######	-75.7972	Peru	Lima	#######	TBA
Lupinus misticola Ulbr.	RJE144	######	-75.8117	Peru	Lima	#######	TBA
Lupinus mutabilis Sweet	RJE151	######	-75.3172	Peru	Junin	#######	TBA
Lupinus breviscapus Ulbr.	RJE154	######	-75.0932	Peru	Junin	#######	TBA
Lupinus huaronensis J.F.Macbr.	RJE155	######	-75.0932	Peru	Junin	#######	TBA
Lupinus fieldii Rose ex J.F.Macbr.	RJE157	######	-75.6350	Peru	Junin	#######	TBA
Lupinus ellsworthianus C.P.Sm.	RJE158	######	-75.5500	Peru	Junin	#######	TBA
Lupinus pinguis Ulbr.	RJE160	######	-75.8380	Peru	Junin	#######	TBA
Lupinus pulvinaris Ulbr.	RJE161	######	-75.8380	Peru	Junin	#######	TBA
Lupinus fieldii Rose ex J.F.Macbr.	RJE163	######	-76.4321	Peru	Junin	#######	TBA
Lupinus humifusus Sessé & Moc. ex G.	RJE165	0.1167	-78.0833	Ecuador	Imbabura	#######	TBA
Lupinus san marcos	RJE168	0.1167	-77.9667	Ecuador	Imbabura	#######	TBA
Lupinus magniflorus C.P.Sm.	RJE169	0.1167	-77.9667	Ecuador	Imbabura	#######	TBA

Lupinus sarmentosus Desr.	RJE174	0.1391	-78.7008	Ecuador	Imbabura	#######	TBA
Lupinus microphyllus Desr.	RJE180	-1.3608	-78.8258	Ecuador	Tungurahua	#######	TBA
Lupinus smithianus Kunth.	RJE183	-1.4000	-78.8667	Ecuador	Tungurahua	#######	TBA
Lupinus santander	RJE187	-1.3708	-79.0528	Ecuador	Bolivar	#######	TBA
Lupinus jamesonianus C.P.Sm.	RJE193	-1.8681	-78.8994	Ecuador	Chimborazo	#######	TBA
Lupinus mutabilis Sweet	RJE197	-1.8100	-78.5778	Ecuador	Chimborazo	#######	TBA
Lupinus nubigenus Kunth	RJE200	-1.4719	-78.8439	Ecuador	Chimborazo	#######	TBA
Lupinus huigrensis Rose ex C.P.Sm.	RJE204	-2.3542	-78.9683	Ecuador	Chimborazo	#######	TBA
Lupinus huancabamba	RJE209	-2.9417	-79.3278	Ecuador	Azuay	#######	TBA
Lupinus tauris Benth.	RJE212	-2.9203	-79.2203	Ecuador	Azuay	#######	TBA
Lupinus jamesonianus C.P.Sm.	RJE214	-3.7928	-79.3031	Ecuador	Loja	#######	TBA
Lupinus arvensis Benth.	RJE221	-3.9850	-79.1872	Ecuador	Loja	#######	TBA
Lupinus arvensis (Silky)	RJE222	-3.6081	-79.2589	Ecuador	Loja	#######	TBA
Lupinus mutabilis Sweet	RJE223	-2.1200	-78.7497	Ecuador	Chimborazo	#######	TBA
Lupinus otto-buchteinii C.P.Sm.	SPS0139	######	-72.0178	Peru	Cusco	#######	TBA
Lupinus chlorolepis C.P.Sm.	SPS1225	######	-72.0175	Peru	Cusco	#######	TBA
Lupinus tominensis Wedd.	SPS1411	######	-72.0095	Peru	Cusco	#######	TBA
Lupinus ellsworthianus C.P.Sm.	SPS1441	######	-72.6550	Peru	Cusco	#######	TBA
Lupinus soratensis Rusby	SPS1466	######	-72.7515	Peru	Cusco	#######	TBA
Lupinus chlorolepis C.P.Sm.	SPS1625	######	-72.8387	Peru	Cusco	#######	TBA
Lupinus ballianus	SPS1755	######	-73.5659	Peru	Cusco	#######	TBA
Lupinus ellsworthianus C.P.Sm.	SPS1839	######	-72.3575	Peru	Cusco	#######	TBA
Lupinus herzogii Ulbr.	SPS2188	######	-72.8397	Peru	Cusco	#######	TBA
Lupinus herzogii Ulbr.	SPS514	######	-72.1407	Peru	Cusco	#######	TBA
Lupinus pulvinaris Ulbr.	SPS970	######	-72.0192	Peru	Cusco	#######	TBA



Click here to download Online Supplemental Appendix_S3 Raw data for seed measurements.xls

Coll_ID	Order	Species	Category	Weight per 25 seeds	weight(g)/1000	Height (mm)	Width (mm)	Depth (mm)
CEH 2333	1	L. pulvinaris	andean	0.11	4.4	1.82	2.34	1.39
CEH 2256	2	L. sp.	andean	0.14	5.6	1.99	2.62	1.41
CEH 2261	2	L. sp.	andean	0.11	4.4	2.00	2.58	1.18
CEH 2324	3	L. sp.	andean	0.08	3.2	1.75	2.59	1.11
CEH 2294	3	L. sp.	andean	0.15	6	2.26	2.96	1.36
CEH 2046	4	L. microphyllus	andean	0.1	4	2.14	2.73	1.32
CEH 2272	4	L. microphyllus	andean	0.15	6	2.15	2.98	1.43
CEH 2319	5	L. sp.	andean	0.16	6.4	-	-	-
CEH 2250	6	L. sp.	andean	0.17	6.8	2.27	3.28	1.39
CEH 2332	7	L. sp.	andean	0.18	7.2	2.15	3.00	1.44
CEH 2276	8	L. sp.	andean	0.19	7.6	-	-	-
CEH 2264	9	L. chrysanthus	andean	0.21	8.4	2.48	3.55	1.48
CEH 2296	10	L. tomiensis	andean	0.21	8.4	2.47	3.42	1.63
CEH 2352	11	L. sp.	andean	0.23	9.2	2.18	3.52	1.69
CEH 2233	12	L. weberbaueri	andean	0.25	10	2.89	4.01	1.54
CEH 2218	13	L. sp.	andean	0.26	10.4	2.59	3.68	1.80
CEH 2247	14	L. sp.	andean	0.26	10.4	-	-	-
CEH 2241	15	L. alopecuroides	andean	0.27	10.8	2.49	3.51	1.90
CEH 2274	16	L. sp.	andean	0.28	11.2	-	-	-
CEH 2248	17	L. sp.	andean	0.33	13.2	2.94	4.07	1.60
CEH 2354	18	L. misticola	andean	0.33	13.2	2.17	3.67	1.46
CEH 2350	18	L. misticola	andean	0.23	9.2	2.23	3.52	1.02
CEH 2365	18	L. misticola	andean	0.18	7.2	2.30	3.35	1.49
CEH 2045	19	L. sp.	andean	0.33	13.2	-	-	-
CEH 2255	20	L. sp.	andean	0.33	13.2	-	-	-
CEH 2239	21	L. prostratus	andean	0.34	13.6	3.22	4.26	1.67
CEH 2000	22	L. austrobicolor	andean	0.32	12.8	-	-	-
CEH 2001	22	L. austrobicolor	andean	0.35	14	-	-	-
CEH 2301	23	L. sp.	andean	0.42	16.8	-	-	-
CEH 2037	24	L. sp.	andean	0.46	18.4	3.09	4.35	2.12
CEH 2325	25	L. sp.	andean	0.46	18.4	3.08	4.45	2.01
CEH 2004	26	L. sp.	andean	0.58	23.2	3.56	4.78	2.05
RJE196	27	L. sp.	andean	0.71	28.4	-	-	-
CEH 2002	28	L. sp.	andean	0.73	29.2	3.75	5.32	2.30

CEH 1997	29	L. piurensis	piurensis	0.81	32.4	3.62	5.31	2.69
RJE113	29	L. piurensis	piurensis	0.77	30.8	-	-	-
CEH 2038	30	L. sp.	andean	0.81	32.4	4.04	5.31	2.18
CEH 2027	31	L. tarus	andean	0.88	35.2	4.31	5.15	-
CEH 2353	32	L. proculaustrinus	andean	1.3	52	-	-	-
CEH 1989	33	L. sp.	andean	1.35	54	5.30	6.85	2.20
CEH 2012	34	L. semperflorens	sempeflorens	0.69	27.6	3.33	4.91	2.44
CEH 2032	34	L. semperflorens	sempeflorens	1.55	62	4.93	6.67	2.82
RJE219	34	L. semperflorens	sempeflorens	1.33	53.2	-	-	-
RJE220	34	L. semperflorens	sempeflorens	0.49	19.6	-	-	-
CEH 2326	35	L. mutabilis	mutabilis	4.99	199.6	7.04	9.33	4.51
CEH 2317	35	L. mutabilis	mutabilis	4.81	192.4	7.21	8.87	4.41
CEH 1999	35	L. mutabilis	mutabilis	5.47	218.8	7.24	8.93	5.13
CEH 2009	35	L. mutabilis	mutabilis	6.79	271.6	7.73	9.25	5.58
CEH 2323	35	L. mutabilis	mutabilis	7.19	287.6	8.25	10.42	4.93
CEH 2302	35	L. mutabilis	mutabilis	6.75	270	8.42	9.93	4.97
RJE151	35	L. mutabilis	mutabilis	8.79	351.6	-	-	-

Appendix S4 - Summary of Archaeological records containing *Lupinus* seeds in the Andean region.

Records listed in chronological order with oldest first.

Date: 10,000 - 8,000 BP.
Location: Ancón, Lima, Peru.
Material: Wild lupin seeds.
Reference: Towle (1961)
Notes: There is no exact date for the *Lupinus* remains, but the date is given for the earliest plant remains at the site. The presence of lupin seeds at the coast suggests movement of humans between the coast and the sierra since all but one of the Andean species of *Lupinus* occur at elevations above 1000 m and the one low elevation species is restricted to southern Peru.

Date: 9,010 - 8,125 BP.
Location: Pachamachay, Junin, Peru.
Material: Wild lupin seeds.
Reference: Pearsall (1980)
Notes: The remains could be the result of intentional or passive moment of seeds into the cave.

Date: 5,000 - 1,400 BP.
Location: Panaulauca Cave, Junin, Peru.
Material: Wild lupin seeds.
Reference: Pearsall (1988)
Notes: *Lupinus* seeds were found infrequently and only from the cave midden. Pearsall (1988) lists lupins as a food source but the basis for this is not explicitly stated. The quantity and location of seed do not suggest that lupins were cultivated or being used for food.

Date: 4,440 - 3,500 BP.
Location: La Chimba, Pichincha, Ecuador.

Material: Wild lupin seeds.

Reference: Pearsall (2003)

Notes: Pearsall (2003) suggests the lupin seeds were harvested from wild species to provide a supplement to the main diet staples of oca, potato and maize.

Date: ca. 3545 BP.

Location: Cotocollao, Pichincha, Ecuador.

Material: Wild or possibly domesticated lupin seeds.

Reference: Pearsall (2003)

Notes: Pearsall (2003) suggests that the presence of lupin seeds represents potential cultivation but there is no information on seed size or abundance.

Date: 1,800 BP.

Location: Mantaro Valley, Peru.

Material: Domesticated *Lupinus mutabilis* seeds

Reference: Hastorf (1993) and personal communication.

Notes: The abundance and size of the seed suggest that domesticated *L. mutabilis* was widely cultivated at the site alongside a set of Andean crops (oca, ullucu, mashua, potato, beans, maize and quinoa). Yields of 119-514 kg/ha have been estimated which is lower than contemporary production (ranges from 300 to 2000 kg/ha, Bellido (1984)).

Date: 1,500 BP.

Location: Nazca tombs, Peru.

Material: Reported as *L. mutabilis* seeds.

Reference: Antúnez de Mayolo (1982)

Notes: Seeds found are large by wild lupin standards but smaller than modern *L. mutabilis*. The size difference could be either and artifact of extreme drying or and indication of early stage domestication.

Date: 540 BP

Location: Mantaro Valley, Peru.

Material: *Lupinus* seeds

Reference: D'Altroy and Hastorf (1984)

Notes: *Lupinus* seeds were recovered associated with ceramic remains in Inca store houses.

Other historical records purportedly referring to *L. mutabilis*

Geyer et al. (2003) reported finding *L. mutabilis* pollen within coprolites taken from three occupants of Tomb 3 at coastal Moche site of Dos Cabezas, Peru. This finding is striking as lupins are insect pollinated and pollen is unlikely to be widely distributed in the environment. Reinhard et al. (2007) commented on the study by Geyer et al. (2003) and suggested that due to analytical errors some or all of the pollen identifications were incorrect. Given this uncertainty, and the difficulty in distinguishing *Lupinus* pollen from the pollen of other Papilionoid legumes, we feel this record lacks credibility.

Two illustrations on Pacheco urns, originating from coastal Nazca valleys and dated at 1,250 - 1,200 BP have been interpreted as depicting *L. mutabilis* (Figs. 27a, b in Yacovleff and Herrera (1934)). Given the lack of an accurate depiction of tarwi seeds in the illustrations, and their lack of overall similarity to tarwi, we feel they do not provide compelling evidence for *L. mutabilis* cultivation in southern Peru.

Literature Cited

Antúnez de Mayolo, S. 1982. Tarwi in ancient Peru. In Proceedings of the1st International Lupin Conference, 1-12, Lima/Cuzco, Peru. International Lupin Association.

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

D'Altroy, T. N., and C. A. Hastorf. 1984. The distribution and contents of Inca state storehouses in the Xauxa Region of Peru. *American Antiquity* 49: 334–349.

Geyer, P. S., T. S. Larson, and L. Stroik. 2003. University of San Diego palynological investigation of the Dos Cabezas giants. *International Journal of Osteoarchaeology* 13: 275–282.

Hastorf, C. A. 1993. Agriculture and the onset of political inequality before the Inka., 1st ed. Cambridge University Press, Cambridge, UK.

Pearsall, D. M. 1980. Pachamachay ethnobotanical report: Plant utilization at a hunting base camp. *In* J. W. Rick [ed.] Prehistoric hunters of the high Andes, 191-231. Academic Press, New York, USA.

Pearsall, D. M. 1988. Interpreting the meaning of macroremain abundance: the impact of source and context. *In* C. A. Hastorf and V. S. Popper [eds.], Current paleoethnobotany: analytical methods and cultural interpretations of archaeological plant remains, 97-118. The Chicago University Press, London, UK.

Pearsall, D. M. 2003. Plant food resources of the Ecuadoran formative: An overview and comparison to the central Andes. *In* J. S. Raymond and R. L. Burger [eds.], Archaeology of Formative Ecuador, 213-257. Dumbarton Oaks, Washington D.C., USA.

Reinhard, K. J., V. M. Bryant, and S. D. Vinton. 2007. Comment on reinterpreting the pollen data from Dos Cabezas. *International Journal of Osteoarchaeology* 17: 531–541.

Towle, M. A. 1961. The ethnobotany of pre-Columbian Peru. Aldine Publishing Company, Chicago, USA.

Yacovleff, E., and F. L. Herrera. 1934. El mundo vegetal de los antiguos peruanos. *Revista del Museo Nacional (Lima)* 3: 241-322.