SYSTEMATIC STUDIES OF THE SWEET POTATO AND ITS WILD RELATIVES

PABLO MUÑOZ RODRÍGUEZ

GREEN TEMPLETON COLLEGE

DEPARTMENT OF PLANT SCIENCES

THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
TRINITY TERM 2018

Cuando nuestra mano libera un objeto y éste cae al suelo, no asociamos la fuerza gravitatoria con la bandera británica. En los libros astronómicos, Ío, Europa, Ganímedes y Calisto, las cuatro lunas jovianas descubiertas por Galileo, no engrosan el territorio de Italia. La relatividad no es judía, la comprensión del sistema sanguíneo circulatorio no lleva patente española y el submarino del figuerense Monturiol pertenece a la historia universal del ingenio humano, no sólo a la del Alto Ampurdán. Al conocimiento le sucede igual que a las aves migratorias: no entiende de fronteras.

~

[*When our hand releases an object and it falls, we do not associate gravity to the British flag. In Astronomical books, Io, Europe, Ganymede and Calisto, the four Jovian moons discovered by Galileo, do not form part of the Italian territory. Relativity is not Jewish, the understanding of the circulatory system is not registered under a Spanish patent, and the submarine designed by Monturiol, born in Figueres, belongs to the universal history of human knowledge, not only to the history of the Alt Empordà region. Is knowledge like migratory birds: it does not recognise borders.*]

Miguel Artime. Revista Naukas, 2013

TABLE OF CONTENTS

ABSTRACT

Phylogenetic studies on the genus *Ipomoea*, and specifically on the species closely related to the sweet potato, have been hindered by limited taxon and character sampling. Any study that aims to investigate the phylogenetic relationships within a megadiverse genus such as *Ipomoea*, with an estimated 800 species, must rely on a comprehensive data set that covers as much diversity as possible, in order to provide accurate and robust phylogenetic inference.

In this thesis, I present a comprehensive phylogenetic study of the genus *Ipomoea* and of the group of species closely related to the sweet potato using genomic-scale data. The aim of this study was to provide a phylogenetic framework with which to understand the diversity existing within the genus and to inform taxonomic decisions. The phylogenies presented here identify several recurrent patterns in the evolution of *Ipomoea*, for instance the multiple origin of storage roots and the existence of multiple episodes of long-distance dispersal by natural means.

I also address several questions pertaining to the sweet potato that have been a matter of debate for decades but remained unanswered until recently. Our genome-scale data facilitated comprehensive phylogenies for all wild species that are most closely related to the sweet potato. Our phylogenies resolve that sweet potato is monophyletic and had a single origin, as well as identified the wild species that is its closest relative: *Ipomoea trifida*, with all other extant species more distantly related. In addition, our studies corroborate the existence of two sweet potato chloroplast lineages and infer that one of them resulted from a hybridisation between sweet potato and *I. trifida* following species divergence.

Finally, I also address a question that has been of interest for over two centuries: the presence of the sweet potato, an American crop, in Polynesia in pre-European times.

ACKNOWLEDGEMENTS

Oxford has been the scene of some of the most rewarding moments of my life. However, this adventure started a long time before I arrived in *the city of the dreaming spires*. It is therefore necessary to express my gratitude to those who have been part of my Oxford experience but also to those who helped me to get here.

Above all, thanks to my parents, Ángeles and Pablo, who shaped the person that I am today. Your support, the education and endless love that you gave me is what got me this far. I will forever be indebted to you.

To my brother, Mario, who (unintentionally) became an example for me and motivated me to go a step further on so many occasions.

To Ana, who years ago decided to change her life to come with me and become the co-author of a beautiful story. Your smile every morning is a gift.

To José María Cardiel, my *botanical father* and now good friend. Thank you for the time you devoted to training the botanist that I was to become and for giving me the opportunity to work alongside you during many good years.

Also, to Elena Dorda, Consuelo Cebolla, Mercedes García-Antón, Juan Carlos Moreno and the other good teachers at the Botany Unit–UAM, whose teaching stimulated my interest in plants, and to Javier Fuertes for his support and advice.

To Juan Carlos Moreno; to Leopoldo Medina and the *Flora Ibérica–Anthos* team; and to Paco Pando and the GBIF Spain team, for their support (and funding) during several years before starting my DPhil. It is likely that this thesis would not have been possible without your support to continue working in science during those years.

To my good friends in Spain: Bea, Isidro, Mike, Fabio, Miguel, Espe and many others, for the moments together and for their encouragement and precious support in moments of uncertainty.

In Oxford, thanks to Lee Sweetlove, Gem Toes-Crichton and Gail Preston for their help to secure a place and funding at Oxford. Thank also to Gail, James Wright and all other people involved in the Interdisciplinary Bioscience DTP programme for their continuous help and the very useful training provided during my thesis.

To the agencies and entities that provided funding for this project: Biotechnology and Biological Sciences Research Council (BBSRC), Manor Fresh, the BBSRC GCRF-IAA fund, The Leverhulme Trust, and The John Fell Fund. Also, the Santander Travel Award contributed towards the costs of my stay at Oregon State University and the Synthesys European Initiative funded my visit to Paris herbarium.

To Gem Toes-Crichton, Nick Kruger, Helen Prescott, Pauline White, Stephen Harris, Serena Marner, Dharshani Perera, John Baker, Timothy Wroe, Sandy West and all the wonderful members of the staff at Plant Sciences for making my everyday life much easier. Every single morning, I felt this department was the place for me.

To Beth Williams, for her help with the basics of lab work and for the extensive and excellent work that she did before I started, which facilitated the success of this project enormously.

To John Wood, whose profound knowledge of *Ipomoea* brought light to the discussions about the results of my analyses. It is a privilege to work alongside you.

To Steve Kelly for his help with bioinformatics during these years and especially at the beginning of the project, when I had no idea how to start working with so much data.

To Aaron Liston, Kevin Weitemier, Brent Kronmiller and Lucas Longway for their help with the sequencing, assembly and analysis of genomic data, and to Aaron and Kevin for hosting me at Oregon State University at the beginning of this project. Also to Richard Cronn, for use of his lab in preparing DNA from the Banks & Solander specimen for sequencing and subsequent analysis.

To Noelle Anglin, Dave Ellis and the staff of the International Potato Center for hosting me during my visit and the continuous collaboration afterwards.

To David Swofford for his help with SVDQuartets and the coalescence methods.

To all herbarium curators for allowing us to sample material hosted at their institutions; especially to the BM herbarium for allowing us to sample Banks and Solander's specimen and to Mark Carine for his help throughout the process.

To the staff of the Bodleian library —*«an Ark to save learning from deluge»*— and especially to Anne Catterall for providing access to, and help with, all the references I consulted during this project.

To Carmen Sánchez, Bruno Catarino, Javier Agustí, Zoë Goodwin, Matteo Tanadini, Louise Hill, Sean Tuck, Cicely Marshall, Gail Stott, Tom Carruthers and Alex Sumadijaya for creating such a nice atmosphere in which so many hours were spent. Also thank you, Tom and Alex, for your help in the lab when time was a limiting factor. Furthermore, Tom, thank you for listening to me talking about the sweet potato again and again, for your incredible help with divergence-time estimates and for the countless cups of not-so-good-coffee. Those were necessary too, I guess.

To the Society of Spanish Researchers in the United Kingdom, to the members of the Oxford constituency and, above all, to Carmen, for all the time and effort devoted to the activities that we organised together. Although, at times, the Society took most of my energy, I always felt it was worth it.

Last but not least, to Robert Scotland, for offering me a place in his group and for his supervision and training. Thank you, Robert, for all the time you devoted to our project, for the continuous support and opportunities and for the endless hours of scientific discussion. There are no words to describe my gratitude.

1 | INTRODUCTION

Carl Linnaeus described the plant genus *Ipomoea* L. and seventeen species in the first edition of his *Species Plantarum* (Linnaeus, 1753: 159). Since then, more than 1,800 names associated with the genus have been published (Missouri Botanical Garden, 2018), making *Ipomoea* the largest genus in the family Convolvulaceae. *Ipomoea* includes herbs, shrubs, lianas and trees and is present in all tropical and subtropical regions of the world, from sea level to 4,000 meters and from tropical rain forests to semi-desert coastal environments. The genus is also present in other more temperate regions as far North as Canada and several widespread species have a worldwide distribution. Two species have global importance as crops: the sweet potato, *Ipomoea batatas* (L.) Lam., and the *kangkong* or water spinach, *I. aquatica* Forssk. Several species are cultivated in gardens —the group is commonly known as morning glory— and some species have become invasive, for example *I. cairica* (L.) Sweet, *I. indica* (Burm.) Merr. and *I. purpurea* (L.) Roth. In addition, multiple studies have identified anti-tumorous properties of different compounds obtained from *Ipomoea* leaves and roots (see review in Meira et al., 2012).

Some genera in the family Convolvulaceae have been subject to comprehensive nomenclatural and/or taxonomic studies in recent times —e.g. *Argyreia* Lour. (Staples and Traiperm, 2017; Chitchak et al., 2018), *Convolvulus* L. (Wood et al., 2015a) or *Seddera* Hochst. (Demissew and Mill, 2009) —, but most genera are poorly known taxonomically, and few studies include molecular data. In the case of *Ipomoea*, hundreds of studies have been published in the last two centuries but few of them addressed the taxonomy of the whole genus. Most taxonomic studies focussed on small groups of species,

frequently related to the sweet potato or widespread species[1], or a single taxon[2]. A few authors also addressed the genus at the national or regional level; although several of those more extensive treatments date back to the 19th century or the first half of the 20th century[3] and most of these, with a few notable exceptions (Austin, 1975a; van Ooststroom, 1953; Wood et al., 2015b), are mere checklists of species or provide only incomplete information about the taxa included. The only global taxonomic review of the genus was published by Choisy more than 170 years ago, in De Candolle's *Prodromus* (Choisy, 1833, 1838, 1845), and included 282 species. Choisy's work is now certainly outdated, considering the current estimate of between 600 and 800 extant species (Austin, 1975a; Austin and Huáman, 1996; and our own work). In summary, information about the genus *Ipomoea* is abundant but fragmentary, with a few species very well known for many aspects of their biology but most other species never having been studied properly. The level of taxonomic knowledge of the group is generally poor and recent estimates indicate that at least 50% of all herbarium collections of *Ipomoea* could be misidentified (Goodwin et al., 2015). In this context, a global, comprehensive taxonomic and phylogenetic review of the megadiverse genus *Ipomoea* is urgently needed.

This D.Phil. study is part of a more extensive study in which we aim to discover and describe the diversity existing within *Ipomoea*. We integrate the analysis of DNA sequence data as part of our taxonomic studies, seeking to accelerate the pace at which

[1] See for example Abdel Khalik, 2013; Abdel Khalik et al., 2012; Austin, 1992, 1997; Austin and McDonald, 2014; Das, 2011; Folorunso, 2013; Folorunso et al., 2013; Gunn, 1969; Jayeola and Oladunjoye, 2012; McDonald, 1987, 1995, 2001; McDonald and Austin, 1990; Ogunwenmo, 2003, 2008; Raya Rao and Leela, 2008; Yatsckievych and Mason Jr., 1984.

[2] See for example Austin, 1991, 2011, 2013a, 2013b; Austin and McDonald, 2014; Howard and McDonald, 1995; Kaur et al., 2016; Kenyan, 1928; Marderosian, 1965; Staples et al., 2005a, 2014; Swamy and Ramana, 2018; Valva and Sabato, 1983.

[3] See for example Austin, 1975a; Austin and Bianchini, 1998; Austin and Huáman, 1996; Austin et al., 2015; Folorunso et al., 2013; Gray, 1878; Grisebach, 1864; Hallier, 1894; House, 1908; Madani and Majbour, 2017; Meisner, 1869; van Ooststroom, 1953; Staples et al., 2005b; Stefanovic et al., 2003; Thulin, 2003; Wood and Scotland, 2017a; Wood et al., 2015b.

taxonomy is carried out (Scotland and Wood, 2012) and to produce what has been termed "Foundation monographs" (Scotland and Wood, 2012; Williams et al., 2014). Foundation monographs prioritise species delimitation and identification of new taxa above other aspects of traditional monographs, while at the same time using molecular data to inform taxonomic decisions and demonstrate phylogenetic relationships. This approach proved effective in a pilot study conducted on *Convolvulus* (~ 190 species) for which a foundation monograph was produced in only twelve months (Wood et al., 2015a). In the case of *Ipomoea*, our studies during the last four years have helped to increase knowledge of the group considerably, especially in the American continent —home to two thirds of the estimated number of species— with ten papers published in the last three years (Muñoz-Rodríguez et al., 2018; Wood and Scotland, 2017b, 2017c, 2017a; Wood et al., 2015b, 2016a, 2016b, 2017a, 2017b, 2017c) and a foundation monograph of the genus in the Americas currently in preparation.

In addition, the comprehensive taxon and character sampling that characterises our studies, together with the extensive knowledge accumulated during these years, allowed us to pay special attention to the most economically important member of the genus: the sweet potato. Despite two centuries of studies and it being one of the most important crops in the world, several questions pertaining to its origin have never been satisfactorily answered. Our integrative approach allowed us to investigate those questions to understand more fully the origin and evolution of the crop.

I joined Robert Scotland's group with the aim of studying the group of species closely related to the sweet potato. For that reason, I have devoted most of my time to the study of this group of species and of the origin and evolution of the crop. In addition, I have also contributed to the overall project, fundamentally with the elaboration of phylo-

genetic tools with which to inform taxonomic decisions and to investigate the evolutionary history of *Ipomoea*. This dissertation presents my contribution to our studies on *Ipomoea* during the last four years.

I structured this thesis in six chapters, including this introduction and general conclusions. Given the diversity of methods that underlie the results, together with the size and characteristics of the data sets analysed, I considered it necessary to dedicate the **second chapter** to explain and discuss the methodology in detail. The following three chapters present a detailed introduction to each of the topics covered, followed by the discussion of our results.

As explained above, one of my objectives was to generate molecular phylogenies that inform taxonomic decisions, but also to provide a phylogenetic framework that allows studying the evolution of *Ipomoea* at multiple levels. In the **third chapter**, I present the phylogenies that constitute this framework and discuss our results in relation to the entire genus *Ipomoea* and the tribe *Ipomoeeae*. Furthermore, I show how the global study of the genus provides important information with which to understand the evolution of the sweet potato.

In the **fourth chapter**, I focus on the origin and evolution of the sweet potato and on the relationship between the species in the group to which this important crop belongs. I first present a review of previous studies on this topic to put our research in context. Subsequently, I present the results of our studies and provide answers to several questions pertaining to the origin and evolution of the sweet potato that have never been properly answered.

Finally, in the **fifth chapter**, I focus on the geographical origin, distribution and dispersal of the sweet potato and, especially, on its alleged presence in Polynesia in an-

cient times. I review the extensive literature, encompassing different disciplines, and present and discuss the implications of our genomic study with the aim of clarifying the current knowledge about these issues. A majority of authors working on the later question have taken for granted the existence of human contacts between America and Polynesia in pre-European times and some claim that their existence is well established (see for example (Clarke, 2009; Heyerdahl, 1952; Horsburgh and McCoy, 2017; Jett, 2017; Matisoo-Smith and Ramírez, 2010). However, definitive confirmation of that theory is still lacking and recent data, including the results of this thesis, pose further doubts about the validity of the dominant theory.

I faced two main challenges during this thesis. The first and most important was to develop a system to manage and analyse the extraordinary amount of DNA data available generated during the project (genomic data for hundreds of specimens, thousands of DNA barcode sequences, the results of the analyses, information gathered from herbarium collections, etc.). The second challenge that I faced during this time was the sheer volume of literature and information generated about *Ipomoea*, especially about the sweet potato, in the last two and a half centuries. Not only taxonomists and systematists but also ethnobotanists, ecologists, geneticists, geographers, archaeologists, linguists, historians and many other scholars of different disciplines have written in regard to the origin, evolution and utilisation of the sweet potato. There were a few references that I was unable to consult and others that surely passed unnoticed. Even so, I hope the reader of this work will acquire a good perspective on the origin and evolution of the sweet potato and related matters.

2 | METHODOLOGY

Ipomoea has been the subject of several phylogenetic studies in the last decades. Despite using different approaches and materials, all these studies are affected by at least one of two problems: limited taxon sampling and/or a small data set of characters. Manos and colleagues (2001), for example, included 45 taxa and one molecular marker in their study, whereas Eserman and colleagues (2014) and Park and colleagues (2018) used whole chloroplast genomes but included only 29 and 35 samples within *Ipomoea* respectively[4]. As explained later, a limited amount of data is the norm in all studies of this group. The consequence is that results are not robust nor able to provide insight into the phylogeny of *Ipomoea* as a whole.

The data sets that underlie this thesis aimed to overcome the limitations that hindered previous studies and constitute the most comprehensive sampling of *Ipomoea* and the sweet potato to date. Our results demonstrate that extensive taxon and data sampling covering as much diversity within *Ipomoea* as possible is necessary to understand better the phylogenetic relationships within the genus.

This chapter provides a detailed account of the methodology used in our phylogenetic analyses, the results of which are discussed in following chapters. In addition to designing and running all phylogenetic analyses except the estimate of divergence times (see section 2.8 in page 43), I generated 600 DNA barcode sequences (*ITS* and *rbcL-rpl32*) from herbarium collections and living material; did the last stages of lab work for Hyb-Seq (blue boxes in Figure 2.5) at Oregon State University (OSU) and worked with Brent Kronmiller from OSU on the assembly of the Hyb-Seq data; and extracted the DNA and assembled the chloroplast genome and the nuclear regions of Banks & Solander's specimen.

[4] In the study by Parker and collaborators (2018), 29 of the 35 samples are those from Eserman's study and only 6 samples were new.

2.1 | PLANT MATERIAL

We estimate that the world's herbaria house more than 100,000 *Ipomoea* specimens[5]. These specimens, held in major institutions but also in smaller regional collections, constitute a remarkable source of data for molecular studies. The use of herbarium material is not only a very important source of plant material for DNA studies, but also is the only practical way to undertake a comprehensive study of a genus the size of *Ipomoea* in a reasonable time.

For these reasons, most material used in our studies was obtained from herbarium specimens from the following 47 herbaria (acronyms according to Thiers, 2018): AAU, ARIZ, BISH, BM, BOLV, CEN, CIP, CUZ, E, F, FCQ, FFTG, FRI, GA, GOT, HAJB, HEPH, HSB, HUEFS, IBSC, IEB, IPA, JPB, K, KEP, KUN, L, LPB, MA, MBM, MEXU, MICH, MO, NY, OXF, P, PC, RB, S, SAN, SELU, SING, TEX, US, USDA, USM, USZ. We successfully sequenced 1,560 herbarium specimens, 88.5% of them collected in the last 50 years. 22 samples came from pre-20th century collections and the oldest collection dates to 1769 (Figure 2.1). This specimen, collected by Joseph Banks and Daniel Solander in the Society Islands, is the first sweet potato collection known from Polynesia and one of the oldest sweet potato collections worldwide[6].

A complementary but essential source of material was the germplasm collection of the International Potato Center (CIP). Founded in 1971 as part of the CGIAR consortium, CIP is a global initiative present in 20 countries in Africa, America and Asia. The germplasm bank at its headquarters in Lima, Peru, hosts the largest sweet potato germplasm collection in the world, with over 5,500 accessions, and is an important resource for any study of sweet potato breeding and improvement. For this reason, CIP

[5] There are c. 50,000 records listed in GBIF and we estimate this is at the very least less than half of the world total.

[6] See Chapter 5 for an extensive discussion involving this specimen.

collections were especially important regarding the sweet potato samples included in our study. We obtained DNA samples from 73 sweet potato varieties from across the American continent and the Old World, including two commercial varieties frequently used in breeding programmes, *Beauregard* and *Tanzania* (Dorcus Gemenet, pers. comm.).

Finally, we obtained 91 samples from collections made during the course of this project (after 2014) and a minimum amount of material from plants grown by ourselves in the greenhouse at the department of Plant Sciences and at the Oxford University Botanical Garden. Also, we incorporated the whole chloroplast genome sequences generated by Eserman and colleagues (2014) of species that we had not sampled for Hyb-Seq. Passport data of all specimens is provided in Supplementary File 1.

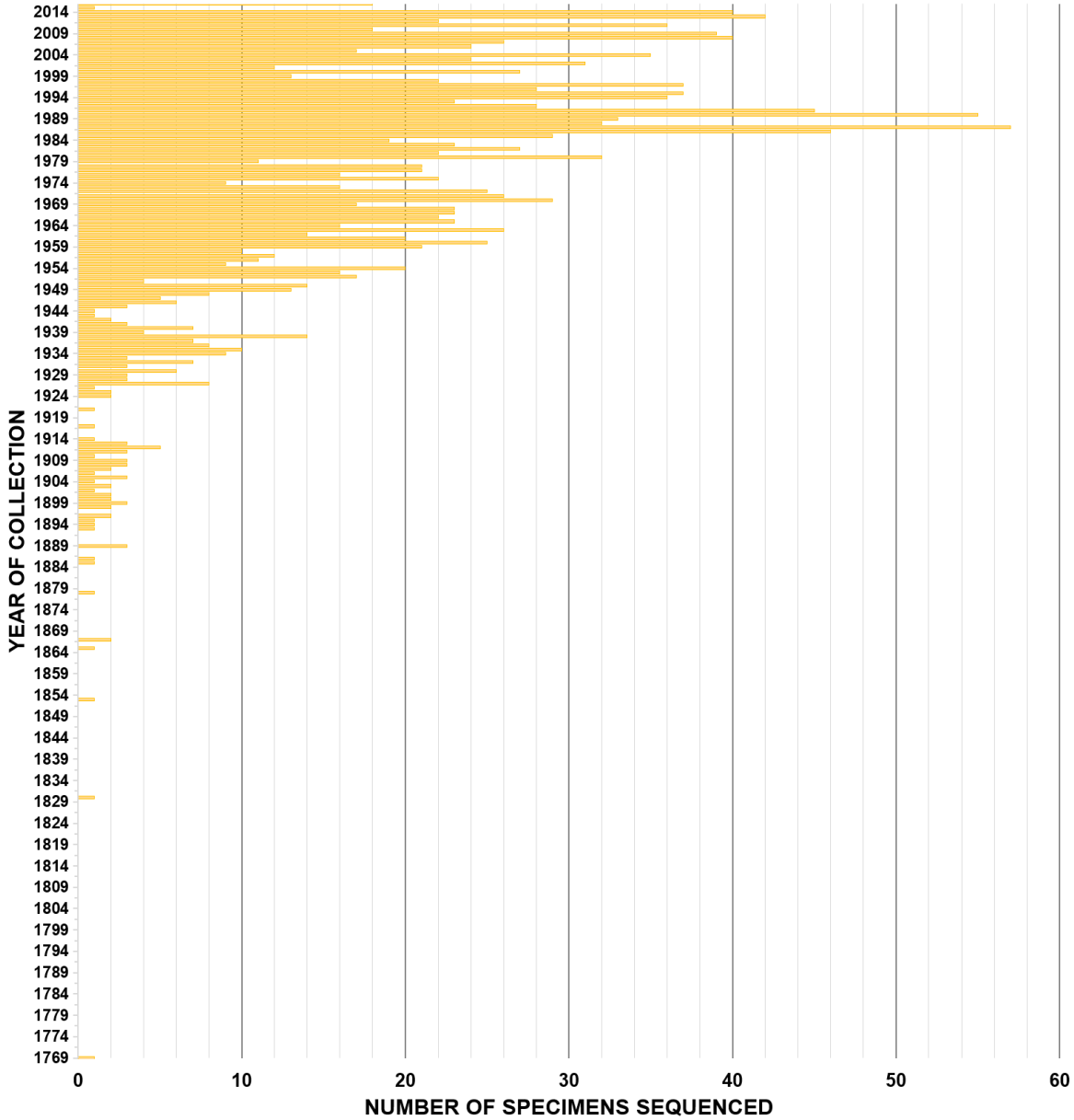


FIGURE 2.1. Number of specimens sequenced per year of collection. N = 1560.

2.2 | DATA SETS

Three different types of data underlie the phylogenetic analyses presented in this thesis: DNA barcodes, regions of nuclear coding DNA and chloroplast genomic data, each of them composed of different sets of samples and designed for different purposes.

2.2.1 | DNA barcodes

The analyses using DNA barcodes are based on 3,035 sequences from 1,560 specimens (Figure 2.2). Before the start of my D.Phil. project, Bethany R.W. Williams tested

the informativeness of different DNA barcode regions for our studies on *Ipomoea* (Baldwin et al., 1995; Feliner and Rosselló, 2007; Shaw et al., 2007). Among them, she found three regions that provide sufficient levels of phylogenetic resolution and between species discrimination: the nuclear ribosomal Internal Transcribed Spacer (*ITS*), either complete or the *ITS2* fragment only (Figure 2.3), the chloroplast gene *Maturase K* (*matK*) and the chloroplast intergenic region *trnH-psbA*. She subsequently obtained a large number of sequences from across *Ipomoea* and related genera.

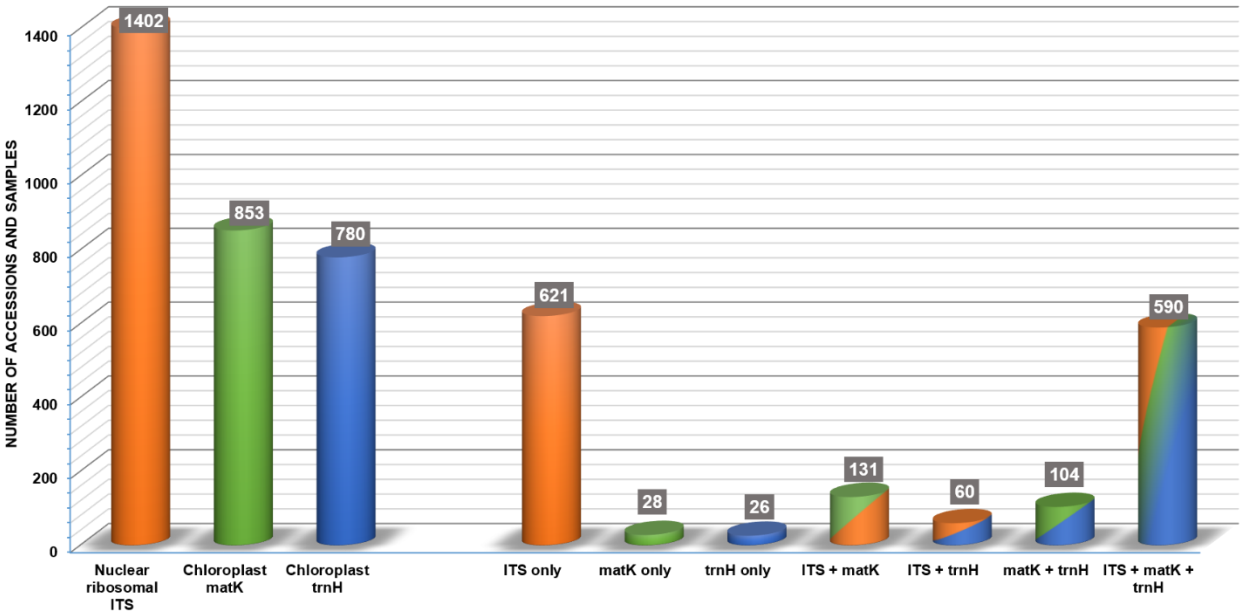


Figure 2.2. We sequenced 1,560 specimens of *Ipomoea* and related genera for one or more DNA barcode regions. This sampling represents around 470 species. The three columns on the left indicate the total number of sequences obtained for each barcode. The seven columns on the right indicate number of specimens for which we obtained one or more barcodes.

Later in the project, we found that the chloroplast non-coding region *rpl32-trnL* provides a good level of variation to estimate the position of a specimen in the group of species closely related to the sweet potato (Batatas group) —as opposed to the three regions aforementioned, which result in largely unresolved polytomies in that group—. Thereafter, we used this region to infer the position of several specimens within this group.

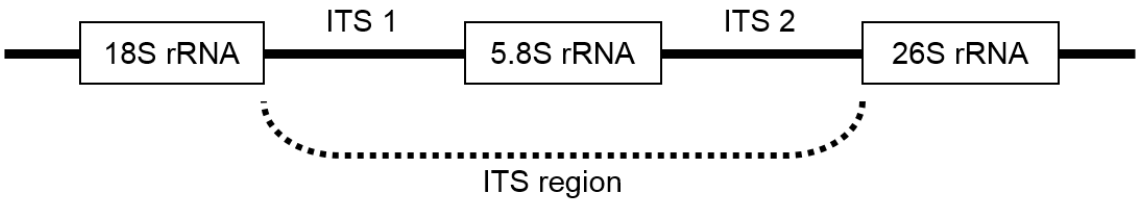


Figure 2.3. Structure of the nuclear ribosomal Internal Transcribed Spacer (*ITS*) in Angiosperms. The non-coding *ITS* region is flanked by two coding regions, 18S and 26S rRNA, and has another coding region within, 5.8s rRNA.

2.2.2 | High-Throughput Sequencing

In addition to DNA barcoding, we sequenced 384 samples (four 96-well plates) using high-throughput sequencing to obtain a much larger amount of sequence data. Several motivations influenced the design of our data set, the most important being the lack of between-species resolution in the sweet potato group using DNA barcodes (see discussion in Chapter 4) and the poor taxon sampling that characterised previous phylogenetic studies of *Ipomoea*. For these reasons, our taxon sampling in this case was centred on the sweet potato group and more than 50% of our 384 samples were from species in this group. In addition, we aimed to cover as much diversity within the whole genus as possible, with two further objectives: to test the accuracy of our studies based on DNA barcodes and to inform our global studies on the genus *Ipomoea*. Figure 2.4 shows a phylogeny constructed using *ITS* sequences; the red tips indicate the samples that we also sequenced using high-throughput sequencing, accounting for around 200 species across *Ipomoea*, a quarter of the estimated total number; furthermore, almost all clades retrieved in the *ITS* phylogeny were sampled for high-throughput sequencing. This large and diverse taxon sampling, complemented with the barcode phylogenies, provides an extraordinary amount of information with which to study the evolution of *Ipomoea*. I included all our 384 samples in the analysis of nuclear coding regions, whereas I included only 379 of them in the chloroplast analyses due to errors detected during the assembly process of five samples. I defined subsets of samples depending on the scope of the analysis.

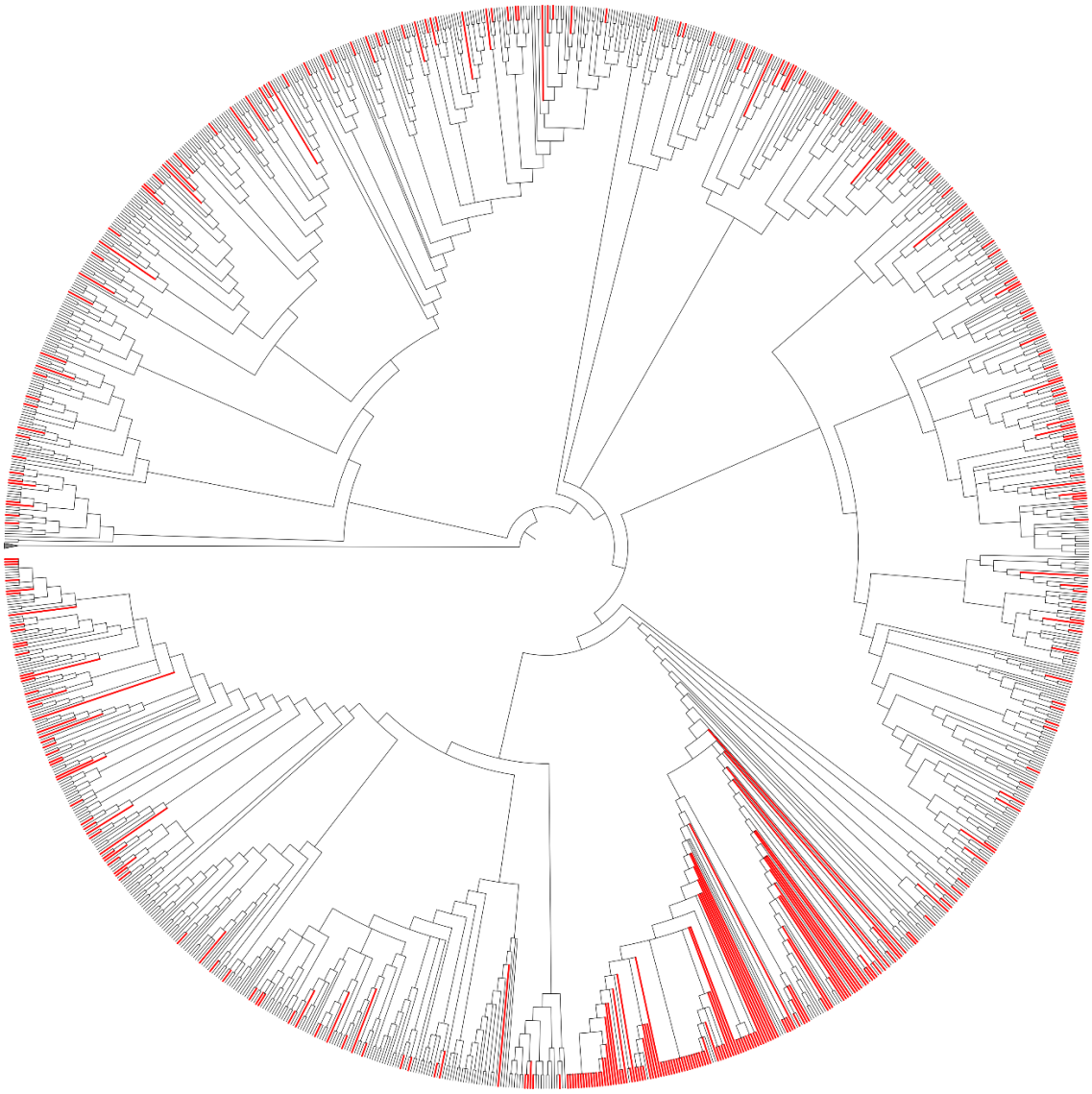


Figure 2.4. Phylogeny of *Ipomoea* inferred using *ITS* sequences**. Tips in red** indicate the 384 specimens from across the whole of *Ipomoea* that were also sequenced using Hyb-Seq.

2.3 | LAB WORK

2.3.1 | DNA extraction

We extracted DNA from herbarium specimens using the Plant Tissue Mini protocol for QIAGEN DNeasy Plant Mini Kit (QIAGEN), and from germplasm collections at CIP using the CTAB method (Doyle and Doyle, 1990). Apart from CIP samples, we extracted DNA from almost 2,500 herbarium specimens. Bethany Williams extracted DNA from almost 1,650 specimens, I added around 600 more and the rest were extracted by other

members of our research group. 1,560 samples, 62% of the total samples extracted, were successfully sequenced for barcoding (Figure 2.2).

We quantified the DNA concentration in the samples using a Qubit fluorometer and only used for high-throughput sequencing those with DNA concentration higher or slightly lower than 5 ng/µl. Because of the relevance of the specimen, we sequenced Banks and Solander's Polynesian collection despite having a DNA concentration lower than desired (2ng/µl).

2.3.2 | Sequencing of barcoding regions

We obtained DNA barcode regions using Sanger sequencing. We obtained the largest number of sequences for *ITS* (1,402) because it is the most informative region for resolution and support among the three barcodes used, but the number of sequences for *matK* and *trnH-psbA* was also considerable (Figure 2.2).

We assembled forward and reverse reads into a consensus sequence, for each barcode region, using either Mega v.5.0 or v.6.0 (Tamura et al., 2013) or Geneious v.9.1.2 (Kearse et al., 2012). All *ITS* sequences are provided in Supplementary File 2.

2.3.3 | High-Throughput Sequencing. Hyb-Seq[7]

Hyb-Seq is a high-throughput sequencing technique that allows the collection of low-copy number regions and high-number copy regions (e.g. chloroplast genomes) simultaneously by combining target enrichment and genome skimming (Weitemier et al.,

[7] The text presented in the 'High-Throughput Sequencing. Hyb-Seq', 'Phylogenetic analysis using whole chloroplast genomes', 'Phylogenetic analysis using nuclear regions', 'Analyses of population structure' and 'Time-calibrated phylogenies' sections, as well as part of the results presented in Chapters 4 and 5, have been incorporated and, in some parts, expanded from what we presented in our April 2018 paper in *Current Biology* (Muñoz-Rodríguez et al., 2018). The content of that paper is included here in agreement with Elsevier's policy that allows authors of papers to include the content of a paper in a thesis, provided it is not published commercially: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights.

2014). We designed the sequencing strategy for our samples in collaboration with Aaron Liston and Kevin Weitemier, developers of the Hyb-Seq strategy (Weitemier et al., 2014). Subsequently, Kevin Weitemier designed the baits and prepared the samples at their lab in Oregon State University (OSU), Corvallis (Oregon, USA) and I spent two weeks there in January 2016 to get involved with the final stages of the lab work. Sequencing was then conducted at the OSU Center for Genome Research and Biocomputing (CGRB) and I worked with Brent Kronmiller, a member of CGRB staff, to produce the assembly of the whole chloroplast genomes and the nuclear regions. Figure 2.5 shows the main steps in the process.

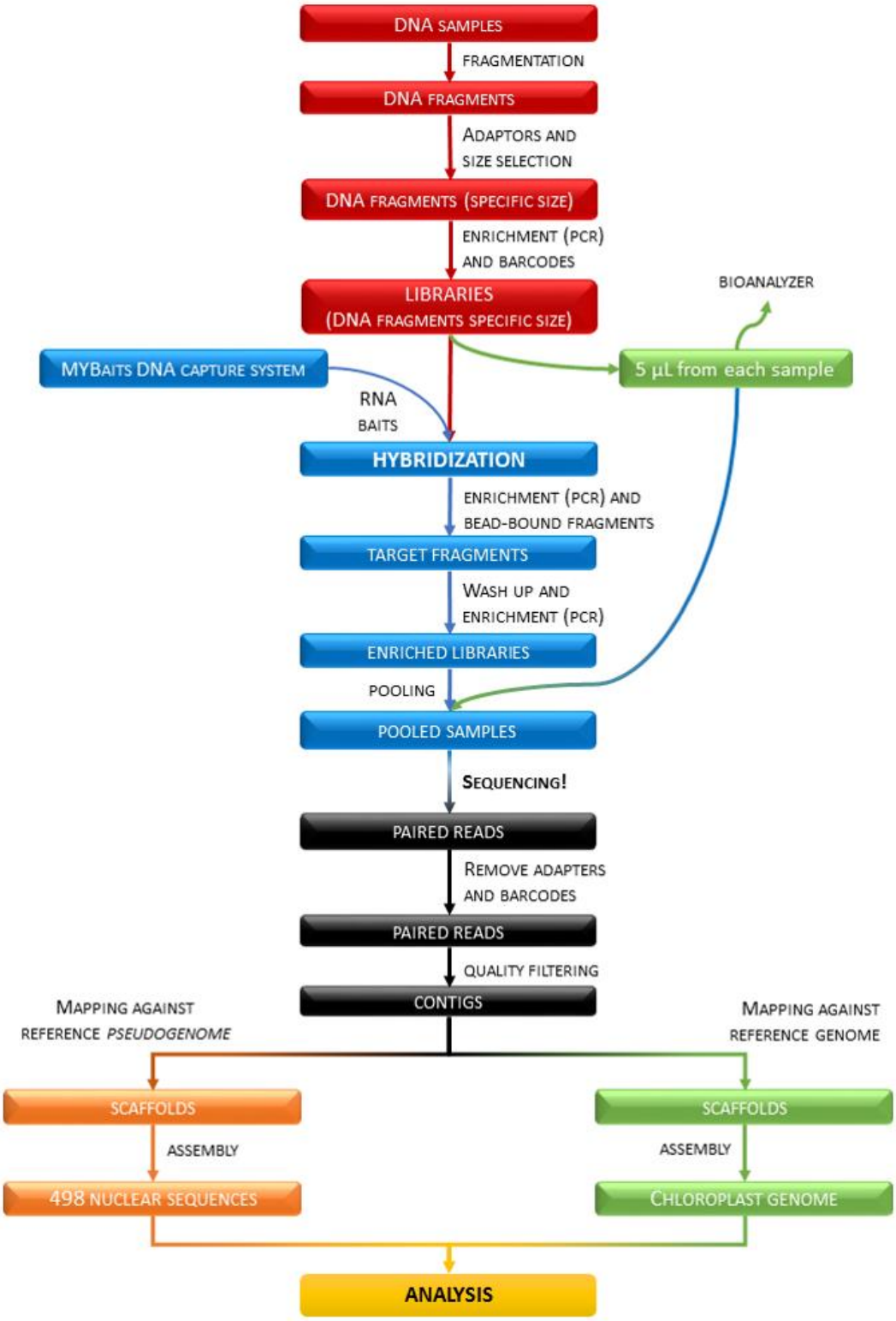


Figure 2.5. Chart flow showing the main stages in the Hyb-Seq workflow as implemented in this study.

2.3.3.1 | Character sampling and target capture probes

We developed probes targeting 605 putative single copy nuclear regions of our samples through comparison of genomic data from *Ipomoea lacunosa* L. (unpublished but kindly provided by Mark Rausher from Duke University; assembled into scaffolds, N50=46047) and of coding sequences (CDS) of *Solanum tuberosum* ITAG2.4 (Fernández-Pozo et al., 2015).

Regions between *Ipomoea* and *Solanum* with a one-to-one match at 70% identity along at least half the length of a *Solanum* CDS were filtered to retain *Ipomoea* loci that were at least 1000 base pairs. Along these loci, 100 bp RNA probes were developed by MycroArray (Ann Arbor, MI), excluding probes with GC content < 25%.

2.3.3.1.1 | Design of target regions

A BLAT comparison was performed between the *Ipomoea* genome and the *Solanum* CDS regions (minIdentity=0.85). From the resulting file, we removed entries where the hit was 50 base pairs or less. We subsequently filtered this trimmed file to find *Solanum* loci that appeared only once. From those single hits, entries were retained where the CDS hit was > 50% the total length of the CDS, and the length of the *Ipomoea* genome was between 959 and 10,000 base pairs. Also, entries where different *Solanum* CDS hit identical spots in *Ipomoea* were removed. This procedure resulted in 605 *Solanum* CDS matching 506 *Ipomoea* scaffolds.

2.3.3.1.2 | Design of baits

We designed *100mer baits with 1x tiling density, 20,020 baits in total. Each bait candidate was blasted against the *I. lacunosa* genome and filtered by melting temperature (the temperature where 50% of the molecules are hybridized to their complementary strand) as follows. For each blast hit, a melting temperature (Tm) of hybridization was predicted using salt and temperature conditions matching the hybridization conditions.

Then, for each bait candidate, the number of blast hits with Tm falling into the following six temperature bins were counted: 40–60 ºC, 60–62.5 ºC, 62.5–65 ºC, 65–67.5 ºC, 67.5–70 ºC and above 70 ºC. During capture, optimal hybridization temperature is 65 ºC. Bait candidates are accepted if they satisfy one of these conditions:

- No blast hit with Tm > 60 ºC.
- No more than two hits at 65-67.5 ºC and ten hits at 62.5-65 ºC, and two neighbour candidates on at least one side being rejected.

No more than a single hit at or above 70 ºC, no more than one hit at 65-67.5 ºC and two hits at 62.5-65 ºC and two neighbour candidates on at least one side being rejected.

2.3.3.2 | Hybridization and DNA sequencing of the main set of samples

We prepared DNA libraries using the NEB's Ultra DNA Library Prep Kit for Illumina v.3.0 (New England BioLabs). We implemented target enrichment using MYBaits (MYcroarray, 2015) to capture the nuclear regions of interest following the protocol described by Weitemier and collaborators (2014) and using Beckman Coulter Agentcourt AMPure XP for product purification. We sequenced a 1:1 mixture of target enriched and unenriched libraries, in order to obtain the chloroplast and nuclear ribosomal *ITS* region with genome skimming (Straub et al., 2012). Sequencing was conducted using the Illumina HiSeq 3000. We trimmed the sequences for Illumina adapters and for quality, Q15 on the left and Q10 on the right of the reads. We obtained 100 bp paired reads. Exact read duplications were reduced down to the expected target coverage level based on sequence coverage.

2.3.3.3 | Assembly of nuclear regions

We conducted a three-stage assembly process to create sequence contigs that correspond to the target sequences: first, we used YASRA (Ratan, 2009) to create gene assemblies that correspond to the target sequences. YASRA cannot utilise paired end

information, but if good overlap is found in the sequence reads, it can extend the assembly beyond the target sequence boundary. We then used PRICE (Ruby et al., 2013) to utilise paired end sequence information to extend the sequence assemblies further, using the output from YASRA as target regions. We then implemented SSPACE (Boetzer et al., 2011) to further extend the gene assemblies. Final assembled contigs were aligned back to the reference sequences using BLASTN (Altschul et al., 1990) to target assembled contig assignments.

Target gene duplication can have several causes and can be found in multiple forms. Among others, multiple assembled contigs can align back to a single target sequence. This can be similar sequence in the genome such as a paralog or can be multiple alleles that had enough sequence difference to create multiple contigs. In addition, alleles can be collapsed in the assembly and we attempted to break these out into the correct number based on the ploidy using the phasing of the shared SNPs (see below). In the duplication processes, we used the target-based trimmed assembled contigs so we have complete sequence alignment and target to assembled contig contiguity between the samples. For each of the 384 samples, we identified the targets that align to multiple assembled contigs. Each case was combined with the multiple genes and the genes assigned to that target sequence from all the other 383 samples. The target sequence was aligned to outgroup genomes tomato and *Mimulus* (Nordberg et al., 2014), matching sequences, when found, where trimmed and combined into this sequence set. The sequences were then multiply aligned with MUSCLE Edgar, 2004a, 2004b) and a distance-based neighbor-joining tree was made using PHYLIP (Felsenstein, 2005). Finally, we "walked" through the tree structure and identified the multiple target matching contigs and the tomato outgroup. If one assembled contig is found within tomato (below in the tree structure) and one is outside, then the inside assembled contig is retained and the other is considered a

paralogous gene. If both are within or outside the tomato outgroup then both are retained. At a later stage in our project, a sweet potato genome assembled into chromosomes and an *Ipomoea trifida* genome became available (Yang et al., 2017 and Hirakawa et al., 2015). I then mapped our nuclear regions to these and confirmed than an overall majority of them map to only one position in those two genomes.

Descriptors of all nuclear regions are provided in Supplementary File 1 and unedited sequence files in Supplementary File 3.

2.3.3.4 | Haplotype identification using nuclear data

We collected information on ploidy levels of the species or specimens in the sweet potato group from the literature and from CIP. Retained assembled contigs from the previous step were put through a haplotype identification process to split the assembled contig into the correct haplotype number based on the provided ploidy. We aligned the nuclear raw reads back to the assembled contigs using Bowtie (Langmead et al., 2009). From this alignment, we created a variant call file that described the SNPs found within the alignment. We then ran Hapcompass (Aguiar and Istrail, 2012) to divide the assembled contig into haplotypes based on SNP phasing. We finally separated assembled contigs that show haplotype-defining SNPs into distinct contigs for downstream analysis.

We ran a coalescent analysis using Astral-II (Mirarab and Warnow, 2015) considering independent alleles for all genes and samples and found no significant intra-specimen variation (see Figure 4.16 in page 117). We conducted all subsequent phylogenetic analyses using consensus sequences.

2.3.3.5 | Assembly of chloroplast genomes and rDNA ITS

We assembled the chloroplast genomes and the nuclear *rDNA ITS* region[8] using SPAdes algorithm (Bankevich et al., 2012). We assembled the chloroplast genomes using as reference the chloroplast genome of *Ipomoea batatas* cultivar Xushu18 (Yan et al., 2015); for the *ITS* region we used the full *ITS* fragment (including the 5.8S coding region, see Figure 2.3 in page 19) of an *I. batatas* sequence (*C. Whitefoord 71*) obtained using Sanger sequencing.

The chloroplast genomes show the same general structure as in most other angiosperms, with one long single copy, one short single copy and two inverted repeats (Figure 2.6). Genome size in our *Ipomoea* samples ranges from 160,382 to 174,715 base pairs, except for *Ipomoea lactifera* J.R.I.Wood & Scotland, in which there have been several large deletions (150,628 base pairs).

We annotated all chloroplast genomes with Geneious v.9.1.8 (Kearse et al., 2012) using the annotated genome of the sweet potato cultivar Xushu18 (Yan et al., 2015) as reference. We also identified the 28 most variable regions in the chloroplast genome according to the tortoise and hare papers (Shaw et al., 2005, 2007, 2014). Variable region 27 corresponds to the *trnL–rpl32* non-coding region (Figure 2.6). All chloroplast whole genome sequences are provided in Supplementary File 4 and the *trnL-rpl32* regions in Supplementary File 5.

[8] We only used this approach to assemble the nuclear *rDNA ITS* of CIP specimens, because herbarium specimens had been already sequenced for ITS using Sanger.

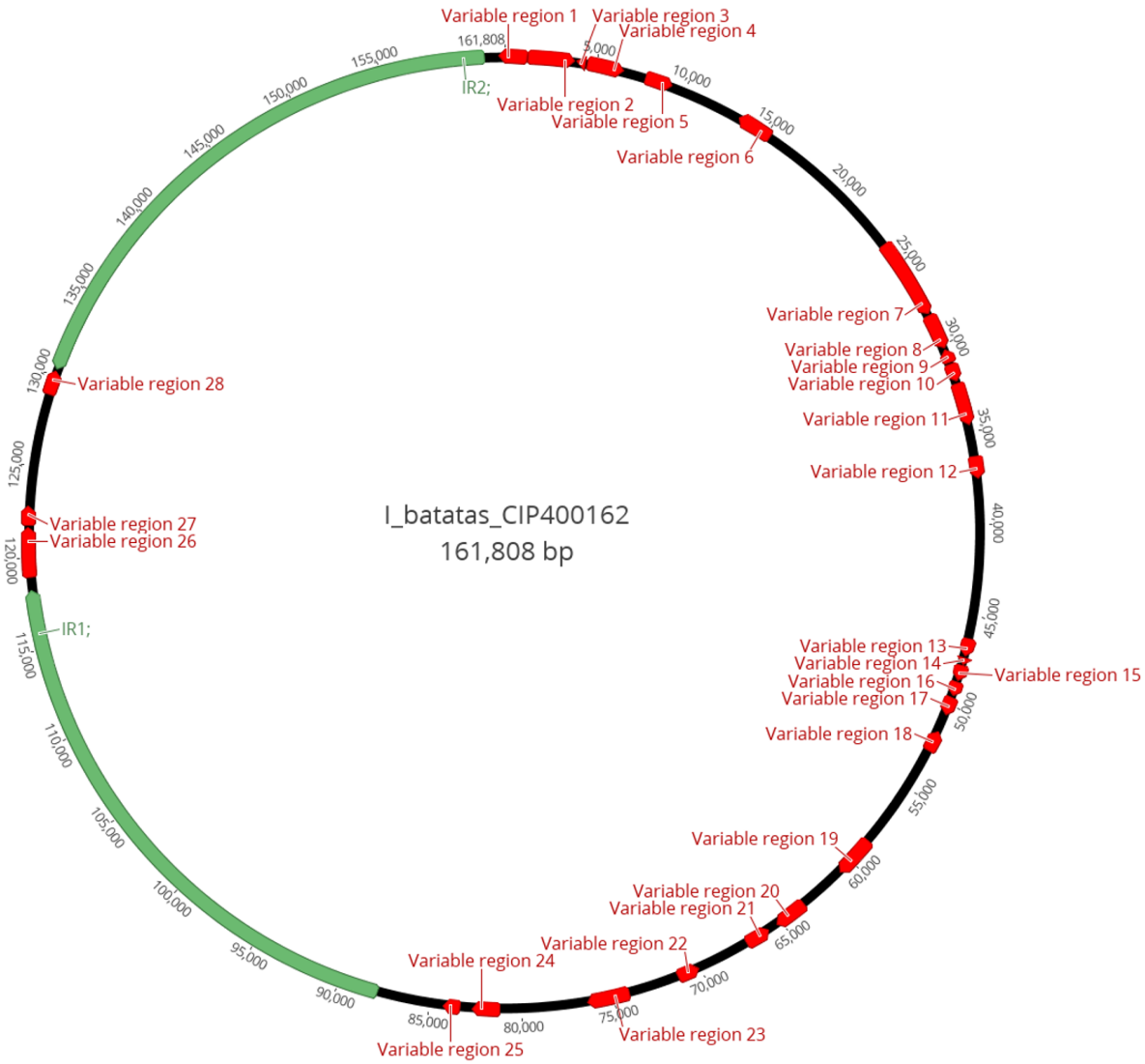


Figure 2.6. The organisation of the chloroplast genome in *Ipomoea* follows the same general structure as in other Angiosperms. Green bars indicate the position of the two Inverted Repeats flanking the Short Single Copy and the Long Single Copy regions. Red bars indicate the 28 most variable regions in the chloroplast genome according to Shaw and colleagues (2005, 2007, and 2014). Variable region 27 is the non-coding region *rpl32-trnL*.

2.3.3.6 | Sequencing and assembly of Banks and Solander's specimen

I extracted DNA from Banks and Solander's herbarium specimen (held at BM, Figure 2.7) and subsequently shipped it to Oregon. Lucas Longway, a member of Richard Cronn's lab at OSU, made the library in a laboratory with no history of involvement with our project and the sequencing process contained no other *Ipomoea* samples. Sequencing was conducted at OSU CGRB using MiSeq and 25 bp paired reads. We evaluated the degree of DNA damage in this specimen using mapDamage 2.0 (Jónsson et al., 2013)

(Supplementary File 6) and calculating DNA base percentages (Supplementary File 1) and found no signs of damage different from levels found in other herbarium specimens.

Figure 2.7. *Ipomoea batatas* specimen collected in 1769 by Banks and Solander in the Society Islands (held at the Natural History Museum in London).

Our aim was to assemble the chloroplast genome of this specimen, which we did using the protocol described in section 2.3.3.5. However, because genome skimming randomly retrieved fragments of the nuclear regions previously analysed in our study (1,016 reads mapped), we decided to try to assemble those fragments and incorporate them to

our nuclear phylogeny, aiming to obtain some insights from the nuclear genome. We assembled into contigs only those read pairs where both reads matched the reference sequence at approximately the expected distance and retained only those positions covered by at least three reads. We then aligned these fragments to all other specimens in this study and discarded all sites with ambiguous nucleotides, as well as all sites where only the Banks and Solander specimen incorporated indels. We finally retained 12,905 sites, 5,735 of which variable positions. I must insist, in any case, that our aim was to assemble the whole chloroplast genome, for which we obtained much higher read coverage, and that we conducted all estimates of divergence times using the chloroplast data only.

2.4 | PHYLOGENETIC ANALYSES USING DNA BARCODES

DNA barcodes are short DNA regions (~400–800 base pairs), easily amplifiable and shared across large groups of organisms, widely used to infer phylogenetic relationship between organisms (Baldwin, 1992; Baldwin et al., 1995; Feliner and Rosselló, 2007; Hebert et al., 2003; Kress, 2017). Phylogenies based on DNA barcodes, mainly *ITS*, have helped to identify several evolutionary patterns in *Ipomoea* and to reveal the phylogenetic position of *Ipomoea lactifera* as a sweet potato crop wild relative (Wood et al., 2015b) and the phylogenetic relationships of other cryptic species (Wood and Scotland, 2017b; Wood et al., 2015b, 2017b, 2017c). Analyses based on *ITS* sequences, however, have to be interpreted cautiously given the general characteristics of this marker.

When sequences are from very closely related taxa (i.e. from species in the same clade, for example in Batatas), and given the short length of the fragment, the topology obtained in subsequent phylogenetic analysis depends largely on the correct alignment of a few highly variable sites (Figure 2.8). If homologies in those sites are not properly identified, the position of a given sequence in the phylogenetic tree can be very different; in some cases, even the change in the alignment of a single position will cause a change in

the topology, hence hindering the inference of the true evolutionary relationships. These highly variable sites are not needed to infer relationships between species that are more distantly related (e.g. New World clade vs Old World clade; see Chapter 3), because the other less variable positions in the alignment (green in Figure 2.8) accumulate enough phylogenetically informative mutations. In that case, those very variable regions can be safely removed and the alignment will still be informative. However, the only way to obtain some resolution when analysing closely related taxa is to take these highly variable regions into account. This becomes a challenging task when we try to build a phylogeny of hundreds of sequences from across *Ipomoea*.

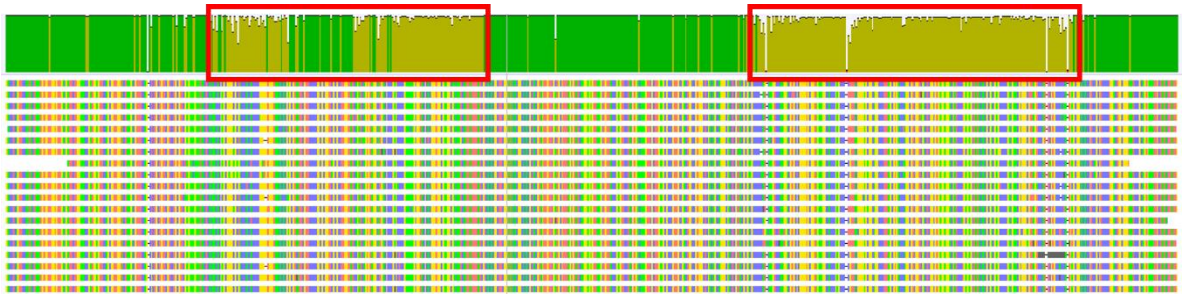


Figure 2.8. Alignment of very closely related *ITS* sequences. Red boxes indicate the highly variable regions in the alignment that accumulate most phylogenetically informative mutations. The correct identification of homologous sites in these regions is essential for the accurate inference of phylogenetic relationships between closely related samples.

Our most extensive *ITS* phylogeny includes 1,402 sequences from across *Ipomoea*. The automatic alignment of such a large number of sequences, despite belonging to the same genus, results in a problematic alignment of the highly variable regions aforementioned —which are needed to infer the relationship between closely related species. There are two ways to address this problem and both are used in this thesis (diagram in Figure 2.9). The first, traditional way is to align the sequences using a computer algorithm and then edit it manually to adjust the alignment. This approach, although very useful in our studies on *Ipomoea*[9], poses two issues that should be considered: first, as in a computer-

[9] All papers published as part of this project until now are based on a manually edited global phylogeny.

aided alignment, manual editing is also subject to generating errors and false homologies. Second, different people will deduce different homologies in the highly-variable regions and subsequently produce different alignments: importantly, a manually-edited alignment of this size cannot be replicated.

An alternative approach, which proved successful in our studies, was to obtain a rough approximation of the position of a given sample in a *backbone* topology and then re-align that sample to the group of most closely related sequences, and infer its specific position within that group (Figure 2.9). The *backbone* topology was inferred using the sequences in the high-throughput sequencing dataset —which retrieves the same main clades as the one obtained using high-throughput data—, whereas the second step included all other barcode sequences available for the group of interest. This two-steps process not only allows a more accurate alignment of the sequences —and hence the more accurate inference of true homologies—, but also facilitates the reproducibility of the analyses. We used this strategy during the late stages of our project, aiming to provide an online tool that allows the automatic identification of DNA sequences inputted by the user (see section 2.8).

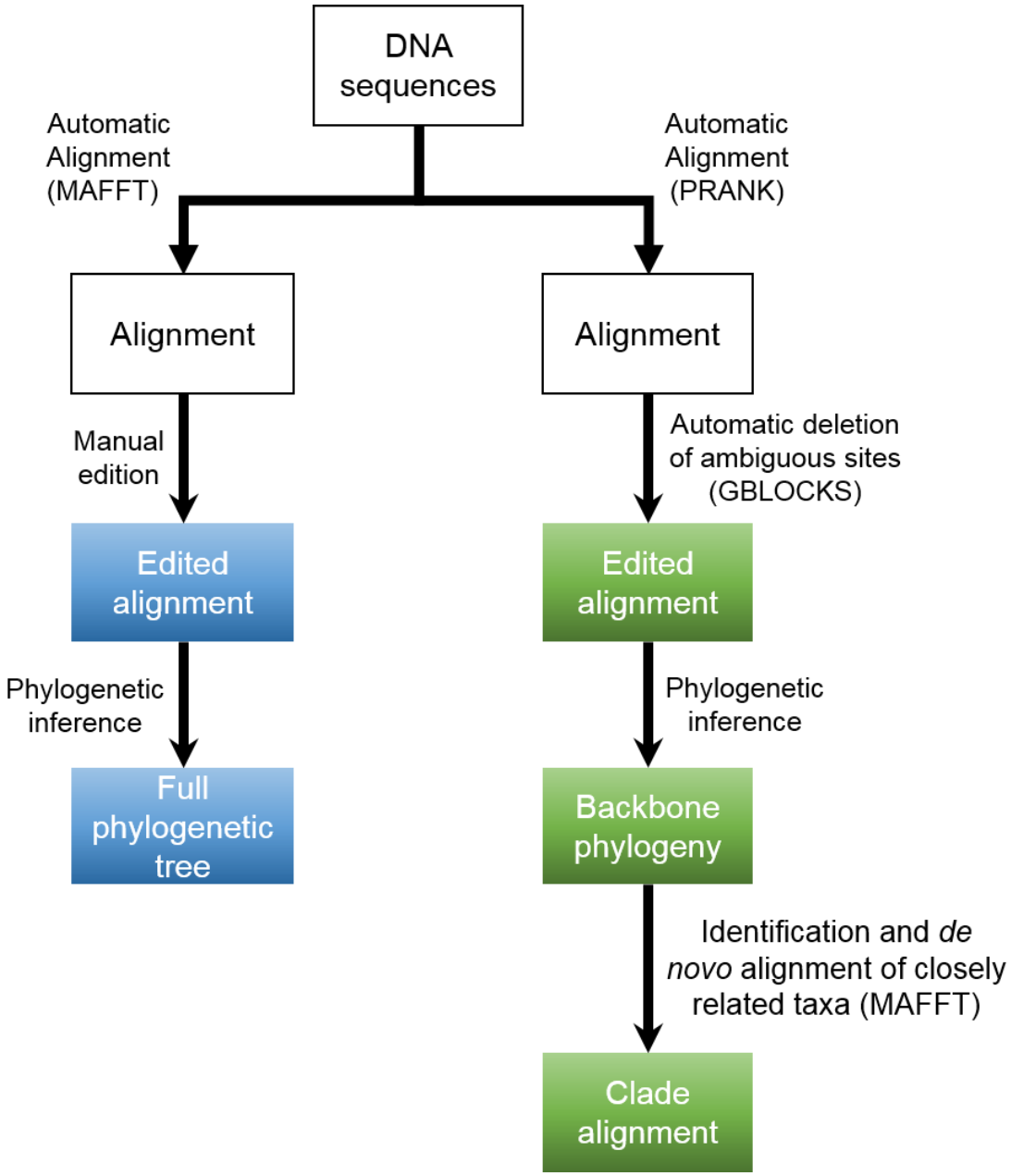


Figure 2.9. Two alternative strategies to deal with DNA barcode sequences from across the diversity of *Ipomoea*. Workflow on the left utilizes all barcode sequences at once. Workflow on the right is a two-step process in which a more sparsely-sampled phylogeny is used to infer the approximate position of a given sample (*backbone* phylogeny); subsequently, that sample is re-aligned to the other samples in the same clade to try to obtain a more accurate position within it.

2.4.1 | Sequence alignment

I conducted all alignments of DNA barcodes using either MAFFT v.7.2.1 (L-INS-I strategy; Katoh, 2002; Katoh and Standley, 2013) or PRANK v.1.140603 (-F option; Loytynoja and Goldman, 2008), two algorithms that follow different procedures and result in different alignments. The most important difference is that MAFFT, a widely used

progressive algorithm, severely penalises single insertion events[10] and combines multiple insertions/deletions (indels) in a single position. On the other hand, PRANK is defined as a "phylogeny-aware" method that differentiates insertions from deletions, assigning a lower penalty for opening gaps. In other words, PRANK supposedly is able to identify single insertion events in one terminal taxon without penalizing the necessary deletion in all other terminals to accommodate it[11]. Nevertheless, despite the fact that I explored many issues related to the alignments used, the vast majority of our analyses produced broadly similar results from different alignments, data sets and included taxa.

2.4.2 | Phylogenetic inference

For phylogenetic inference I retained those positions in the alignments with less than 30% gaps. I ran phylogenetic analyses using Maximum Likelihood in RAxML v.8 (Stamatakis, 2014), Approximate Maximum Likelihood in FastTree 2 (Price et al., 2009, 2010) and Bayesian inference in MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; Ronquist et al., 2012). All three methods recover the same major clades within *Ipomoea*.

The relationship between taxa within clades is mostly congruent across phylogenies that are generated from different data sets. Those instances of topological incongruence between different data sets and with the analyses using genomic data are discussed in the relative chapter.

[10] Due to the fact that gaps have to be opened in all other sequences to accommodate those insertions.

[11] PRANK's development team provided a more detailed explanation on their official website.

2.5 | PHYLOGENETIC ANALYSIS USING CHLOROPLAST GENOMES

Regardless of the methods and subset of samples used, the analysis of chloroplast data follows a general pipeline:

SEQUENCE ALIGNMENT
↓
REMOVAL OF POORLY ALIGNED REGIONS
↓
MODEL SELECTION
↓
PHYLOGENETIC INFERENCE

However, different algorithms and computer programmes offer multiple alternatives on each of these steps and I spent a good amount of time to explore all of them in depth. For that reason, this section is divided in sub-sections explaining each step and ends with an overview of the strategy followed (section 2.5.3).

2.5.1 | Sequence alignment

I took two different approaches dependent on scale to align chloroplast sequences. For our studies on the sweet potato and close relatives, I aligned the whole genomes as a single block. In contrast, for our studies on the entire genus, the growing genetic distance between species made it more difficult to align the whole genomes as a block, due to the amount of variability that had accumulated the non-coding regions. In this second case, I extracted the 81 coding regions and the 28 non-coding regions from our previously annotated genomes, aligned them independently and then concatenated the alignments.

2.5.1.1 | Alignment of whole chloroplast genomes

Among all available alignment tools, only MAFFT (Katoh, 2002; Katoh and Standley, 2013) and PRANK (Loytynoja and Goldman, 2008) were able to align whole chloroplast genomes. Mauve (Darling, 2004; Darling et al., 2010), Muscle (Edgar, 2004a,

2004b) and Clustal (Sievers et al., 2011) crashed a few minutes after starting due to insufficient computational capacity. As expected, PRANK was much slower than MAFFT (FFT-NS-2 method)[12] and produced much longer alignments, but the results of the subsequent phylogenetic analyses were almost identical. Therefore, in subsequent analyses of whole chloroplast genomes I used only MAFFT, which is faster.

I aligned the sequences using MAFFT (FFT-NS-2) with default parameters (gap penalty = –1.53). I then edited the alignments using Gblocks (Castresana, 2000; Talavera and Castresana, 2007), a programme that eliminates poorly aligned positions and ambiguous regions that could lead to wrong conclusions. I tested two different edition strategies: 1) removing all positions with a gap in any sequence and 2) removing all positions with more than half the sequences with a gap. Although the topology remains unchanged regardless the number of positions removed, removing all positions with a gap in the alignment resulted in slightly lower support values in some nodes, as otherwise expected.

2.5.1.2 | Alignment of independent chloroplast regions

I used PRANK for the alignment of independent chloroplast regions used to build the phylogeny of the entire genus *Ipomoea*. I aligned each region independently and then concatenated all together, defining two partitions for coding and non-coding regions in subsequent phylogenetic analyses.

2.5.2 | Phylogenetic analyses

I used jModelTest 2 (Darriba et al., 2012) to evaluate what model of sequence evolution is appropriate for the whole chloroplast genome data. The GTR+I+G model received the lowest –lnL value and was also supported by AIC and BIC criteria.

[12] It took PRANK several days to complete the alignment of whole chloroplast genomes, whereas MAFFT took just a few hours

I subsequently tested different methods of phylogenetic inference: Neighbor-Joining and Parsimony analysis as implemented in PAUP 4.0 (Swofford, 2002), Maximum Likelihood using RAxML v.8.0. (Stamatakis, 2014), approximate Maximum Likelihood using FastTree (Price et al., 2009, 2010) and Bayesian analysis using MrBayes (Huelsenbeck and Ronquist, 2001). The results are highly consistent regardless of the method of phylogenetic inference and the particular subset of specimens included.

As explained in Chapter 4, the chloroplast topology in our analyses of the sweet potato group disagrees with the nuclear topology and identifies a non-monophyletic sweet potato divided into two lineages, one of them more closely related to *Ipomoea trifida* than to the other sweet potato lineage. I evaluated the robustness of the chloroplast topology that identifies a non-monophyletic sweet potato using the approximately unbiased test (Shimodaira, 2002) as implemented in IQ-Tree 1.5.0a (RELL method with 100,000 resamplings) (Nguyen et al., 2015). The test provides significant support to the original topology compared to a constrained topology in which the two sweet potato lineages form a clade (Table 2.1).

Table 2.1. Statistical comparison of two phylogenetic trees with two different topologies. 1) Tree with two independent sweet potato groups (sweet potato not monophyletic), and 2) tree with a monophyletic sweet potato (*a priori* constraint).

Tree	-lnL	Diff -lnL	s.d.	T	KH	SH	wtd-SH	AU
1	80945.57215	(best)						
2	81145.72418	200.15203	55.506	3.606	0.0003*	0.0002*	0.0002*	~0*

KH: Kishino-Hasegawa test using normal approximation, two-tailed test. SH: Shimodaira-Hasegawa test using RELL bootstrap (one-tailed test). AU: Shimodaira approximately unbiased test. *Values for KH, SH and AU are P values for null hypothesis of no difference between trees ($P < 0.05$).

2.5.3 | Overall workflow

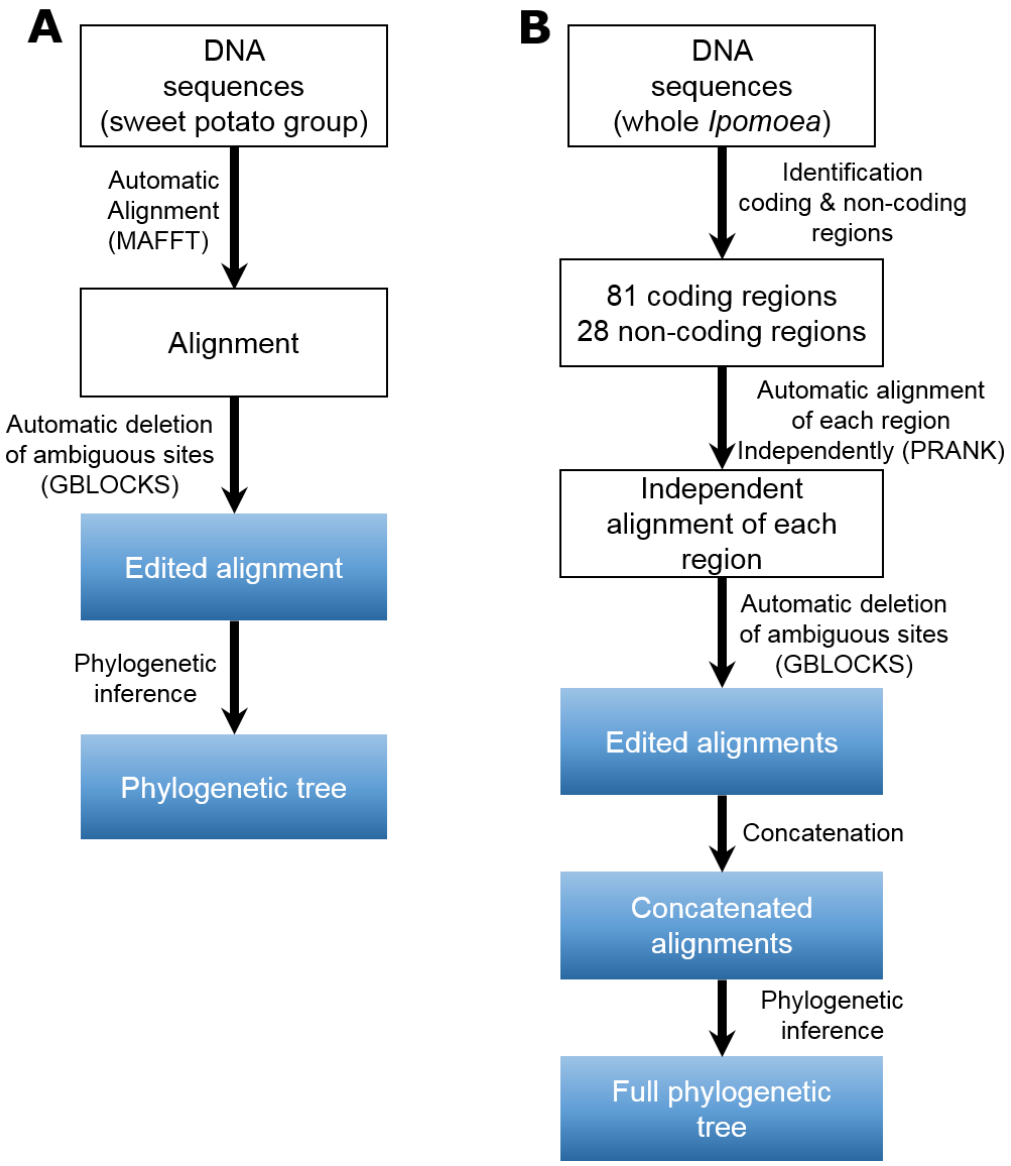


Figure 2.10. Overview of the strategy followed to analyse chloroplast sequences of A) members of the sweet potato group, and B) all *Ipomoea* specimens.

2.6 | PHYLOGENETIC ANALYSES USING NUCLEAR CODING REGIONS

2.6.1 | Sequence alignment

I aligned every nuclear region individually using L-INS-I strategy in MAFFT (gap penalty = 1.53) (Katoh, 2002; Katoh and Standley, 2013) and used default parameters in Gblocks (Castresana, 2000; Talavera and Castresana, 2007) to remove poorly aligned

positions from the alignment. Retaining either all gaps, half gaps or no gaps in Gblocks did not change the result of the analysis.

2.6.2 | Phylogenetic analyses

In a data set this large, neither intralocus recombination, Incomplete Lineage Sorting (ILS) nor reticulation can be discounted (Folk et al., 2017). I ran multiple analyses using different methods to evaluate the effect of these processes in our analyses. First, to reduce the possible effect of recombination, I ran the PHI statistical test (Bruen, 2005) to identify those nuclear regions in our data set likely to contain recombination. I ran all subsequent analyses using two data sets in parallel: one including all 605 regions and another including only the 434 regions that did not show evidence of recombination according to the PHI test.

To explore the effect of ILS, I ran three different types of phylogenetic analyses using coalescent-based and concatenated methods. I estimated evolutionary models for each region using jModelTest 2 (Darriba et al., 2012) and obtained independent gene trees using default parameters in FastTree 2.1.9 (Price et al., 2009, 2010). I then used gene trees as input to infer the species tree using Astral-II (Mirarab and Warnow, 2015).

In addition, I concatenated the alignments and conducted Approximate Maximum Likelihood as implemented in FastTree (Price et al., 2009, 2010) through the CIPRES Science Gateway (Miller et al., 2010).

Finally, I also used SVDQuartets (800,000,000 random quartets) (Chifman and Kubatko, 2014, 2015), a coalescent-based method available in PAUP 4.0 (Swofford, 2002). I used the supercomputer at University of Oxford Advanced Research Computing to run SVDQuartets.

Approximate ML of concatenated alignments and species tree inferred from independent gene trees retrieved the same general topology of *Ipomoeeae*, with one difference

in the relationship between clades A–D: the concatenated tree retrieves [[A,B],C,D], whereas the coalescent analysis retrieves [[A,B],[C,D]]. This and other examples of topological incongruence are discussed in detail in Chapter 3.

2.7 | ANALYSES OF POPULATION STRUCTURE AND GENETIC DIVERSITY

I conducted several population structure analyses, mainly of the species in the Batatas group. I randomly extracted 3,000 variable positions from the alignments of nuclear regions and used them as input for STRUCTURE (Falush et al., 2003; Pritchard et al., 2000) with 150,000 MCMC replications and 100,000 burn-in repetitions, using an admixture model and assuming independent allele frequencies among populations (λ = 0.4469; K = 1–5; 3 runs for each K value). I also ran independent analyses with the same parameters using 16 variable positions from the alignment of ITS sequences (λ = 0.4605; K = 1–4; 3 runs), 522 variable positions from the chloroplast alignment (λ = 0.0.3081; K = 1–5; 3 runs) and 5,735 variable positions from the nuclear alignments including Banks and Solander's specimen (λ = 0.3483; K = 1–5; 3 runs). I used Clumpak (Kopelman et al., 2015) to summarise the output from STRUCTURE.

2.7.1 | Genetic diversity

After identifying two independent sweet potato chloroplast lineages (see Chapter 4), I further explored the chloroplast and nuclear data to test whether any of the two lineages had more genetic diversity than the other. I divided the samples in two groups according to whether they belonged to lineage 1 or 2 in the chloroplast topology. I then estimated pairwise distances (p-value) between pairs of accessions in each lineage using Mega v.6.0 (Tamura et al., 2013) and used SPSS v.24 (IBM Corp, 2016) to obtain descriptive statistics for each lineage and to estimate whether genetic diversity was significantly different between sweet potato lineages.

I first conducted Kolmogorov-Smirnov and Shapiro-Wilk tests of Normality in both data sets and observed that the nuclear data follows a Normal distribution but the chloroplast data does not. I ran two subsequent tests with the same null hypothesis ("the distribution of genetic distances is the same across lineages", and defined a confidence interval of 95%): on the nuclear data I ran a t-test considering independent samples (parametric test), whereas on the chloroplast data I ran an independent-samples Mann-Whitney U test (non-parametric test). The output of the analysis of genetic diversity is in Appendix 1.

2.8 | TIME-CALIBRATED PHYLOGENIES[13]

We generated time-calibrated phylogenies to estimate the time of divergence of the sweet potato. Although I participated in the design of the analyses, these were conducted by Tom Carruthers, another D.Phil. student with Robert Scotland. These analyses are described here because they provided the necessary time frame for the discussion below.

We implemented divergence time estimation in RevBayes (Hohna et al., 2014, 2016), a graphical modelling framework enabling highly flexible model specification. Because of a lack of previous divergence time estimates in Convolvulaceae, we constructed a supermatrix of three chloroplast genes (*matK*, *rbcL*, *atpB*), the chloroplast *trnL-trnF* intergenic spacer, and the nuclear ribosomal *ITS* region which incorporates a balanced sample of taxa from across both Convolvulaceae and its sister family Solanaceae (obtained from GenBank. Accession numbers in Supplementary File 1). This matrix covers a sufficiently broad phylogenetic scale to enable the implementation of temporal calibrations. In our analyses, we used a single normally distributed calibration (mean =

[13] The methodology in this section has been incorporated from what we presented in our April 2018 paper in *Current Biology* (Muñoz-Rodríguez et al., 2018). The content of that paper is included here in agreement with Elsevier's policy that allows authors of papers to include the content of a paper in a thesis, provided it is not published commercially: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights. See also footnote 7.

67.34 million years, standard deviation = 9.980 million years) for the divergence between Convolvulaceae and Solanaceae. This calibration age is derived from a previous study which simultaneously implements 132 fossil calibrations across angiosperms (Magallon et al., 2015). This calibration is likely to represent an underestimation of the true age of the divergence between the two families because many of the 132 fossils that were used are likely to be significantly younger than the true age of the node which they were used to calibrate. In turn, this is likely to result in the age estimates inferred in this study to be biased toward younger ages. Despite this apparent limitation, we believe this approach is appropriate for the purposes of our study—namely to infer whether the origin of the sweet potato occurred in human or pre-human times.

The utility of our pragmatic calibration approach is further highlighted by recent work which demonstrates apparent conflict within the Solanaceae fossil record (the closest relatives to *Ipomoea* with a fossil record) (Särkinen et al., 2013; Wilf et al., 2017). Although our approach was useful for the purposes of this study, extreme caution should be taken if using dates inferred in this study as secondary calibrations in future studies which aim to answer different questions.

We used this matrix and age calibration to infer a time-calibrated phylogeny for Convolvulaceae and Solanaceae. A GTR+I+G model of DNA substitution was implemented, and branch-specific substitution rates were inferred using an uncorrelated lognormal relaxed clock with a standard deviation 0.2972 (corresponding to 0.5 orders of magnitude). We partitioned the supermatrix such that separate parameters for nucleotide substitution and branch-specific substitution rates were inferred for the chloroplast and *ITS* data. A constant rate birth-death branching process was implemented as the time prior in this analysis.

We then used a matrix containing samples from throughout *Ipomoea* based on 21 nuclear genes for which there was high coverage (99%) to infer divergence times within the genus, such as the crown nodes for *Ipomoea* series *Batatas* and the Tuboides clade (see discussion in Chapter 5). We implemented a GTR+G+I model and inferred branch-specific substitution rates using an uncorrelated lognormal relaxed clock with a standard deviation 0.2972. We estimated a single set of parameters for nucleotide substitution and branch-specific substitution rates for the entire 21 gene matrix. We implemented a constant rate birth-death branching process as the time prior. The age for the root node of this tree is determined by the sampled ages for the equivalent node in the Convolvulaceae and Solanaceae time-calibrated phylogeny.

Based on the inferred ages for the crown node of *Ipomoea* series *Batatas* and the Tuboides group, we inferred three more time-calibrated phylogenies: two for series *Batatas*—one based on plastome data and one based on a matrix of the 21 nuclear genes for which there was 100% coverage, and one for the Tuboides group—based on the same 21 nuclear genes. In each of the three separate trees, we implemented a GTR+G+I model and inferred branch-specific rates of DNA substitution with an uncorrelated lognormal relaxed clock with a standard deviation of 0.2972. Neither the chloroplast plastome data set nor the nuclear data sets were partitioned. Therefore, we estimated a single set of parameters for nucleotide substitution and branch-specific substitution rates for each of the three time-calibrated phylogenies.

2.8.1 | Divergence time for Banks and Solander's specimen

We performed two subsequent analyses using all *Ipomoea trifida* and *I. batatas* specimens to estimate when the specimen collected by Banks and Solander diverged from

its closest relative. These analyses were performed exclusively using chloroplast data because the nuclear data we recovered from Banks and Solander's specimen was fragmentary.

In one analysis, we constructed a time-calibrated phylogeny in a manner similar to that described above. We implemented a GTR+G+I model and inferred branch specific substitution rates for each chloroplast lineage using an uncorrelated lognormal relaxed clock with a standard deviation 0.2972. We implemented a birth-death branching process as a prior for the divergence times and calibrated the root node with a normal distribution, with a mean of 0.7 million years and a standard deviation of 0.18 million years (corresponding to the age inferred for this node in our chloroplast time-calibrated phylogeny for *Ipomoea* series *Batatas*). We implemented this younger age calibration (compared to the equivalent age inferred for this node from nuclear data) to provide the most robust challenge to the hypothesis that Banks and Solander's specimen dispersed to Polynesia in pre-human times. Based on this time-calibrated phylogeny we were able to infer a divergence time for the split between Banks and Solander's specimen and its closest relative.

We also performed a coalescent analysis for the same data set. In this case, we assumed the same species tree as that in our analysis to infer ancestral population sizes: specifically, a three-taxon tree of *Ipomoea batatas* CL1, *I. batatas* CL2 and *I. trifida* in which *I. batatas* CL2 is sister to *I. trifida*. We also assumed a fixed gene tree (representing the relationships between the plastomes of all sampled specimens of both species) which was based on that inferred in our previous analyses of the chloroplast data but also including the specimen collected by Banks and Solander. We implemented a GTR+G+I model of sequence evolution and assumed overall rates of sequence evolution to be constant among different branches of the gene tree. Effective population sizes on the species

tree were assigned an exponential prior distribution with a rate parameter of 0.1, and the species tree was assumed to evolve under a constant rate of speciation and extinction. We calibrated the root node using the same parameters as in the previous analysis. This second analysis infers coalescent times between different chloroplast lineages in the gene tree, and therefore allows us to infer when the Banks and Solander specimen is likely to have diverged from its closest relative. This analysis makes several different assumptions compared to that of our more conventional time-calibrated phylogeny – most notably the rate of coalescence between different chloroplast samples in the gene tree is influenced by the relative effective population size of the relevant branch in the species tree. It, therefore, enables us to test the sensitivity of our conclusions to the assumptions inherent in different analytical methods. We ran each analysis for 500,000 generations, sampling every 50 generations.

2.9 | IPOMOEA PROJECT WEBSITE

An aim of our studies on the sweet potato and its wild relatives is to provide genomic tools that help identify entities and traits of interest for crop improvement. For that reason, we aim to make the data in our studies and different tools freely available to sweet potato breeders and researchers. During the last months of my D.Phil., I worked to create a website to host all these resources. This website, the *Ipomoea Project Website*, is available through the website of the Department of Plant Sciences, University of Oxford.

The website features the result of four years of taxonomic and phylogenetic research, not only on the sweet potato but on the entire genus *Ipomoea*. We include a BRAHMS database with descriptions of all specimens studied together with photographs and maps, as well as phylogenetic trees and all the molecular data generated during these four years.

Also, I have developed an online tool that allows the quick and reliable identification of all species recognised in *Ipomoea*, including the sweet potato and its relatives, by using simply a small DNA barcode sequence. This tool allows users to upload the sequences they obtained as part of their studies to identify their position in our phylogenetic trees. The code underlying this pipeline is provided, together with example files, in Supplementary File 7.

3 | PHYLOGENETIC STUDIES OF *IPOMOEA*

3.1 | INTRODUCTION

Recent years have seen an increase in the number of phylogenetic studies on the genus *Ipomoea*. This is mostly explained by the relevance of the sweet potato (*Ipomoea batatas*) and the growing interest of several other species as plant models in developmental biology and genomics (Baucom et al., 2011; Christoff et al., 2012; Duncan and Rausher, 2013a). However, limited taxon and character sampling hindered most of those previous studies, even those focused on the origin of the sweet potato as explained in the next chapter.

A common feature of previous phylogenetic studies on the entire genus *Ipomoea* is that, despite the large number of species and the global distribution of the genus, most papers — with the exception of Wilkin's (1999) work— have most often included a similar set of taxa: some widespread species such as *Ipomoea purpurea*, *I. nil* or *I. pes-caprae* (L.) R.Br.; the sweet potato and one or two of its closest wild relatives (often *I. lacunosa*, *I. triloba* L. or *I. umbraticola* House [=*I. splendor-sylvae* House]); and some other well-known species, for example *I. alba* L., *I. aquatica* or *I. cairica*. In contrast, most species have never been sampled for any phylogenetic study.

A detailed account of the evolution of the taxonomy and nomenclature of *Ipomoea* and *Ipomoeeae* will be presented in the monograph that is currently under development (Wood, Muñoz-Rodríguez *et al*., in prep.). In this chapter, I review the previous phylogenetic studies on the genus *Ipomoea* and the tribe *Ipomoeeae* and summarise the main contributions to the taxonomic knowledge of the genus. I then present the results of our studies using genomic-scale data and discuss several evolutionary patterns that characterise these plants. The results presented here illustrate how any study that aims to understand the relationships between taxa, to identify evolutionary patterns or to inform taxonomic decisions and species classification, in a megadiverse group of plants such as *Ipomoea*, must include a comprehensive sampling of

as much diversity within the group as possible, not just the most representative or well-known species.

3.1.1 | Traditional circumscription of *Ipomoeeae*

There have been many attempts to classify the diversity within *Ipomoea*. Already in the 19th century, Choisy (1833, 1838, 1845), Grisebach (1864), Meisner (1869), Gray (1878) and Hallier (1893, 1894) arranged the members of *Ipomoea* in various sections or series, or proposed to split it in multiple genera. Those early treatments were followed by others in the 20th century (Austin, 1975a, 1975b, 1979, 1980; Austin and Huáman, 1996; House, 1908; van Oostroom, 1953; Verdcourt, 1957).

Hallier (1893, 1894) defined two groups in the Convolvulaceae family based on spiny or smooth pollen, Echinoconiae Hallier f. and Psiloconiae Hallier f. respectively. He further divided **Echinoconiae**, the group of taxa with spiny pollen, into two groups: *Ipomoeeae* Hallier f. and *Argyreieae* Hallier f. ***Ipomoeeae*** *sensu* Hallier included *Ipomoea* (arranged in six sections) and four other genera[14]. Some of these genera are no longer recognised, whereas four genera that Hallier placed in *Argyreieae* are now considered members of *Ipomoeeae*[15]. Argyreieae has been subsequently considered a subgroup within *Ipomoeeae* (see for example Eserman et al., 2014). Later authors described and added other smaller genera to the group, for example *Lepistemonopsis* Dammer (1895) or *Paralepistemon* Lejoly & Lisowski (1986), or transferred some genera previously recognised by Hallier as sections of *Ipomoea*, for example *Quamoclit* (O'Donell, 1959).

The infrageneric classification of *Ipomoea* proposed by Austin based on morphological characters (Austin, 1975a, 1975b, 1979, 1980; Austin and Huáman, 1996) is the most widely

[14] *Astroclaena* Hallier f. (*nom. illeg.*, *non* Corda) [= *Astripomoea* A.Meeuse], *Calonyction* Choisy, *Lepistemon* Blume and *Quamoclit* Mill

[15] *Argyreia* Lour., *Rivea* Choisy, *Blinkworthia* Choisy and *Stictocardia* Hallier f.

used. Austin divided *Ipomoea* in three subgenera (*Ipomoea* L., *Quamoclit* (Moench) Clarke and *Eriospermum* (Hallier f.) Verdc.) and these, in turn, in various sections and series. Austin's infrageneric delimitation of the genus is the scheme to which most authors compared the results of their phylogenetic studies using molecular data. It is presented, for comparison purposes, in Appendix 2.

As currently circumscribed, the tribe *Ipomoeeae* comprises *Ipomoea* and nine smaller genera: *Argyreia*, *Astripomoea*, *Blinkworthia*, *Lepistemon*, *Lepistemonopsis*, *Paralepistemon*, *Rivea*, *Stictocardia* and *Turbina* Raf. All ten genera share the characteristic spiny pollen, as opposed to a smooth pollen in the other tribes in Convolvulaceae (Hallier, 1893; Manos et al., 2001; Stefanovic et al., 2003).

3.1.2 — Phylogenetic studies on *Ipomoeeae* and *Ipomoea*

In 1992, McDonald and Mabry (1992) studied restriction site variation in chloroplast DNA of 31 species to test whether the classification proposed by Austin was correct. This was the first study to use DNA data to build a phylogeny of *Ipomoea*. Their results provide support for some groups recognised by Austin (*Batatas*, *Calonyction*, *Mina*, *Pharbitis*, Ser. *Jalapae* and Ser. *Microsepalae*) and do not support other groups (several species in Austin's Ser. *Setosae* belong elsewhere, for example). McDonald and Mabry identified some interesting preliminary results that would be later confirmed by other authors and supported with our own extensive sampling, for instance the placement of *Ipomoea setosa* Ker Gawl. as sister to the Batatas group in the chloroplast phylogeny.

The most complete phylogeny of the tribe *Ipomoeeae* to date included 138 species of *Ipomoea* and related genera in a morphological cladistic analysis (Wilkin, 1999). Still, it is worth noting that this study covered less than 20% of the estimated number of species in the

group and included only one specimen per species. Even though the relationships between species presented by Wilkin have not been supported in subsequent molecular analyses, the results of his study suggested that the smaller genera in *Ipomoeeae* are part of *Ipomoea*, rather than independent genera. Subsequent studies using random amplified polymorphic DNA markers (Dhillon and Ishiki, 1999), two nuclear DNA markers (Manos et al., 2001; Miller et al., 2002), four chloroplast DNA regions (Stefanovic et al., 2002, 2003) and the whole chloroplast genome (Eserman et al., 2014) provided further support to Wilkin's (1999) hypothesis that at least some of these genera are nested within *Ipomoea*. Furthermore, those papers show that several taxonomic groups recognised by Austin are not monophyletic. However, their limited taxon and data sampling[16] demands a cautious interpretation of the results. Specifically, a comprehensive sampling of the diversity within *Ipomoea* and all the smaller genera is necessary to produce a phylogenetic framework that allows an accurate interpretation of the relationships within the group.

Finally, the monophyly of most species of *Ipomoea* has never been realistically assessed. Even for those species that are always included in molecular studies, they normally consist of one specimen only (see for example Wilkin, 1999; Miller et al., 1999; or Miller et al., 2002 among others), with species monophyly assumed. Some studies, focussed on specific sections and series, did include multiple specimens per species (for example Eserman, 2012; Miller et al., 2004), but the limited amount of data (ITS sequences) provided only an incomplete picture of the relationships within those groups and does not make for robust conclusions.

Here, I present a comprehensive phylogenetic study of *Ipomoeeae* using genomic-scale data. Our taxon sampling includes 455 species (about 60% species known in *Ipomoea*), three

[16] 29 and 40 species of *Ipomoea* in (Dhillon and Ishiki, 1999) and (Miller et al., 1999) respectively; 45 from across *Ipomoeeae* in (Manos et al., 2001); 112 (only 10 of them *Ipomoea*) and 20 from across Convolvulaceae in (Stefanovic et al., 2002) and (Eserman et al., 2014) respectively.

times more than the next most complete molecular study on the genus. In terms of character sampling, the 605 putative single-copy coding nuclear DNA regions that we sampled total over two million base pairs. This, together with the whole chloroplast genomes of 384 samples and the three DNA barcodes for 1,176 additional specimens, constitute an extraordinary resource from which to study the genus *Ipomoea* and its allies.

3.2 | RESULTS AND DISCUSSION

I produced several phylogenies of *Ipomoea* using different data sets and methods of phylogenetic inference: phylogenies using genomic-scale data (single-copy nuclear regions and whole chloroplast genome) and phylogenies using DNA barcodes, with a larger taxon sampling but a smaller amount of data (Figure 3.1) (see Methods in Chapter 2). These phylogenies are complementary in many ways. First, as explained below, all trees are largely congruent and identify the same main clades except one. Second, and as a result of this general agreement, the barcode phylogenies provide a good estimate of the position of many taxa for which no genomic-scale data has been generated yet. The three main phylogenies are:

- **Nuclear phylogeny** (Figure 3.2 in page 52): species tree inferred using gene trees from 434 putative single-copy coding nuclear DNA regions that did not show evidence of recombination, as well as from concatenated alignments of those regions. Includes 384 samples. Clade-specific phylogenies extracted from the phylogeny inferred using concatenated alignments are presented to best illustrate the discussion.

- **Chloroplast phylogeny** (Figure 3.3 in page 53): inferred from whole chloroplast genomes. Includes the same samples as the nuclear phylogeny, except five that were removed for problems with the assembly. Clade-specific phylogenies extracted from the phylogeny inferred using Bayesian inference are presented to best illustrate the discussion.

- ***ITS* phylogeny** (Figure 3.4 in page 54): inferred from 1,242 *ITS* sequences using Approximate ML: 1,199 from *Ipomoeeae*, 42 from other genera in *Convolvulaceae* and *Solanum tuberosum* as outgroup. Clade-specific *ITS* phylogenies are provided in Appendix 3.

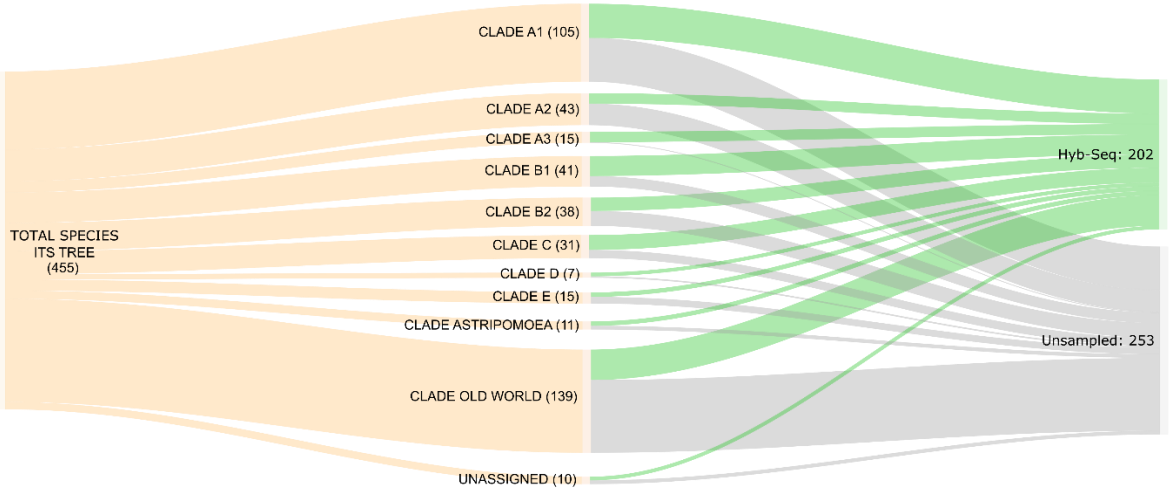


Figure 3.1. Sankeymatic diagram showing number of species represented in the *ITS* (orange) and Hyb-Seq (green) phylogenies for each clade. We estimate that the tribe *Ipomoeeae* consists of over 800 species and, therefore, our phylogenetic analyses using DNA barcodes include roughly 60% species known in the tribe, whereas the Hyb-Seq analyses include around a quarter of species.

The discussion that follows takes into account all phylogenies. I use the names above to refer to the three phylogenies throughout the discussion (i.e. nuclear, chloroplast and *ITS*). Rather than a detailed discussion of every single species in our phylogenies, I focus on a few specific patterns that can be inferred from them. I also discuss certain species and groups of species for various aspects such as their economic importance, their unexpected position in the phylogeny or their relevance to the next chapters of this thesis. Also, considering the limitations of an A4 document and to aid visualization, in the main text I include summary phylogenies identifying the main clades (Figures 3.2–3.4), clade-specific phylogenies and phylogenies highlighting specific questions to illustrate the results of our analyses. Complete, unedited phylogenies in NEWICK format are provided in Supplementary Files 8 (nuclear phylogenies) and 9 (chloroplast phylogenies).

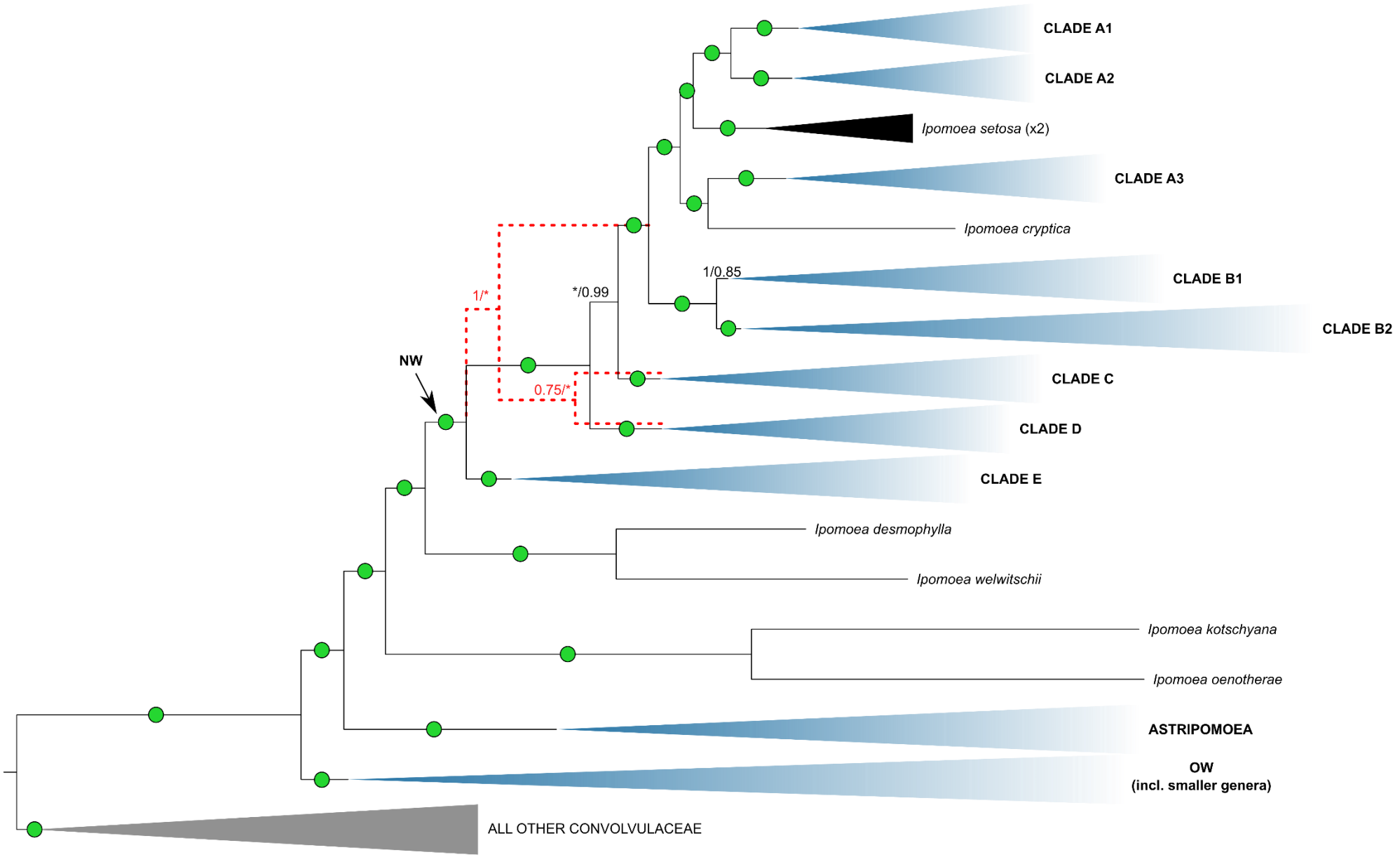


Figure 3.2. Phylogeny of *Ipomoea* inferred using Approximate Maximum Likelihood and Astral-II (coalescence) from 434 non-coding single copy regions concatenated and independent gene trees respectively. Red dashes indicate the alternative topology obtained with the coalescence method. Green dots indicate 100% support in both analyses; when support ≠ 100%, first number indicates Astral support values and second number ML support (Shimodaira test, 1,000 replicates).

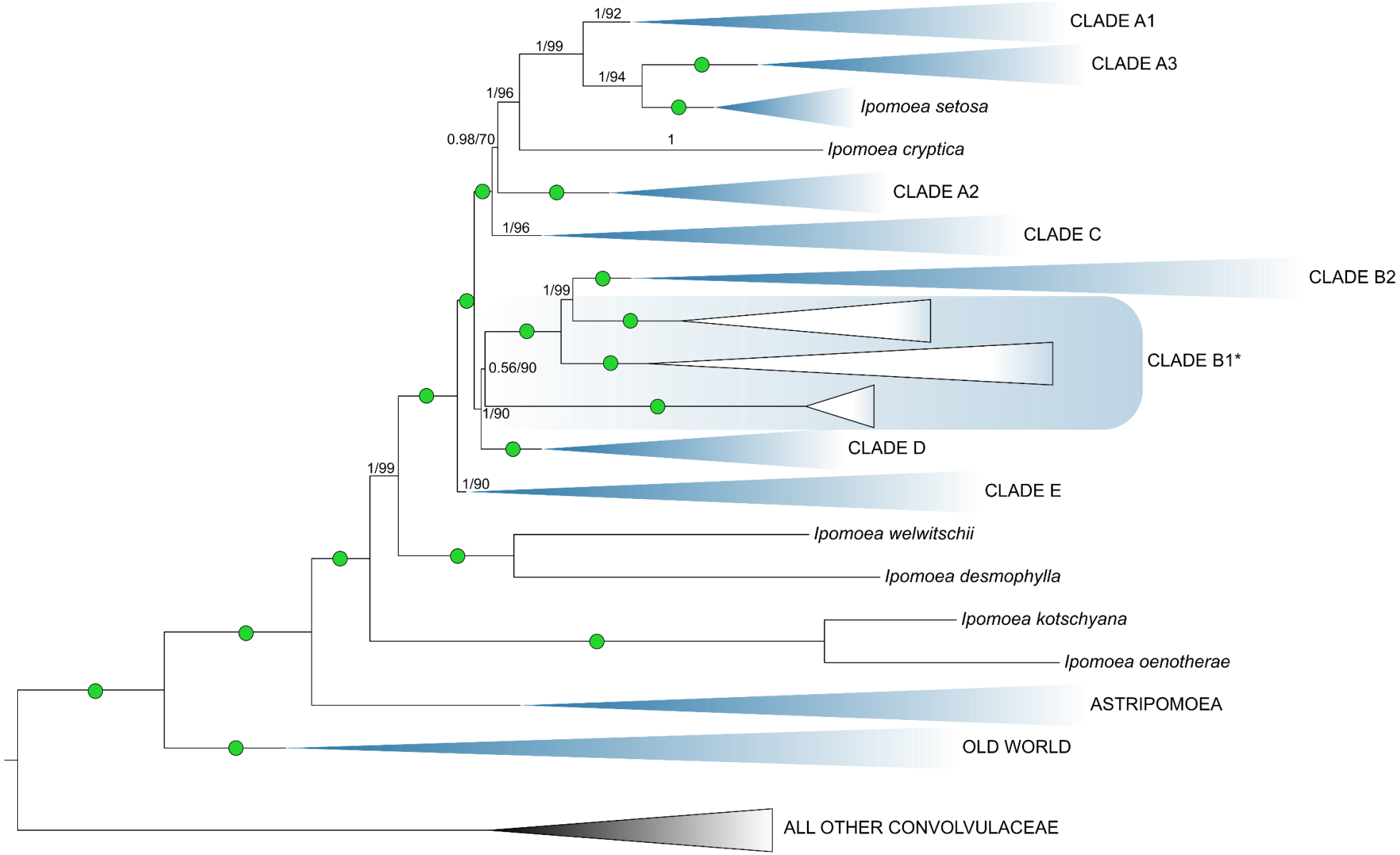


Figure 3.3. Chloroplast phylogeny inferred using Bayesian and Maximum Likelihood (ML). Analysis using whole chloroplast genome data, divided in coding and non-coding regions, each of which was aligned independently and then concatenated and analysed as two partitions. Green circles indicate 100% support in both analyses; when support ≠ 100%, first number indicates, first number indicates Bayesian posterior probability and second value is ML bootstrap support.

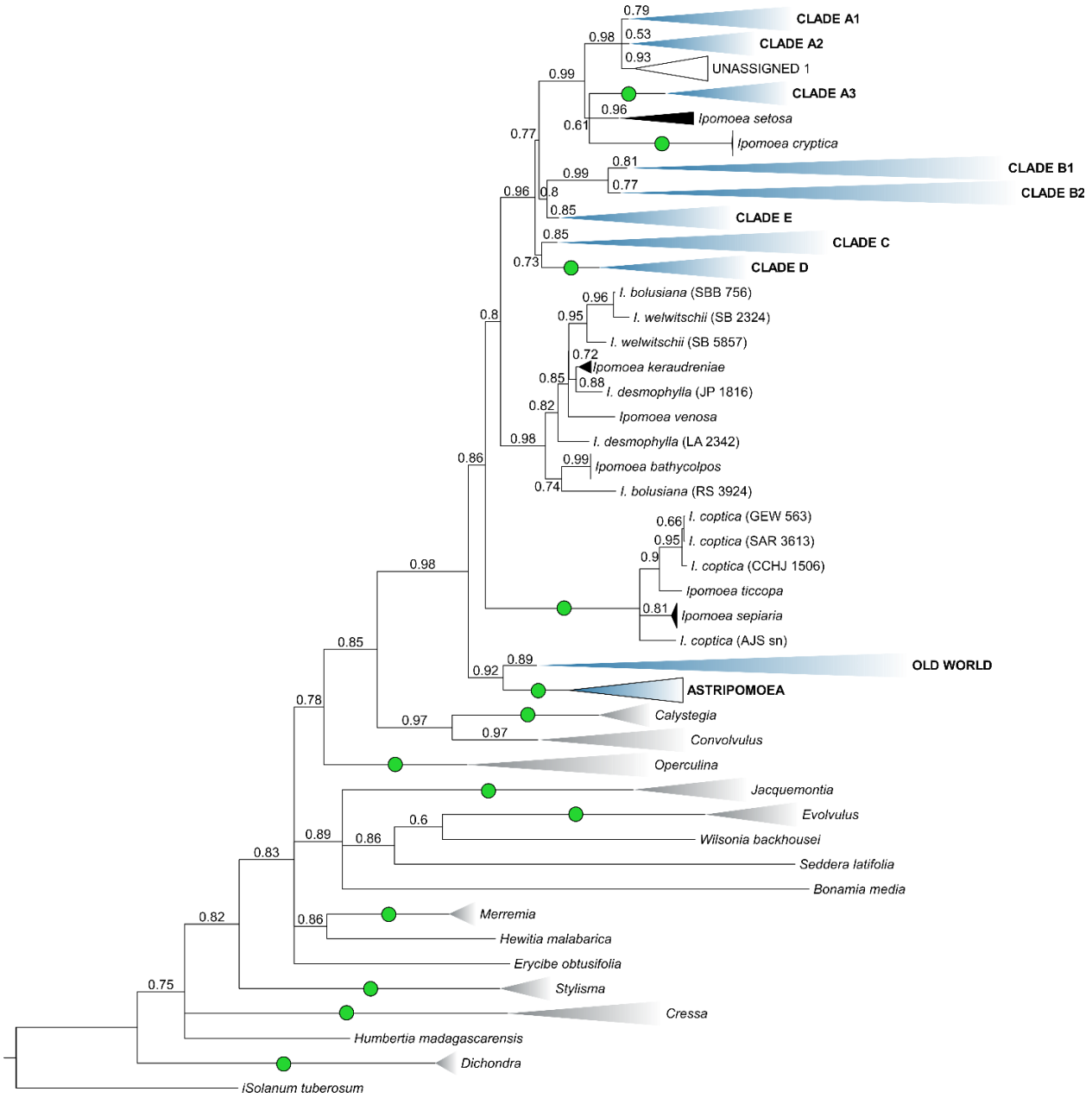


Figure 3.4. Approximate Maximum Likelihood *ITS* phylogeny of *Ipomoea*. Number above the branches represent support values (Shimodaira test, 1,000 replicates); Green dots indicate 100% support.

3.2.1 | Identification of the material

As explained in the introduction of this thesis, we study herbarium specimens and conduct molecular analyses in parallel, using DNA barcoding to corroborate, refute or augment taxonomic hypotheses —mainly species delimitation— based on morphology (Figure 3.5). As a result of our studies, around 30% specimens across *Ipomoea* have changed their identification during the last four years (Figure 3.6). The names given to the specimens included in this dissertation are the most up to date identifications after John Wood, a taxonomic expert on *Ipomoea* involved in the project, reviewed the specimens taking into account both morphological observations and molecular sequence data.

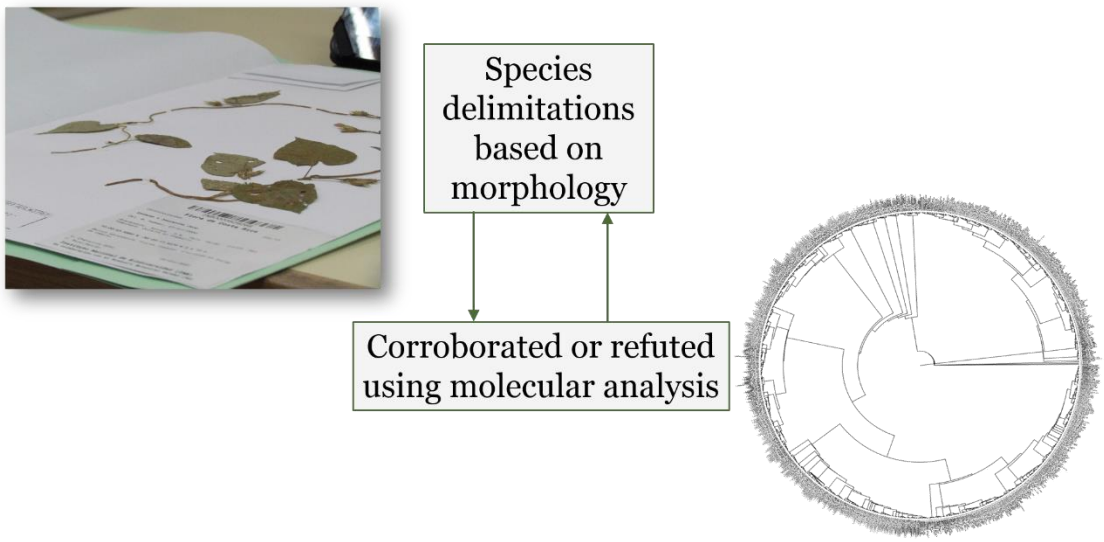


Figure 3.5. We conduct morphological and molecular studies in parallel, reciprocally illuminating each other.

It is important to note that we are conservative in the conclusions drawn from DNA barcodes. If the results of the phylogenetic analysis using these markers (mainly *ITS*) agree with the species boundaries defined using morphology —that is, if the phylogeny places all specimens of that species in a highly supported monophyletic group—, we accept that the delimitation of the species is consistent in both morphology and molecular data.

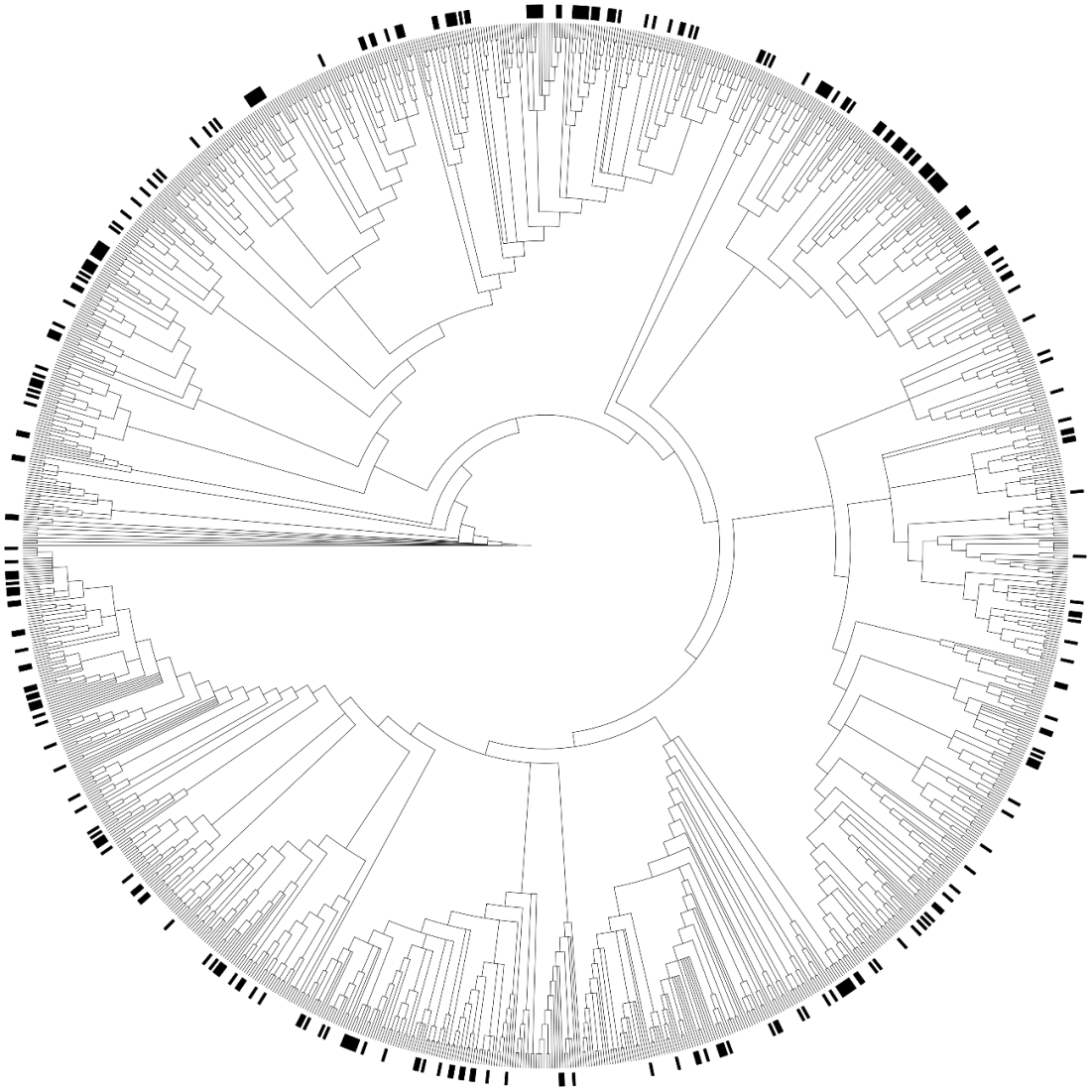


Figure 3.6. *ITS* phylogeny of *Ipomoea* indicating the specimens that changed their identification during this project. Each black bar in the outer circle indicates a specimen that received a new identification during our studies, different from the original identification on the herbarium label.

Also, if specimens from the same species are distantly related in the barcode phylogeny, this suggests that they may represent different evolutionary entities that have not been identified in morphological studies, and we therefore acknowledge that it is necessary to re-consider the delimitation of the species. However, if the barcode phylogeny does not place all specimens of a species in the same clade, but these are closely related, we do not make a decision about the species boundaries but understand that more data is required to clarify whether the entity in question represents a good species. In other words, we used the *ITS* phylogeny in a similar way to any other taxonomic character:

when the *ITS* phylogeny agrees with an existing hypothesis based on other data, then we interpret the hypothesis as more strongly supported. When the *ITS* phylogeny strongly disagrees (specimens from the same species belonging to distantly related clades) with existing hypotheses, then we re-examine all data to check whether the conflict is real or simply, as in many cases, a wrongly identified specimen. In a third situation that was more prevalent across the *ITS* phylogeny, multiple accessions of a species were closely related but not monophyletic. In these occasions, we simply ignored the non-monophyly of species based solely on *ITS* sequences, as we could not be sure whether it was an accurate pattern or it was simply due to the small size of the *ITS* fragment and its inability to corroborate monophyly.

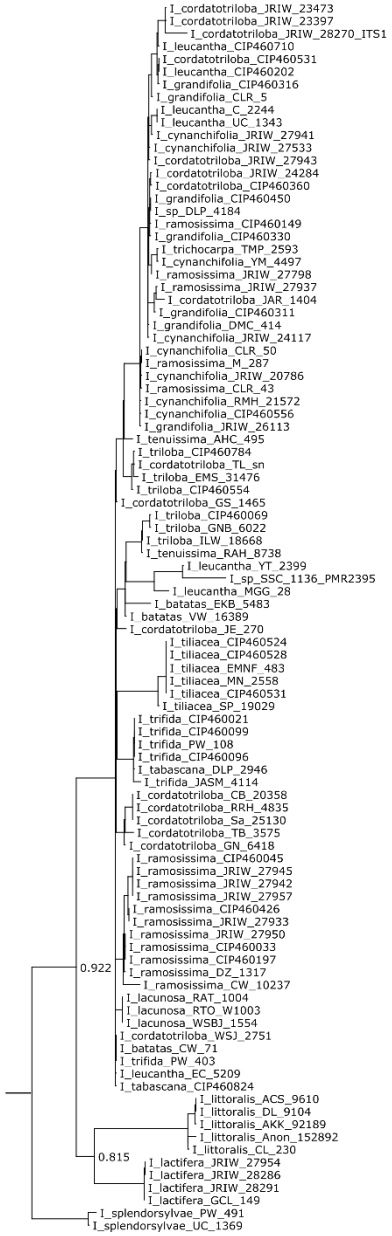


Figure 3.7. The phylogeny inferred using *ITS* sequences results in a largely unresolved polytomy, whereas the phylogenies using nuclear and chloroplast data identify most species monophyletic.

We adopted this approach based on our experience with the group of species closely related to the sweet potato (discussed in next chapter). In the *ITS* phylogeny (Figure 3.7), most species are not monophyletic but result in a polytomy. However, the nuclear and chloroplast phylogenies, with more data, supported the monophyly of most existing species. In this case, we would have made a mistake if we had decided to redefine the boundaries between species based on the results of the *ITS* phylogeny only.

For this reason, in this thesis I only refer to specific taxa in the *ITS* phylogeny if they are monophyletic or if they are clearly polyphyletic —i.e., the specimens split in two or more groups distantly related. I do not discuss the

non-monophyletic taxa in cases when the specimens, despite not forming a clade, are closely related; in such cases, it is assumed that *ITS* is of limited value to interpret species boundaries.

3.2.1.1 | Species concept

Whilst a goal of systematics in general, and species delimitation in particular, is the recognition of natural groups/monophyletic taxa, recent literature on this matter has acknowledged the inevitability that some morphologically-defined species pairs actually represent an ancestral-descendent lineage (see for example Rieseber & Brouillet, 1994; Knowles & Cartens, 2007; Yang & Smith, 2014; Naciry & Linder, 2015; Pennington & Lavin, 2015). In such cases, if the process of speciation is not yet complete to result in clearly distinct species (incomplete lineage sorting), one of the species (ancestral species) will be paraphyletic in the tree with the other species nested within.

Conceptually, we follow the framework proposed by de Queiroz (2005, 2007), which includes the idea that disagreements about species boundaries are especially prevalent in those species at the active *interface* of speciation (grey zone in Figure 1 in de Queiroz, 2007). This explains why many species have universally agreed boundaries, whereas others are more difficult to interpret. This is biological reality. In the context of taxonomy, and for those cases where species delimitation is particularly problematic, we take the view that it is best to flag up the situation for future studies, make a pragmatic, discursive and explicit taxonomic decision and move onto those other species delimitation problems that taxonomy can readily solve. This is especially true in large tropical groups such as *Ipomoea*, in which many of the taxonomic problems can be readily solved by having a good sample of specimens combined with good knowledge of the group's literature and nomenclature. Our view is that species and species delimitation can be seen as a heuristic, allowing an approach to problem-solving or discovery that employs

a practical method not guaranteed to be optimal or perfect, but sufficient for the immediate goals.

3.2.2 | Monophyly of the genus *Ipomoea* and other previously recognised groups

Our results confirm that the genus *Ipomoea*, as currently defined, is not monophyletic. All other genera in the tribe *Ipomoeeae* are nested inside *Ipomoea*, as suggested by previous studies using smaller data sets (Manos et al., 2001; Miller et al., 1999, 2002; Stefanovic et al., 2002, 2003). What previous studies did not show is that, among the smaller genera, only ***Astripomoea*** forms a monophyletic group: elements belonging to all other genera intermingle with various species of *Ipomoea* (Figure 3.8). This lack of monophyly provides further support for the most logical taxonomic decision: to expand the genus *Ipomoea* to include the smaller genera and make it monophyletic. In its new circumscription, *Ipomoea* includes all members of Convolvulaceae with spiny pollen and is the only genus in the tribe *Ipomoeeae*.

Regarding the subgenera and sections defined by Austin (Austin, 1975a, 1975b, 1979, 1980; Austin and Huáman, 1996), most of them are not supported in our phylogenies, which agrees with the results of Miller and colleagues (1999). In addition, here we show that the re-definition of the three subgenera by Miller *et al.* (1999) is also artificial and does not define monophyletic groups (Figure 3.9).

Our results show that most sections and series traditionally recognised in *Ipomoea* (see Appendix 2) are not monophyletic. On the contrary, the members of most groups recognised by Austin and others (Austin, 1975a, 1975b, 1979, 1980; Austin and Huáman, 1996; House, 1908; Miller et al., 1999; van Ooststroom, 1953; Verdcourt, 1957) are scattered in several different clades.

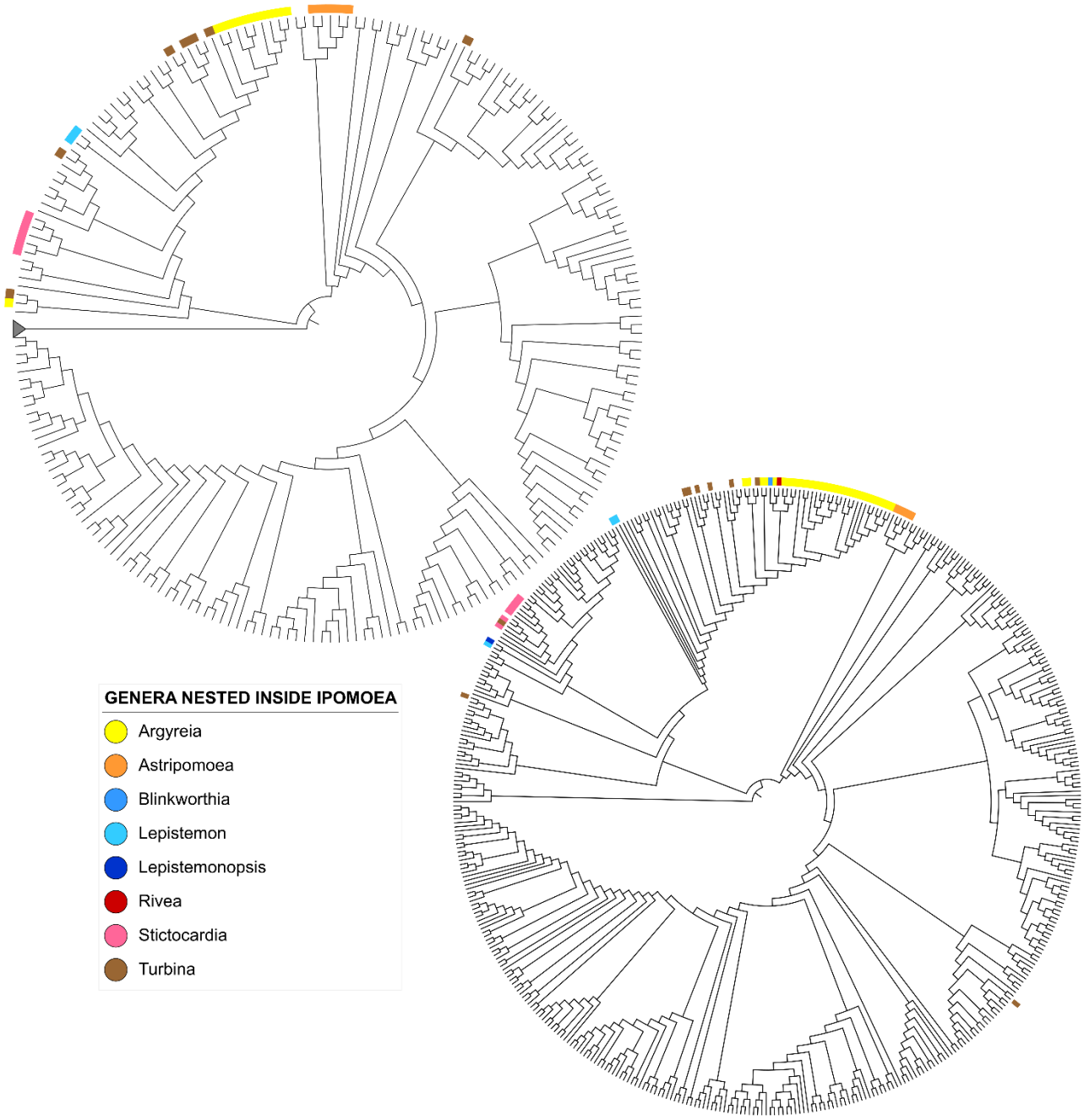


Figure 3.8. All smaller genera in *Ipomoeeae* are part of *Ipomoea*. Summary phylogenies of *Ipomoeeae* using (A) nuclear regions and (B) *ITS* sequences highlighting the position of specimens belonging to species from other genera in *Ipomoeeae*. Only one specimen per species is included in these trees. Colours in the outer circle indicate the position of a specimen belonging to a different genus, according to the colour code in the legend.

This is, for example, the case of Sections *Exogonium* and *Leptocallis* or Series *Jalapae*, *Setosae* and *Tyrianthinae*, some of the most frequently cited taxa in the literature. This, together with the fact that some of the clades recognised and strongly supported in our phylogenies lack diagnostic morphological characters, indicates that it

may not be possible to classify the diversity within *Ipomoea* into mutually exclusive taxa at the same rank as in a Linnean system. This reflects the impossibility of obtaining monophyletic, diagnostic and mutually exclusive groups (Carine and Scotland, 2002) following the principles of diagnosability of the traditional Linnaean system. The names that I use throughout this dissertation for the clades identified in our studies, aimed at facilitating the discussion of our results, are informal names. We did not attempt any formal recognition of infrageneric ranks as in Linnean taxonomy, although this has been attempted in other megadiverse genera (see for example Moonlight et al., 2018).

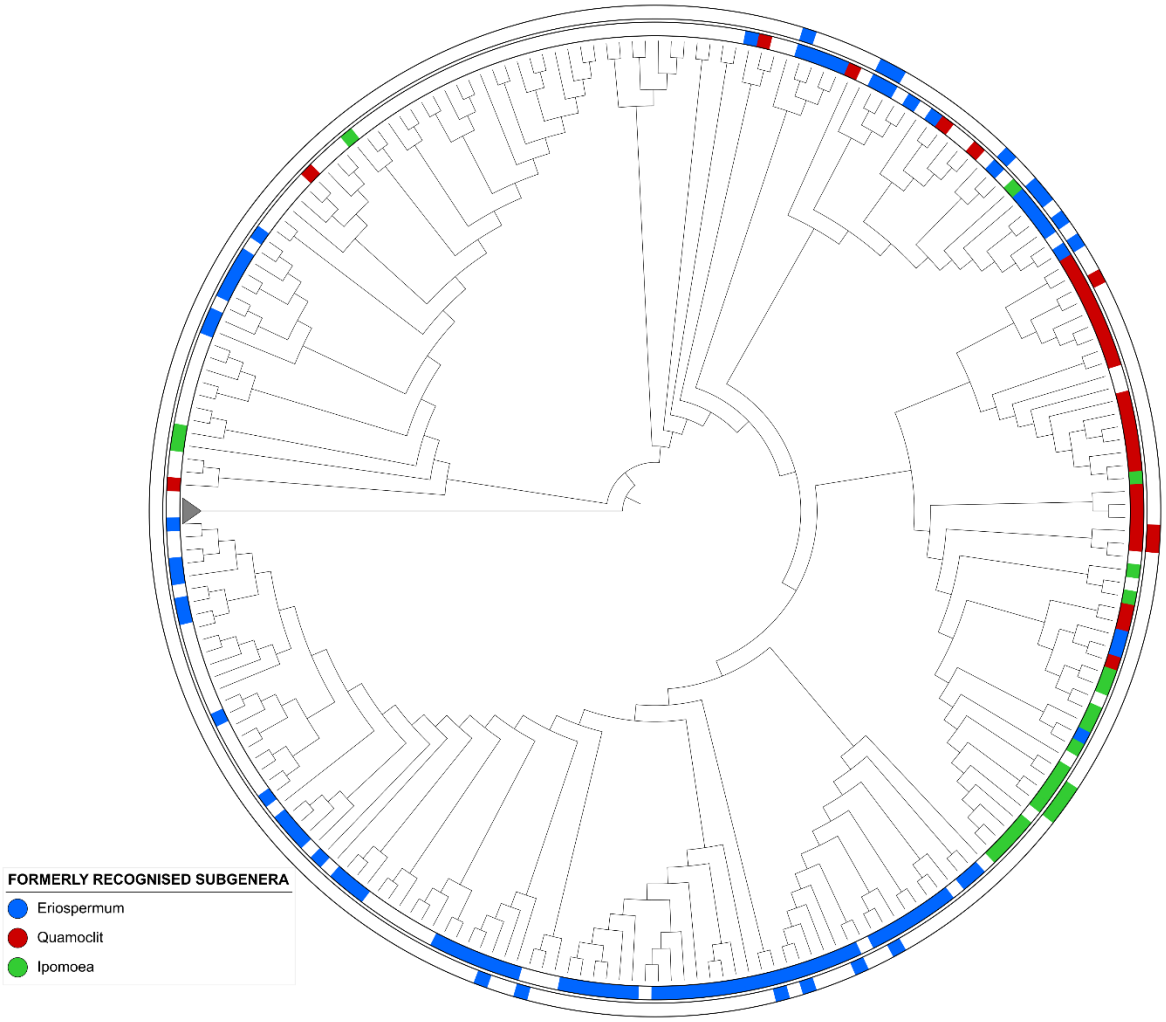


Figure 3.9. Nuclear phylogeny of *Ipomoea* showing the position of the species in the subgenera recognised by Austin (1977-1996) (inner circle), and Miller (1999) (outer circle). Not all species in the phylogeny were assigned to a specific subgenus by these authors, hence the lack of colour in some tips.

3.2.3 | Phylogenetic structure within *Ipomoea*

The phylogeny of *Ipomoea* is consistent between methods and genomic data sets, in the sense that the same main groups are monophyletic in all phylogenies (except B1, paraphyletic in the chloroplast phylogeny). Furthermore, the topology inferred from a given data set is constant regardless of the method of phylogenetic inference. In contrast and despite notably high overall topological congruence between data sets, there is incongruence between the nuclear and the chloroplast topologies in terms of species relationships within certain clades.

Ipomoea is divided in two major clades (Figure 3.10). The first one is dominated by taxa from the Old World (hereafter and in the figures **OW**) and the second clade includes a mainly African paraphyletic grade (*sensu* Huxley, 1957, 1958) and a New World (**NW**) clade. In the African grade, the largest clade is formed by all taxa previously included in *Astripomoea* and some species of *Ipomoea* (**ASTRIPOMOEA**)[17], whereas the two smaller clades include only *Ipomoea* species (see section 3.2.2.2). All species in this grade are restricted to Africa, hence I refer to it as the **AFRICAN GRADE**.

The position of the OW clade sister to ASTRIPOMOEA+NW, together with the fact that the closest relatives of *Ipomoea* in Convolvulaceae are mainly distributed in the Old World, strongly suggests that the genus *Ipomoea* had its origin somewhere in this region. This is congruent with the recent identification of *Ipomoea* Palaeocene fossil leaves (58.7-55.8 million years old) from India which, according to the authors and if their identification is correct, would support an Eastern Gondwana origin of the genus (Srivastava et al., 2018).

[17] In the *ITS* phylogeny, **ASTRIPOMOEA** is sister to the OW clade. However, both nuclear and chloroplast phylogenies, using much more data, place it next to the NW clade and we take that as its correct position.

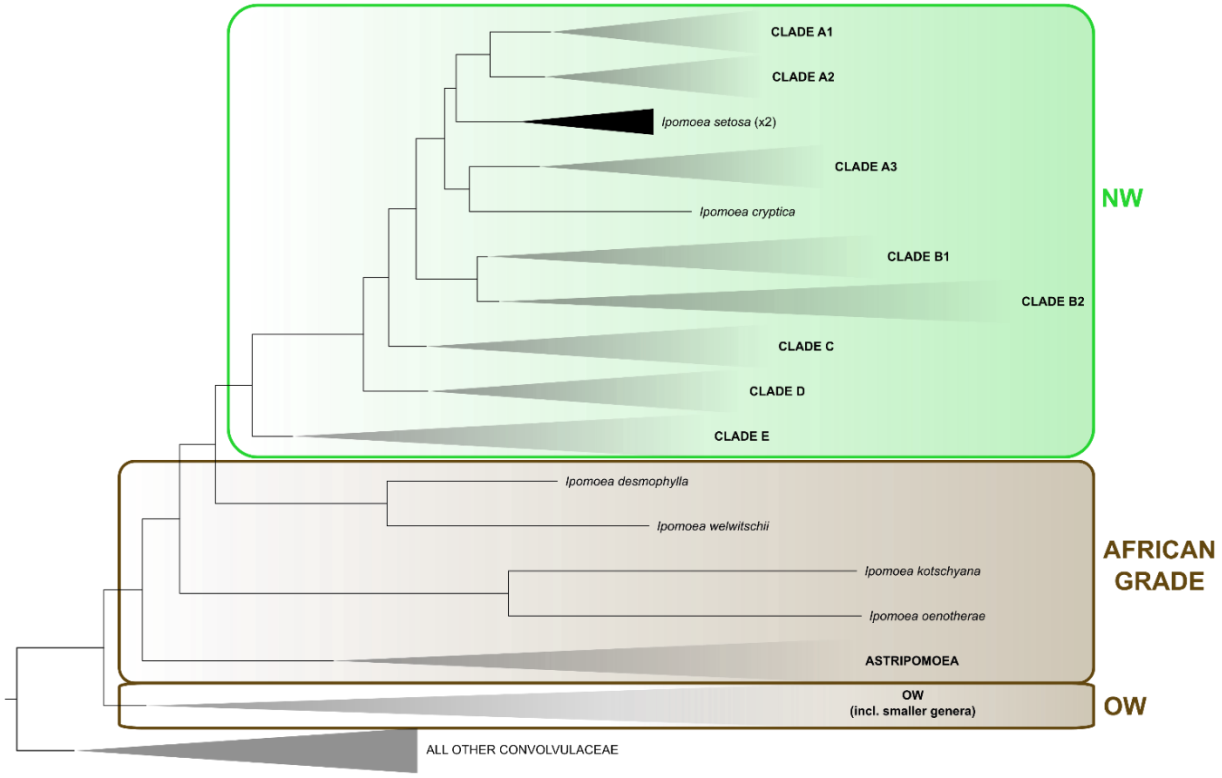


Figure 3.10. General geographical pattern in the nuclear phylogeny of *Ipomoea*. The early diverging position of the OW clade and African grade in this and the chloroplast phylogeny supports a Palaeotropical origin of the genus and a subsequent major split to the New World.

3.2.3.1 | OW CLADE

The OW clade includes around a third of the species in *Ipomoea* (Figure 3.11). Until now, the taxonomic studies conducted in Robert Scotland's group have largely focused on the American species of the genus and, for that reason, our sampling is certainly biased towards the New World.

Our study of the OW species is less advanced and I will not discuss this clade in as much detail as the NW clade. However, there are some patterns than can be observed in the OW clade that are worth highlighting, for example the fact that all genera formally recognised as independent except *Astripomoea*[18] belong to this clade[19] and that none of them is monophyletic (see Figure 3.8 and 3.11 and Figure A in Appendix 3).

[18] *Argyreia*, *Blinkworthia*, *Lepistemon*, *Lepistemonopsis*, *Paralepistemon*, *Rivea*, *Stictocardia* and *Turbina*.
[19] Among all representatives of these genera included in our study, only *Turbina amazonica* D.F.Austin & Staples is in the NW clade.

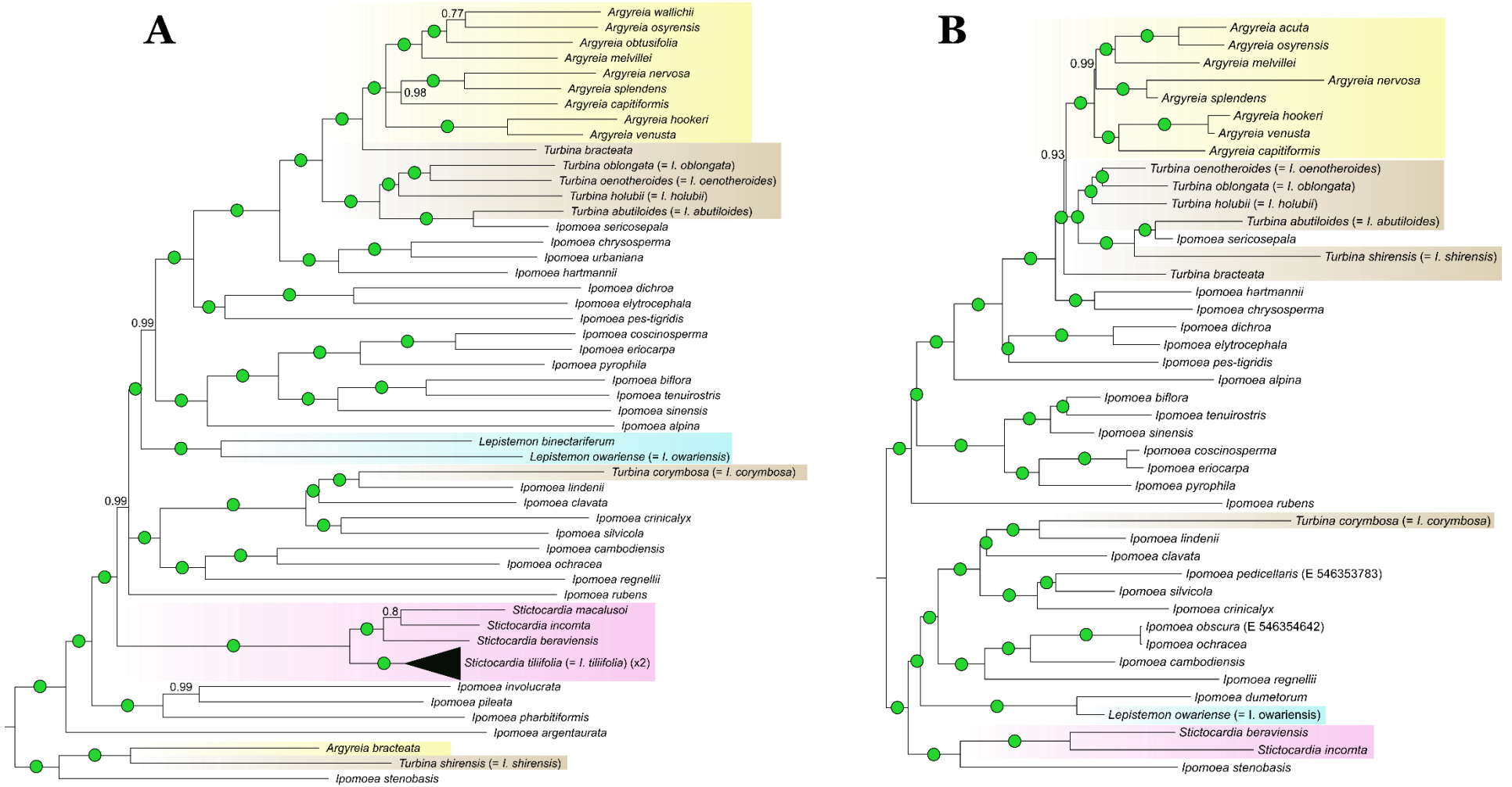


Figure 3.11. Detail of the OW clade in A) nuclear and B) chloroplast phylogenies, inferred using Approximate ML and Bayesian respectively. Numbers on the nodes represent support values (bootstrap support in A and posterior probability in B); green circles indicate 100% support. Colours indicate the species formerly recognised in other genera: yellow = *Argyreia*, blue = *Lepistemon*, pink = *Stictocardia*, brown = *Turbina*.

The fact that these segregate genera are nested within the OW clade is not surprising because those genera are almost entirely restricted to the Palaeotropical region. This relationship is, in any case, confirmed here for the first time. As indicated above, multiple nomenclatural changes are required to accommodate these species in *Ipomoea*. This new arrangement of *Ipomoea* is in preparation and will be published elsewhere.

Although the OW clade is dominated by species restricted to the Palaeotropics, at least one group consists entirely of American species and at least two other species have a distribution restricted to the New World. These cases are discussed in more detail in section 3.2.5.

The position of several species in the OW clade changes between phylogenies (Figure 3.11 and Figure A in Appendix 3), which could be a consequence of the sparser sampling for this part of the tree compared to the NW clade; in which case, the inclusion of more species that possibly belong to the OW clade will likely provide a more accurate and consistent placement for these species. Also, some groups such as *Stictocardia* have never been studied and are poorly known, and therefore little can be said about them at this point; a detailed study of this and other groups will identify other species belonging to this clade.

3.2.3.2 | AFRICAN GRADE

We identified three small clades that form a grade between the OW clade and the NW clade. These three clades are found in both nuclear and chloroplast phylogenies and the extra taxon sampling in the *ITS* phylogeny adds several species to each of them. All species in this grade are restricted in their distribution to the African continent and Madagascar.

One of the clades includes all taxa formerly recognised in *Astripomoea* (forming a sub-clade) and five species of *Ipomoea*: *I. hackeliana* (Schinz) Hallier f., *I. leucanthemum* Hallier f., *I. papilio* Hallier f., *I. simonsiana* Rendle and *I. vagans* baker (Figure 3.12).

The other two clades in the African grade consist of six[20] and three[21] species respectively. Our *ITS* phylogeny includes multiple samples for most species[22], retrieving three of them as monophyletic (*I. bathycolpos, I. keraudreniae* and *I. sepiaria*) (Figure 3.4 in page 54). In summary, there is strong molecular support for the existence of these two clades, although additional studies are required to clarify the boundaries between species.

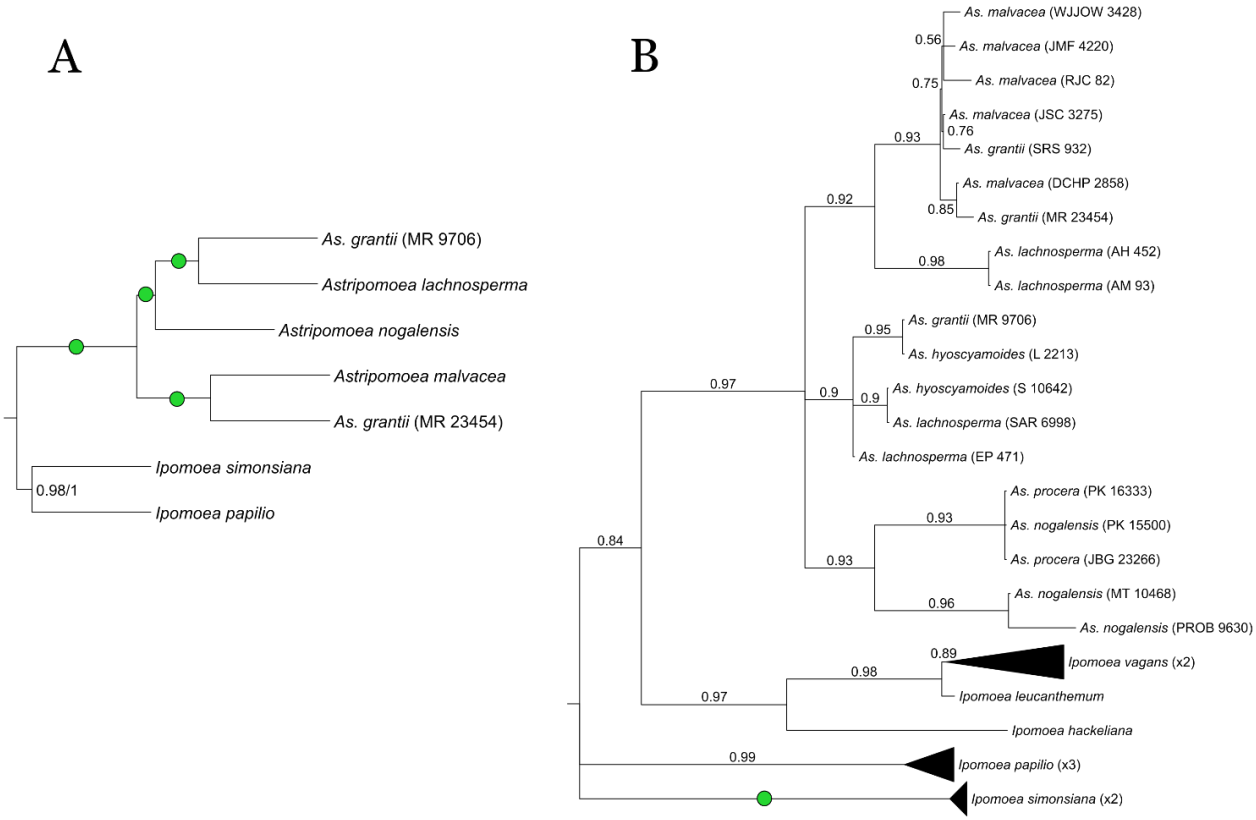


Figure 3.12. ASTRIPOMOEA clade in A) the nuclear phylogeny and B) the *ITS* phylogeny, inferred using Approximate ML. Numbers on the nodes represent bootstrap support; green circles indicate 100% support.

[20] *Ipomoea bathycolpos* Hallier f., *I. bolusiana* Schinz, *I. desmophylla* Bojer ex Choisy, *I. keraudreniae* Deroin, *I. venosa* (Desr.) Roem. & Schult. and *I. welwitschii* Vatke ex Hallier f.

[21] *Ipomoea coptica* (L.) Roth ex Roem. & Schult., *I. sepiaria* Koenig ex Roxb. and *I. ticcopa* Verdc.

[22] Only *Ipomoea ticcopa* and *I. venosa* are represented by a single specimen.

3.2.3.3 | NW CLADE

The NW clade includes about two thirds of the total number of species of *Ipomoea*. In addition, several subclades can be identified and diagnosed morphologically. I refer to these clades as **CLADE A** (further divided in A1, A2 and A3), **CLADE B** (B1 and B2), **CLADE C**, **CLADE D** and **CLADE E**. Although the subclades within CLADE A are always retrieved, the relationship between them changes in the nuclear and the chloroplast phylogenies: [[A1,A2],A3] and [[A1,A3],A2] respectively. Similarly, the position of CLADES B1, B2 and C changes between the nuclear and the chloroplast phylogenies (Figures 3.2–3.4).

3.2.3.3.1 | CLADE A

With around 250 species, this is by far the largest clade in NW *Ipomoea*. It can be further divided in three sub-clades: A1, A2 and A3, and species belonging to any of these sub-clades can be identified on the basis of diagnostic morphological characters.

CLADE A1 is the largest of the three sub-clades and received strong support in all our phylogenies. In addition, the *ITS* phylogeny, although poorly resolved compared with other parts of the tree (Figure B in Appendix 3), places in this clade 52 additional species that were not sampled for Hyb-Seq. A more comprehensive sampling using Hyb-Seq data would help clarify the relationships in some parts of this clade, although our results reveal several interesting patterns. For example, all species of *Ipomoea* that have a tree habit form a clade (Table 3.1 and Figure 3.13), and the relationship between them is confirmed by both nuclear and chloroplast phylogenies with 100% support.

The clade of tree species is closely related to a puzzling group, represented by only three species in our nuclear and chloroplast phylogenies but complemented with 10 more in the *ITS* phylogeny (Table 3.2). This clade is interesting because it includes a heteroge-

neous group of species from different parts of the world: several species with a widespread distribution in North America and/or South America; two species endemic to Brazil; one species from SE Asia; and several species endemic to Australia[23].

Table 3.1. Tree species in *Ipomoea*

Ipomoea arborescens (Humb. & Bonpl. ex Willd.) G.Don
Ipomoea intrapilosa Rose
Ipomoea murucoides Roem. & Schult.
Ipomoea pauciflora M.Martens & Galeotti
Ipomoea populina House
Ipomoea rzedowskii E.Carranza, Zamudio & Murguia*
Ipomoea seaania Felger & D.F.Austin*
Ipomoea teotitlanica McPherson*
Ipomoea wolcottiana Rose

* Species included only in the *ITS* phylogeny.

Table 3.2. Species in the clade sister to the trees (see also Figure 3.13).

I. abrupta R.Br.	Endemic to Australia
I. calobra Hill. & F.Muell. ex Benth.	Endemic to Australia
I. campanulata L.	SE Asia
I. connata J.R.I.Wood & L.V.Vasconc.	Endemic to Brazil
I. costata F.Muell. ex Benth.	Endemic to Australia
I. graniticola J.R.I.Wood & Scotland	Endemic to Brazil
I. longifolia Benth.	Mexico and United States
I. polpha R.W.Johnson	Endemic to Australia
I. pterocaulis J.R.I.Wood & L.V.Vasconc.	Brazil
I. reticulata O'Donell	Widespread in the Americas
I. rosea Choisy	South America
I. saopaulista O'Donell	Brazil
I. sumatrana (Miq.) Ooststr.	China, Taiwan and SE Asia

A group of species within CLADE A1 is identified as monophyletic in all phylogenies. This species-rich clade is termed NORTHERN SOUTHERN CONE in Figure 3.13 because it has a distribution restricted to the northern part of the Southern Cone in South America.

CLADE A1 also includes several species of *Ipomoea* recently described from Brazil and Bolivia, for example *Ipomoea graniticola* J.R.I.Wood & Scotland, *I. pterocaulis* J.R.I.Wood & L.V.Vasconc. and *I. queirozii* J.R.I.Wood & L.V.Vasconc.

[23] Although Austin and colleagues (1993) identified a close relationship between these species, they wrongly placed them as allies of *I. gracilis* R.Br., which as our phylogenies reflect is in CLADE C.

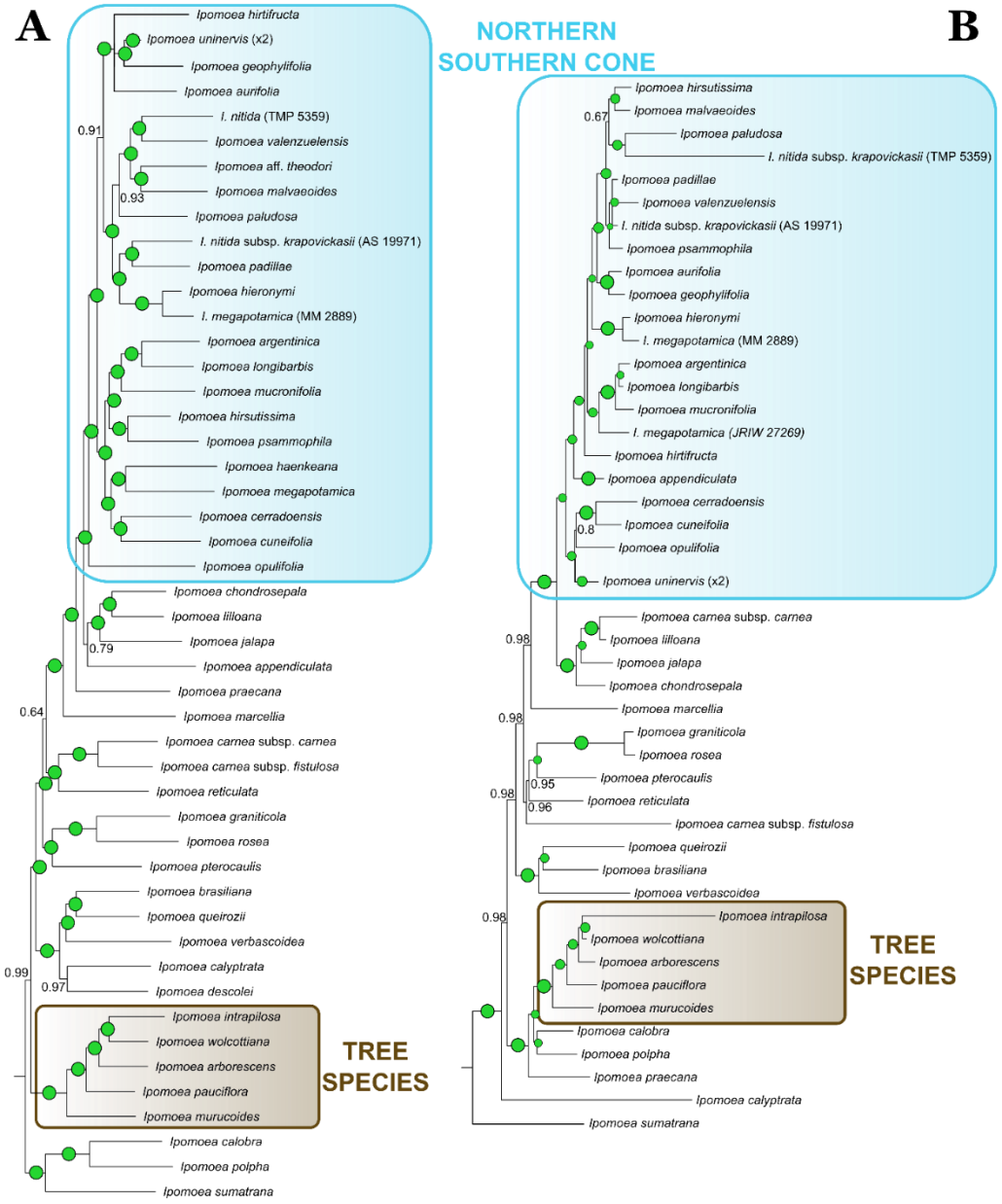


Figure 3.13. Detail of CLADE A in A) Nuclear and B) chloroplast phylogenies, inferred using Approximate ML and Bayesian respectively. Numbers on the nodes represent support values (bootstrap support in A and posterior probability in B); green circles indicate 100% support.

Next in size within CLADE A is **CLADE A2**. Represented by 15 species in our nuclear and chloroplast phylogenies (Figure 3.14) and 25 additional species in the *ITS* phylogeny (Figure C, Appendix 3). We estimate that this clade consists of around 80 species, most of them restricted to the Americas but at least two from the Palaeotropical region (*Ipomoea asterophora* Ooststr. and the widespread *I. mauritiana* Jacq.). The species in this clade can be identified from morphology, as they have coriaceous sepals, and

the highest diversity is found around the Caribbean region and South America (Wood et al., 2015b: 71).

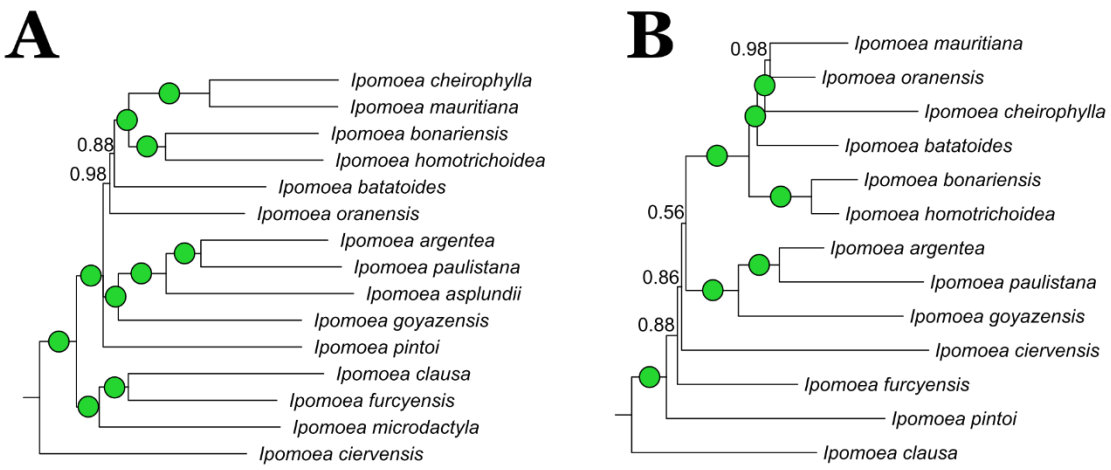


Figure 3.14. Detail of CLADE A2 in A) nuclear and B) chloroplast phylogenies, inferred using Approximate ML and Bayesian respectively. Numbers on the nodes represent support values (bootstrap support in A and posterior probability in B); green circles indicate 100% support.

It is difficult to study evolutionary relationships within CLADE A2, partly because our Hyb-Seq phylogenies only include a small representation of the species in the group. However, we can observe some patterns common to both nuclear and chloroplast phylogenies. For example, a group of five species (*Ipomoea batatoides* Choisy, *I. bonariensis* Hook., *I. cheirophylla* O'Donell, *I. homotrichoidea* O'Donell and *I. mauritiana*) form a highly supported group in the nuclear and chloroplast phylogenies, and the *ITS* tree adds several other species to the group. In the *ITS* tree, fifteen species represented by two or more specimens are monophyletic, whereas seventeen are represented by one specimen only and their monophyly cannot be assessed. A more comprehensive taxon sampling using Hyb-Seq will allow a better understanding of the relationships within this group and the assessment of the delimitation of other species.

Especially relevant for the following chapters of this dissertation is **CLADE A3**. This clade includes the **sweet potato**, *Ipomoea batatas*, and 14 other species and receives 100% support in all our phylogenies, regardless of the data set or method of phylogenetic

inference. The wild species in this clade that share a recent evolutionary history with the sweet potato are the **crop wild relatives (CWR)** (Castañeda-Álvarez et al., 2016; Maxted et al., 2006). I discuss this clade in detail in Chapter 4.

Interestingly, two species in CLADE A are not included in any of the sub-clades aforementioned but represent independent lineages. These species are *Ipomoea setosa* (including *I. sepacuitensis* Donn.Sm.) and *I. cryptica* J.R.I.Wood & Scotland. *Ipomoea setosa* (Figure 3.15) is a species widespread in tropical America up to Mexico, characterised by its stems and sepals with fleshy spines and frequently used as indicator plant for sweet potato grafting (Wood et al., 2015b). *Ipomoea cryptica* was recently described from Bolivia (Wood et al., 2015b) and later found in Brazil (John Wood, pers. comm.). Plants now assigned to *I. cryptica* were previously thought to be *I. squamosa* Choisy, which is morphologically identical. Our extensive phylogenies showed that in fact these specimens belong to a new, cryptic entity that seems to be only distantly related to the rest of taxa in CLADE A. Both *I. setosa* and *I. cryptica* have different position in the Hyb-Seq phylogenies. In the nuclear phylogeny, *I. cryptica* it is sister to CLADE A3 and *I. setosa* sister to clades A1+A2, whereas in the chloroplast phylogeny *I. cryptica* is sister to a clade formed by A1, A3 and *I. setosa*.

Figure 3.15. *Ipomoea setosa* in Bolivia.

Finally, the *ITS* phylogeny retrieves one other clade that I labelled as "UNASSIGNED" in Figure 3.4. This clade comprises around ten Mexican species[24], but has not been sampled for Hyb-Seq. It may be part of CLADE A1 or A2, but it requires further investigation using additional molecular data.

3.2.3.3.2 | CLADE B

CLADE B includes some well-known species in the morning glory family for their use in gardening and also as invasive species in different parts of the world, for example ***Ipomoea alba*** (white morning glory), ***I. indica*** (blue morning glory), ***I. nil*** (Japanese morning glory), ***I. purpurea*** (common morning glory) or ***I. tricolor* Cav.** (Mexican morning glory). Species in this clade have been traditionally classified in subgenus *Quamoclit*, section *Pharbitis* and various other sections and series, although as I showed before these taxa are not monophyletic groups and their use must be avoided.

The nuclear and chloroplast phylogenies consist of 41 and 44 species respectively (Figure 3.16), to which the *ITS* phylogeny adds 35 species (Figure D in Appendix 3), giving a total of 79 species sequenced from this group. Although the nuclear phylogeny can be further divided in two reciprocally monophyletic **CLADES B1** and **B2**, in the chloroplast topology CLADE B1 is paraphyletic (Figures 3.3). All species represented by multiple specimens in the nuclear phylogeny are monophyletic, whereas neither *Ipomoea dumosa* (Benth) L.O.Williams nor *I. costellata* Torr. are monophyletic in the chloroplast phylogeny. In addition, thirty-six species are monophyletic in the *ITS* phylogeny, whereas nineteen are represented by a single specimen and their monophyly cannot be assessed.

Several species in CLADE B have a widespread distribution. The Hyb-Seq phylogenies include only one specimen for each of those species except for *I. indica*,

[24] *Ipomoea ciervensis* Painter, *I. cuprinacoma* E.Carranza & J.A.McDonald, *I. durangensis* House, *I. hartwegii* Benth., *I. lenis* House, *I. lottiae* J.A.McDonald, *I. peteri* (Kuntze) Staples & Govaerts, *I. proxima* M.Martens & Galeotti, *I. scopulorum* Brandegee and *I. simulans* D.Hanb.

whereas the *ITS* phylogeny[25], largely congruent and including several specimens for each species, provides additional information.

The combination of both sources allows further interpretation of the relationships between species, for example, it is possible that plants identified as *Ipomoea indica* represent two distinct entities. The nine specimens in our *ITS* phylogeny segregate in two clades: one of them includes four specimens (the same four specimens that we sequenced using Hyb-Seq and form a monophyletic clade with 100% support) and several other species: two other widespread species, *I. purpurea* and *I. pubescens* Lam., both monophyletic; the Jamaican endemic *I. jamaicensis* G.Don; and *I. neurocephala* Hallier f. and *I. thurberi* A.Gray, both restricted to south-western USA (Austin, 2006). The other five *I. indica* specimens form a clade with the widespread species *I. nil* and the North American species *I. hederacea* Jacq. and *I. lindheimeri* A.Gray, restricted to northern Mexico and south-western USA (Austin, 1992). The close relationship between these three species had been already indicated (Austin, 1975a; Austin et al., 2001; Miller et al., 2004), and Austin (1975a) even hypothesised that *I. nil* and *I. hederacea* could be different forms of the same species, although other studies identified certain reproductive barriers between species in this group (Ennos, 1981; Smith and Rausher, 2007). Additional studies are necessary to corroborate these relationships and to define boundaries between species.

In CLADE B2, *Ipomoea quamoclit* L. and *I. hederifolia* L., two other well-known species, are both monophyletic. The latter is in a sub-clade with three Mexican species, which could indicate that this species, the geographical origin of which was not known, originated somewhere in this region.

[25] The results of our *ITS* phylogeny largely agree with Miller and colleagues (Miller et al., 2004), which is unsurprising because they also used *ITS*.

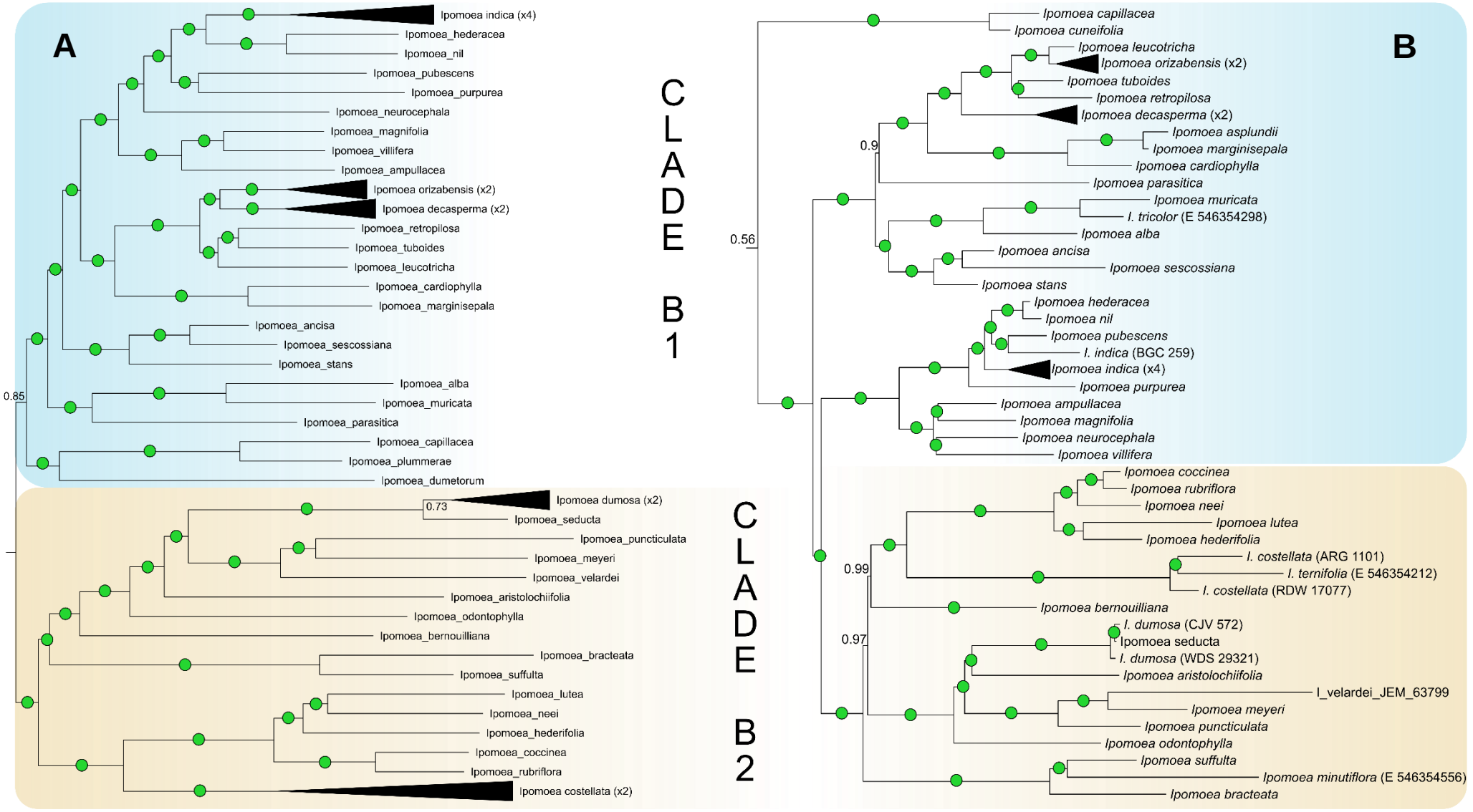


Figure 3.16. Detail of CLADE B in A) nuclear phylogeny and B) chloroplast phylogeny, inferred using Approximate ML and Bayesian respectively. Numbers on the nodes represent support values (bootstrap support in A and posterior probability in B); green circles indicate 100% support.

3.2.3.3.3 | CLADE C

CLADE C contains fewer species than CLADES A and B. The nuclear and chloroplast phylogenies include 20 and 22 species respectively (Figure 3.17), whereas the *ITS* phylogeny adds nine more (Figure E in Appendix 3). In the *ITS* phylogeny, 16 species are monophyletic.

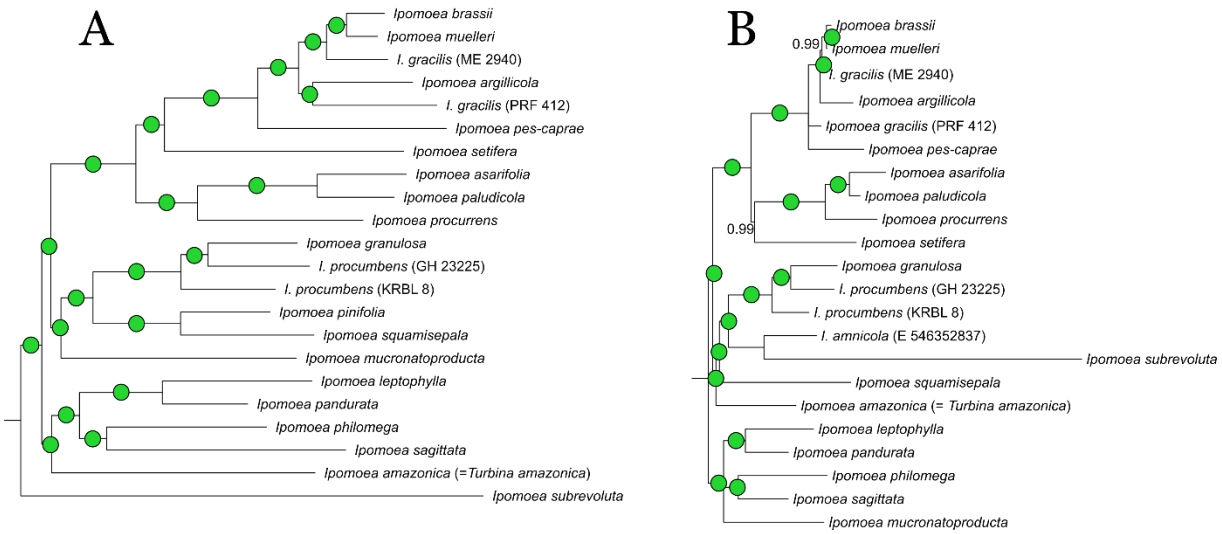


Figure 3.17. Detail of CLADE C in A) nuclear phylogeny and B) chloroplast phylogeny, inferred using Approximate ML and Bayesian respectively. Numbers on the nodes represent support values (bootstrap support in A and posterior probability in B); green circles indicate 100% support.

The relationship between *I. procumbens* Mart. ex Choisy, *I. granulosa* Chodat & Hassl. and *I. rupestris* Sim-Bianch. & Pirani is worth exploring further: *I. granulosa* and *I. rupestris*, two species with a restricted distribution in Brazil and Paraguay (Chodat and Hassler, 1905; Simão-Bianchini and Rubens Pirani, 2005) are nested within the more widespread *I. procumbens* in all our phylogenies (which is not monophyletic); this could indicate either conspecificity or recent divergence of these entities.

Also of interest is the position of *Turbina amazonica* (= *I. amazonica*), distributed across South America, which is the only species of *Turbina* in the NW clade.

Finally, an outstanding element of CLADE C is the sub-clade formed by *Ipomoea pes-caprae*, a coastal species from sandy areas present in most tropical regions of the

world, and several species restricted to Australia: *Ipomoea brassii* C.T.White, endemic to a small area in northern Australia; *I. gracilis* R.Br., endemic to northern Australia[26]; *I. muelleri* Benth., more widespread but still restricted to Australia; and *I. argillicola* R.W.Johnson, endemic to north-eastern Australia (Austin, 1991; Austin et al., 1993). To these species, the *ITS* phylogeny adds two other species: *I. fimbriosepala* Choisy in A.DC., a species widespread in the Americas, Tropical Africa, Australia and New Guinea, and *I. setifera* Poir., also widespread in the Americas and Tropical Africa. This clade is the largest clade of Palaeotropical species nested within the NW.

3.2.3.3.4 | CLADE D

CLADE D is a small, strongly supported clade of five species: *Ipomoea acanthocarpa* (Choisy) Asch. & Schweinf., *I. bahiensis* Willd. ex Roem. & Schult., *I. imperati* (Vahl) Griseb., *I. longeramosa* Choisy and *I. squamosa* Choisy. The relationship between species is the same in both nuclear and chloroplast phylogenies except for *I. acanthocarpa* (Figure 3.18). The *ITS* phylogeny (Figure F in Appendix 3), which includes multiple samples for each species, adds two more species to this clade: *I. eriocalyx* Mart. ex Choisy) Meisn. and *I. kraholandica* J.R.I.Wood & Scotland, a species recently described (Wood et al., 2017b). The *ITS* phylogeny shows that all species are monophyletic with > 94% support except *I. squamosa*, which is paraphyletic.

[26] *Ipomoea gracilis*, a species frequently confounded with the sweet potato CWR *I. littoralis*, is not monophyletic. Our studies show that plants referred to as *I. gracilis* may comprise two different entities. This species needs further, more comprehensive molecular and morphological studies.

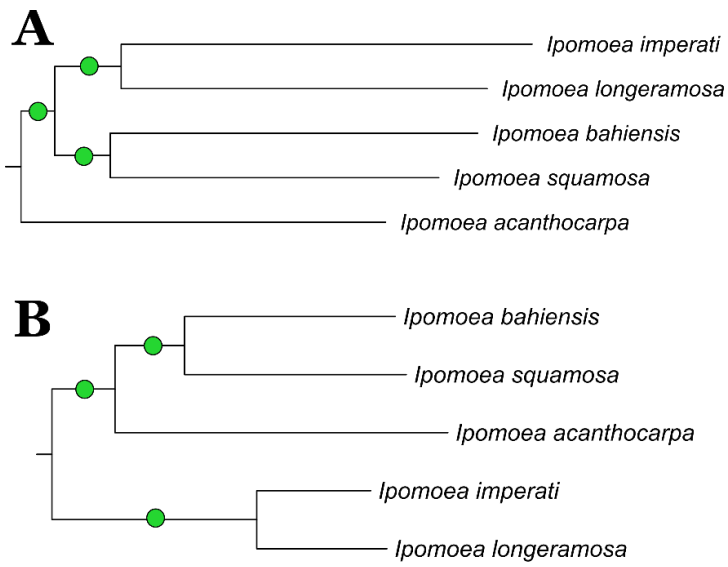


Figure 3.18. Detail of Clade D from A) nuclear and B) chloroplast phylogenies, inferred using Approximate ML and Bayesian respectively. Green circles indicate 100% support (bootstrap support in A and posterior probability in B).

3.2.3.3.5 | CLADE E

CLADE E is a small, 100% supported clade in all our phylogenies. The nuclear and chloroplast phylogenies include only 3 and 4 species respectively in this clade (Figure 3.19), whereas the *ITS* phylogeny contributes 10 species more that were not sampled for Hyb-Seq (Figure G in Appendix 3). *Ipomoea kotschyana* and *I. oenotherae*, two species in the African grade, are placed in this clade in the *ITS* phylogeny.

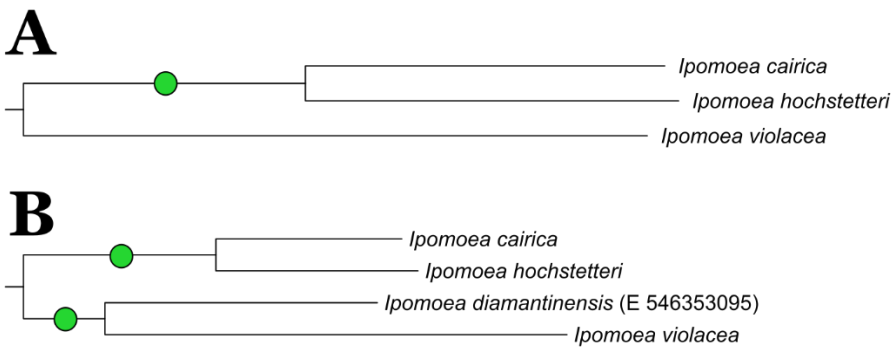


Figure 3.19. Detail of Clade E from A) nuclear and B) chloroplast phylogenies, inferred using Approximate ML and Bayesian respectively. Green circles indicate 100% support (bootstrap support in A and posterior probability in B).

Ipomoea aquatica, the water spinach, is the other crop species in *Ipomoea* and a member of this clade. Nowadays, it is distributed worldwide for its weedy behaviour, but it is thought to be native to China, SE Asia and the Western Pacific islands (Austin, 2007a;

Kaur et al., 2016). All species in this clade except *I. habeliana* Oliv. are Palaeotropical, which further supports a Palaeotropical origin of the water spinach. This is also the possible origin of *I. cairica*, another member of this clade with a widespread distribution, introduced in some regions and now invasive (Srivastava and Shukla, 2015).

3.2.4 | Evolution of storage roots in *Ipomoea*

Numerous species of *Ipomoea* produce enlarged storage roots, in some cases the size of the sweet potato roots or even bigger. Table 3.3 is a list of 48 species that are mentioned in literature as developing storage roots. This list is not comprehensive, so the number of species with storage roots is likely to increase if a detailed literature review is conducted and also if future collectors pay attention to the underground structures when collecting *Ipomoea*. For example, the root of *Ipomoea lilloana* shown in Figure 3.20 had never been dug up and its existence was therefore unknown (John Wood, *pers. comm.*). These 48 species are not grouped in a clade but are scattered across the phylogeny of *Ipomoea* (Figure 3.21). Also, in most cases the species with enlarged roots are sister to species that do not have storage roots. The evidence thus strongly indicates that the development of storage roots is a homoplastic character in *Ipomoea* and evolved multiple times independently.

As in the sweet potato, several other species have edible roots and are consumed in different parts of the world (Table 3.3). *Ipomoea albivenia* Sweet and *I. bolusiana* Schinz, for example, are eaten in South Africa, Mozambique and Zimbabwe, especially in times of famine (Hargreaves, 1994; Jacobs, 2000; Mongalo and Makhafola, 2018; South African National Biodiversity Institute, 2018). In Mexico, Gentry (1942) found that the roots of *I. bracteata* Cav. are *«in high repute among the natives for their edibility»*. Also, a group of at least five Brazilian species known by local people as "batata da serra" for their similarity with the sweet potato (*batata* in Brazil), are consumed and sold in markets: *I.*

longistaminea O'Donell, *I. pintoi* O'Donell, *I. rupestris* Sim.-Bianch. and the recently described *I. ana-mariae* L.V.Vasconcelos & Sim.-Bianch. and *I. serrana* L.V.Vasconcelos & Sim.-Bianch. (Vasconcelos et al., 2016). In Australia, several species of *Ipomoea* have edible roots, for example *I. polpha* R.W.Johnson, the specific epithet of which derives from the Greek *polphos*, "valuable food" (Johnson, 1986). Finally, several authors report consumption of several species of *Ipomoea* in the Pacific islands; for example, the roots of *I. cairica* and *I. pes-caprae* var. *brasiliensis* were cooked and consumed across all islands in Hawaii, especially in times of famine (Chock, 1968; Handy, 1940; Malo, 1898). Despite being eaten and traded by humans, there is no record of active cultivation for most of these species; rather, they seem to be gathered from the wild (Handy, 1940; Vasconcelos et al., 2016).

Figure 3.20. Storage roots of *Ipomoea lilloana* O'Donell, a South American species.

Table 3.3. Species of *Ipomoea* with storage roots. Names in bold indicate species recorded as food.

SPECIES*	REFERENCE
I. albivenia	Jacobs, 2000; Mongalo and Makhafola, 2018
I. ampullacea	Felger et al., 2012
I. ana-mariae	Vasconcelos et al., 2016
I. ancisa	McDonald, 2001
I. argillicola	Johnson, 1986
I. batatas	Felger et al., 2012
I. bolusiana	Hargreaves, 1994; Mongalo and Makhafola, 2018; Welman and Meeuse, 1998
I. bracteata	Felger et al., 2012; Gentry, 1942
I. cairica	Chock, 1968; Deroin, 1999
I. capillacea	Felger et al., 2012; McDonald, 1995
I. conzatii	Lipp, 1971
I. digitata	Meira et al., 2012
I. elongata	McDonald, 1987
I. funicularis	Johnson, 2012
I. holubii	Hargreaves, 1994
I. jalapa	Linajes et al., 1994
I. jicama	Brandegee, 1889; McDonald, 1987
I. leptophylla	Meira et al., 2012
I. lilloana	John Wood, *pers. comm.*
I. lindheimeri	Eserman, 2012
I. longiflora	Felger et al., 2012
I. longifolia	Felger et al., 2012
I. longistaminea	Vasconcelos et al., 2016
I. macrorhiza	Austin, 2011
I. madrensis	Felger et al., 2012; McDonald, 1995
I. mauritiana	von Jacquin, 1790; Powell, 1979
I. muricata	Gentry, 1942
I. oenotherae	Welman and Meeuse, 1998
I. orizabensis	McDonald, 2001
I. pandurata	Horak and Wax, 1991
I. pedicellaris	Felger et al., 2012
I. pintoi	Vasconcelos et al., 2016
I. plummerae	Austin, 1998; Felger et al., 2012; McDonald, 1995
I. polpha	Johnson, 1986
I. pubescens	Eserman, 2012; Felger et al., 2012
I. purga	Hanbury, 1871; Linajes et al., 1994; McDonald, 1987
I. repanda	Powell, 1979
I. rupestris	Vasconcelos et al., 2016
I. serrana	Vasconcelos et al., 2016
I. sescossiana	Felger et al., 2012; McDonald, 2001
I. simulans	Hanbury, 1871; McDonald, 1987
I. stans	McDonald, 2001; Meira et al., 2012
I. tastensis	McDonald, 1987
I. tenuiloba	Austin, 1998; Felger et al., 2012; McDonald, 1995
I. thurberi	Austin, 1998; Felger et al., 2012
I. tolmerana	Johnson, 2012
I. violacea	Cheeseman, 1903
I. welwitschii	Welman and Meeuse, 1998

*in addition to these species, *I. jacalana*, *I. laeta*, *I. parasitica* and *I. scopulorum* are cited by McDonald (2001) and Felger *et al*. (2012) as having enlarged roots, but this information is not clear and needs confirmation.

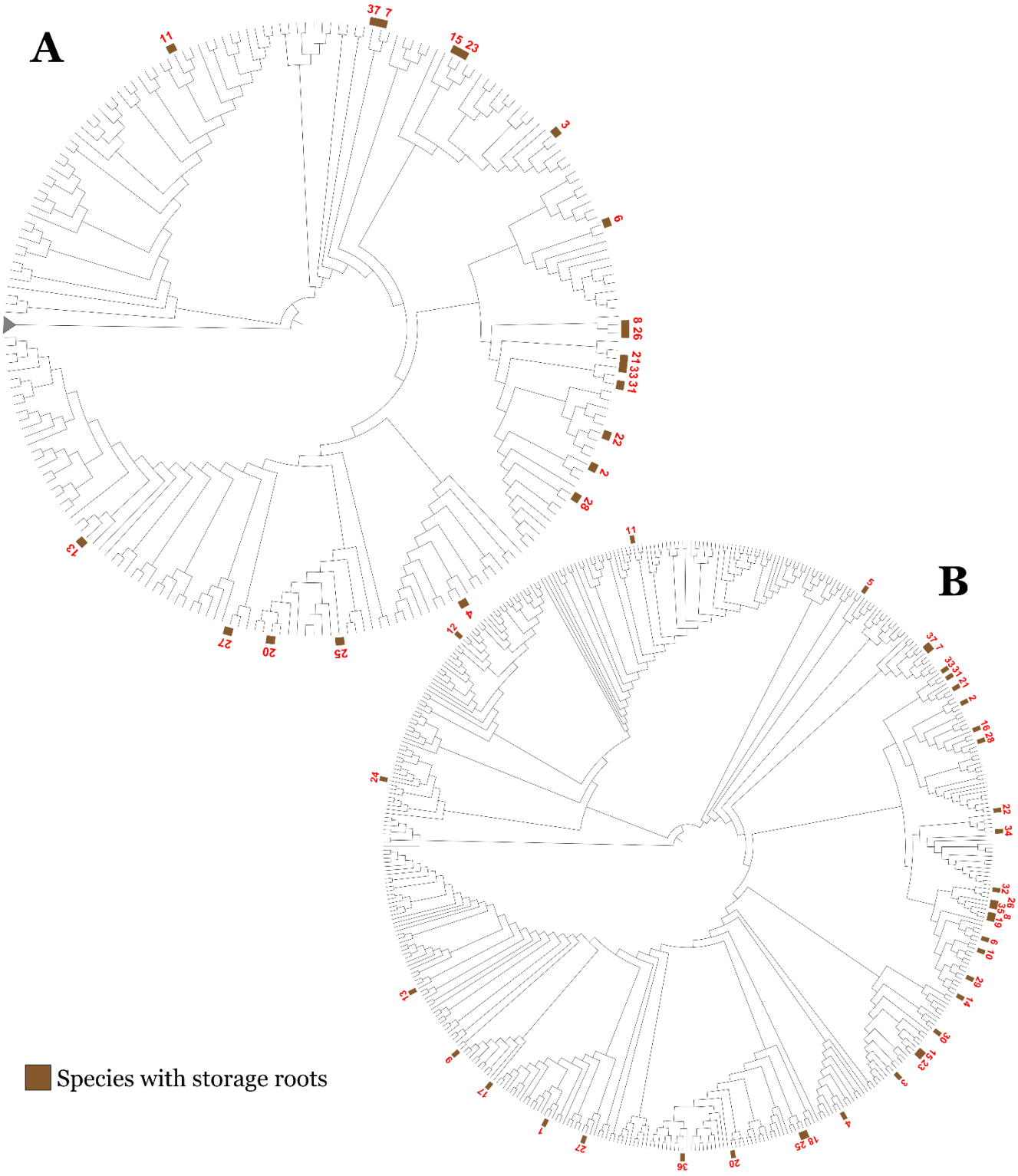


Figure 3.21. Summary phylogenies of *Ipomoeeae* using (A) nuclear regions and (B) *ITS* sequences highlighting the position of species with storage roots. Each brown bar indicates a species for which storage roots have been recorded: 1. *I. albivenia*; 2. *I. ampullacea*, 3. *I. argillicola*, 4. *I. batatas*, 5. *I. bolusiana*, 6. *I. bracteata*, 7. *I. cairica*, 8. *I. capillacea*, 9. *I. elongata*, 10. *I. holubii* (= *Turbina holubii*), 11. *I. jalapa*, 12. *I. jicama*, 13. *I. leptophylla*, 14. *I. lindheimeri*, 15. *I. longifolia*, 16. *I. longistaminea*, 17. *I. madrensis*, 18. *I. mauritiana*, 19. *I. muricata*, 20. *I. orizabensis*, 21. *I. pandurata*, 22. *I. pedicellaris*, 23. *I. pintoi*, 24. *I. plummerae*, 25. *I. polpha*, 26. *I. pubescens*, 27. *I. purga*, 28. *I. rupestris*, 29. *I. sescossiana*, 30. *I. simulans*, 31. *I. stans*, 32. *I. tastensis*, 33. *I. tenuiloba*, 34. *I. thurberi*, 35. *I. violacea*.

In terms of their geographical distribution, there is no common pattern among the species that produce storage roots, either edible or not. Some species are endemic of certain environments and are restricted to a specific area (e.g. *Ipomoea ana-mariae*, *I. serrana*) or country (e.g. *I. ampullacea* Fernald, *I. argillicola*, *I. conzattii* Greenm., *I. elongata* Choisy, *I. jicama* Brandegee, *I. lindheimeri*). Some other species are more widely distributed across a single geographical region, for example Tropical Africa (e.g. *I. albivenia*), Southern Africa and Madagascar (*I. bolusiana*, *I. holubii* Baker) or the Caribbean (e.g. *I. digitata* L.). Finally, other species have a widespread distribution across one or more continents (e.g. *I. cairica*, *I. capillacea* (Kunth) G.Don, *I. jalapa* (L.) Pursh, *I. mauritiana*, *I. plummerae* A.Gray or *I. violacea* L.).

3.2.5 | Long distance dispersal in *Ipomoea*

As its name suggests, long-distance dispersal by natural means is the diffusion of plants across long distances without the mediation of humans. If the disjunct distribution pattern of a given taxon is not a consequence of human activity, it is logical to assume that it had to be mediated by other agents, most likely sea currents, wind or animals. For example, the distribution of the coconut palm (*Cocos nucifera* L.) thanks to its floating fruits is a classic example of long-distance dispersal mediated by sea currents (Harries and Clement, 2014).

There are few studies on dispersal ability and long-distance dispersal in *Ipomoea*, but those published are meticulous and provide extensive information on the topic. For example, flotation of seeds across the ocean has been convincingly presented to explain the pantropical distribution of *Ipomoea imperati*, *I. indica*, *I. littoralis* Blume and *I. pes-caprae* (Austin et al., 2001; Carlquist, 1967; Miryeganeh et al., 2014; Ridley, 1930).

A pioneer on studies of long-distance dispersal was Henry B. Guppy, a British naturalist who published a series of books with his experimental observations on plant dispersal in the Pacific region. In his most extensive book, Guppy (1906) studied twelve species of *Ipomoea* and concluded that the seeds of coastal species float (he listed *I. pes-caprae*, *I. violacea* and *Stictocardia tiliifolia* (= *I. tiliifolia* (Desr.) Roem. & Schult.), whereas those of inland species (*I. alba*, *I. batatas* or *I. indica*, among others) sometimes float and sometimes do not. However, as Guppy and others admit, this pattern is not absolutely clear, especially in Convolvulaceae; some inland species are also frequent in more coastal environments and their seeds can sometimes float:

> *«This behaviour of the Convolvulaceae becomes yet more intelligible, and more in accordance with the principle, when we reflect that the cause of buoyancy is not concerned with the seed-coats or with the nucleus, neither of which are able to float, but with the air-spaces left by the incomplete filling-up of the seed-cavity by the crumbled embryo. The extent to which the seed-cavity is filled up varies not only between different genera and between different species of the same genus, but also amongst individuals of the same species. Even the seeds of* Ipomoea pes-caprae*, amongst the most typical of floating seeds, display this variation, and they show it also in their floating power, since about a third of the seeds usually sink during the first month or two of the flotation experiments. We can thus explain also why in the case of Ipomoea indica seeds from Fiji floated for months, whilst those from Hawaii had no floating power»*
>
> (Guppy, 1906: 20).

In addition, Guppy (1906) was able to germinate seeds of several species of *Ipomoea* after weeks and months of them floating in sea water, which further demonstrates that long-distance dispersal, in this case by sea currents, is a plausible mechanism to explain the disjunct distribution of certain species of *Ipomoea*.

We have identified at least 27 instances of closely related species with highly disjunct distribution patterns in *Ipomoea*, that is, a taxon or group of taxa nested in a clade where all other taxa inhabit a different geographical region (Figure 3.22). These cases are

most easily explained as the result of long-distance dispersal by natural means. In some cases, species with a distribution explained by long-distance dispersal have subsequently speciated (for example number 1 in Figure 3.22).

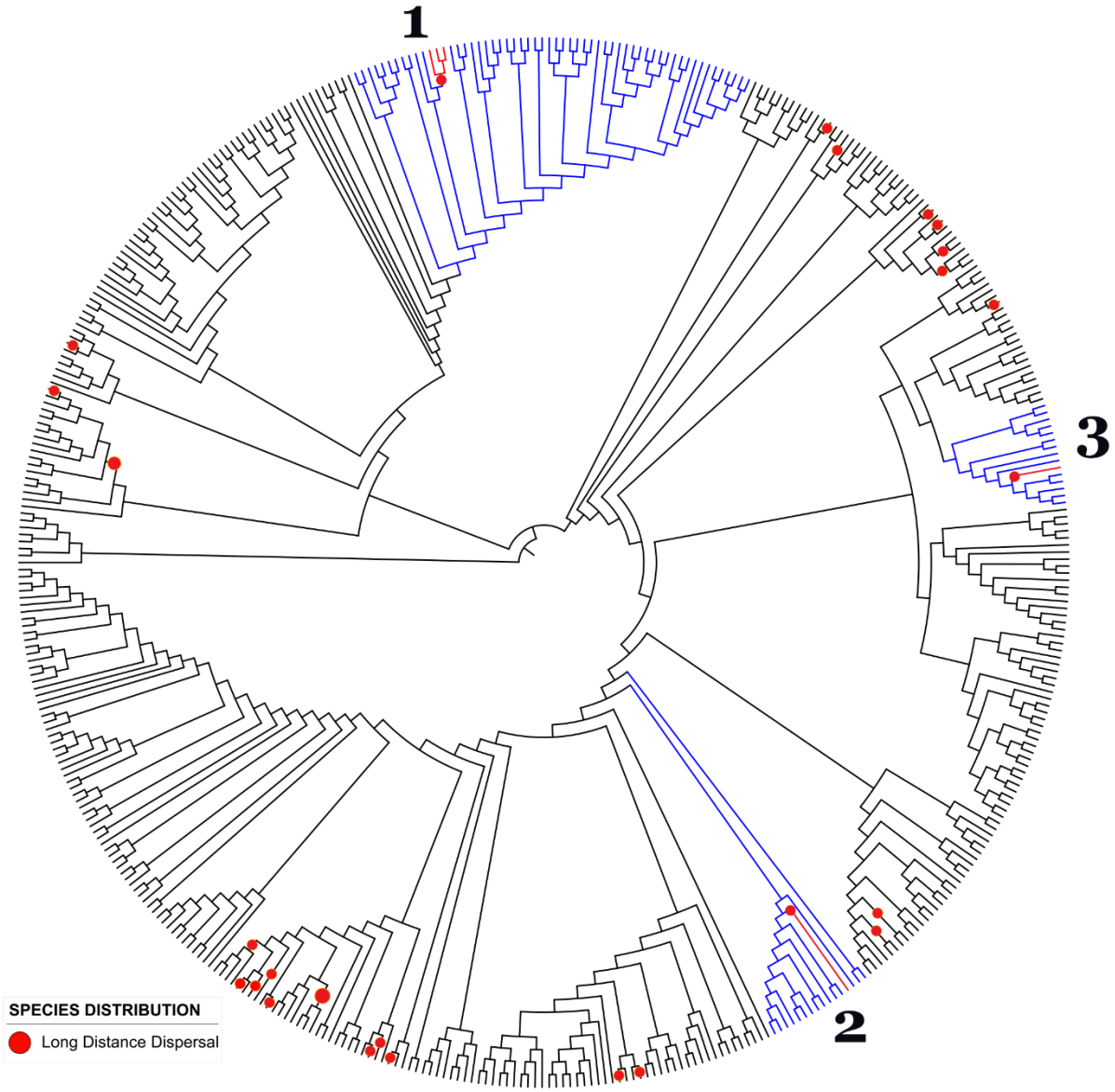


Figure 3.22. Summary phylogeny of *Ipomoea* inferred from *ITS* sequences using one sample per species. Long-distance dispersal events in *Ipomoea* are scattered across the phylogeny of the genus. Red dots indicate events of long-distance dispersal in a phylogeny of *Ipomoea* inferred using *ITS* sequences. Numbers 1, 2 and 3 indicate three examples of trans-oceanic dispersal discussed in the text. Red branches indicate the species with a geographical distribution different from their closest relatives (in blue): 1, divergence occurred after dispersal; 2, *Ipomoea littoralis* is the only sweet potato CWR in the Old World and absent from the Americas; 3, *I. tuboides* is an endemic Hawaiian species nested within a clade of Mexican taxa.

Our results show that long-distance dispersal by natural means is commonplace in *Ipomoea* and thus its potential relevance in the distribution of species with a widespread

distribution must be considered. Especially relevant for the discussion of sweet potato presence in Polynesia are two cases of trans-oceanic long distance dispersal, *Ipomoea littoralis*, a sweet potato CWR, and *I. tuboides* O.Deg. & Ooststr. ***Ipomoea littoralis*** is the only sweet potato CWR restricted to the Old World and absent from the Americas (number 2 in Figure 3.22), whereas ***I. tuboides*** is a Hawaiian endemic nested in a clade of Mexican species (number 3 in Figure 3.22). Both these species are present in the Pacific islands but absent from the American continent, where all other members of their groups inhabit. I discuss these distribution patterns in more detail in Chapter 5.

3.3 | CONCLUSIONS

In this chapter, I have presented the most comprehensive phylogenetic study on the genus *Ipomoea* and the tribe *Ipomoeeae* to date. Using phylogenies inferred from 605 single copy nuclear coding regions and from the whole chloroplast region, in combination with a densely-sampled *ITS* phylogeny, we identified several patterns that characterise this group of plants.

First, the genus *Ipomoea* as traditionally considered is polyphyletic, with all other genera with spiny pollen (tribe *Ipomoeeae*) nested inside. Furthermore, among the smaller genera only *Astripomoea* is monophyletic, with all other genera polyphyletic. For these situations there are generally two solutions: to recognise an expanded *Ipomoea* that includes all other genera that are nested within or, alternatively, split *Ipomoea* into a series of smaller, more manageable, monophyletic groups. Our research clearly demonstrates that the latter solution of splitting *Ipomoea* into smaller clades is unrealistic, as many parts of the *Ipomoea* phylogeny have few or no diagnostic characters. In our view, it is most appropriate to sink all these genera into a broader, monophyletic *Ipomoea*. This would make the tribe *Ipomoeeae* monogeneric, and the spiny pollen a synapomorphy of the group within Convolvulaceae. Our phylogenies also show that, except for *Ipomoea* Ser.

Batatas, all infra-generic ranks described in *Ipomoea* are polyphyletic and the species traditionally assigned to them occur in different parts of the tree. The use of these ranks to identify groups within *Ipomoea* should be discontinued.

Secondly, I have presented a fully-resolved phylogeny of *Ipomoea* in which two main clades can be identified. One clade comprises around one-third of the species, mostly from the Old World, and the other one comprises the other two-thirds, mostly species from the New World. Between these two clades, a series of smaller clades of African taxa form an evolutionary grade. Our results suggest that the genus *Ipomoea* originated and first diversified in the Old World, subsequently dispersing to the New World. Two species-rich clades (A and B) are restricted to the New World.

In addition, we identified several species-rich sub-clades within the NW clade— some of them with a narrow geographic distribution— that can be identified morphologically. Among these, we have identified the group of fourteen species that share a recent evolutionary history with the sweet potato: its crop wild relatives. These wild species are a potential source of genetic diversity for crop improvement, and thus their identification is the first step for prospective studies on sweet potato improvement that could use these wild plants as reservoirs of genetic diversity.

Finally, our densely sampled phylogenies reveal interesting patterns in the evolution of *Ipomoea*. I have shown that at least 48 species of *Ipomoea* develop enlarged storage roots as in the sweet potato that are often edible. I have demonstrated that species with storage roots, which appear in all continents, are scattered across the phylogeny of the genus. The most likely explanation for this pattern is that storage roots have evolved multiple times independently.

Also, we have identified multiple instances of long-distance dispersal by natural means in *Ipomoea*. This shows that diffusion to distant places by natural means is not

uncommon in the genus and, therefore, must be taken into account when trying to explain distribution and biogeographical patterns in the genus.

4 | PHYLOGENETIC STUDIES OF THE SWEET POTATO AND ITS WILD RELATIVES

4.1 | INTRODUCTION

Ipomoea batatas (L.) Lam., the sweet potato, is the most well-known and economically important member of the genus *Ipomoea*. Cultivated in all tropical and subtropical regions of the Globe (The International Potato Center, 2006) (Figure 4.1), it is among the ten most consumed crops worldwide (Food and Agriculture Organization of the United Nations, 2018) and a staple in over twenty developing countries (Bovell-Benjamin, 2007; The International Potato Center, 2006). In addition, its orange-fleshed varieties are rich in β-carotene, a major source of vitamin A for most people (Burri, 1997), and their consumption helps to address deficiencies related to this important nutrient affecting millions of children worldwide (Jones and de Brauw, 2015; Kurabachew, 2015; Baafi et al., 2016).

Despite its importance as a staple and after decades, if not centuries[27], of studies, many aspects of sweet potato evolution remain poorly understood. It is not known, for example, whether the sweet potato had a single or a multiple origin. In addition, and taking into account that *Ipomoea batatas* is a hexaploid species (with all its wild relatives either diploid or tetraploid), it is not known whether it originated by direct autopolyploidization from its wild ancestor or it is the result of hybridisation between different species. Furthermore, it is not known which wild species is/are most closely related to the sweet potato nor which of them is sweet potato's closest relative. The lack of knowledge about the species closely related to the sweet potato has hindered their utilisation in breeding programmes to improve the quality of the crop (Martin and Jones, 1972; Ting and Kehr, 1953).

In this chapter, I first review the current knowledge about the systematics of sweet potato and the group of species closely related to it. I then present and discuss the results

[27] Studies started with Choisy (1838, 1845) and Candolle (1883), among others.

of our genomic analyses, which provide answers to several questions pertaining to the origin and evolution of the crop that have never been satisfactorily answered. Our results resolve the relationship between the sweet potato and its close relatives, identify the wild species that is most closely related to the sweet potato and provide new insights into the evolution of the crop.

Figure 4.1. Sweet potato field at CIP experimental field station in Ica, Peru.

4.1.1 | Taxonomy of the sweet potato and related species

Linnaeus originally described the sweet potato as *Convolvulus batatas* L. in 1753, in the first volume of his *Species Plantarum*[28]. In those pages, Linnaeus described two other species that we now know are closely related to the sweet potato: *Ipomoea lacunosa*[29] and *I. triloba*[30], although he did not indicate any relationship between them. Four

[28] Linnaeus (1753), p. 154.
[29] Ibid, p. 161.
[30] Ibid.

decades later, Lamarck (1793) transferred some of Linnaeus' *Convolvulus* to the genus *Ipomoea*, including *C. batatas* as *Ipomoea batatas*. Still, he did not discuss any relationship between species.

It took another forty years to see the first formal attempt to delimit a group of species closely related to the sweet potato. Choisy, in 1833, described the genus *Batatas* based on Rumphius' (1747) pre-Linnaean name "Batata" and transferred Lamarck's *Ipomoea batatas* to *Batatas edulis* (Thunb.) Choisy, as well as including three other species in the newly described genus. Curiously, Choisy's first definition of *Batatas* as "plants with four-loculed ovaries" actually excludes *Ipomoea batatas*, which only has two locules (Austin, 1978). Choisy corrected this in his revision of the family Convolvulaceae for De Candolle's *Prodromus* and included up to twelve species in *Batatas*[31] (Choisy, 1845), although only three of them are currently included in the group (Table 4.1).

In subsequent years, some authors accepted the group at the rank of genus (Candolle, 1855; Darlington, 1847; Miquel, 1856; Sagra, 1850) or subgenus (Clarke, 1885; Roberty, 1952), but most authors considered it a section or series within *Ipomoea*, especially after Grisebach's (1864) review of *Ipomoea* for the *Flora of the British West Indian Islands*. Grisebach included eight species in *Ipomoea* Sect. *Batatas* (Choisy) Griseb., but only *Ipomoea batatas* and *I. fastigiata* Sweet (=*I. tiliacea* (Willd.) Choisy) are currently included in the section (Table 4.1). House (1908) divided the section *Batatas* into eleven subsections and included *Ipomoea batatas* in the subsection *Aequisepalae* with twenty-six other species, only six of them currently ascribed to the sweet potato group (Table 4.1).

[31] Only two of these species, apart from the sweet potato, are currently considered closely related to the sweet potato: *B. littoralis* (L.) Choisy (= *Ipomoea littoralis* Blume) and *B. triloba* (L.) Choisy (=*I. triloba* L.).

As currently accepted, Section *Batatas* was circumscribed by van Ooststroom (1953) in his revision of the family Convolvulaceae for *Flora Malesiana*. Importantly, van Ooststroom provided a detailed description of the sepals, the most useful morphological trait to assign a species to the group:

> «[…] *subcoriaceous sepals, often oblong or lanceolate, acute, with ciliate margins, often glabrous, often attenuate from a stiff pale base into a herbaceous green, recurved acumen, rarely obtuse or entirely glabrous.*»
>
> (Van Ooststroom, 1953: 468)

Two other species, *Ipomoea lacunosa* and *I. trichocarpa* Ell. were added to *Ipomoea* section *Batatas* in the revision of the group by Martin and Jones (1972) (Table 4.1). In that revision, they also cite *I. gracilis* R.Br. as closely related to the sweet potato. This was an error frequently repeated in subsequent works because *I. gracilis* was later shown to be a misapplied name for *I. littoralis* (Austin, 1991) (see footnotes 23 and 26).

The last global revision of section *Batatas* was carried out by Austin (1978, 1988a, 1988b) and most subsequent works are based on Austin's descriptions and identification keys. Austin included 13 species in *Ipomoea* series *Batatas*, two of them "natural hybrids" (*I.* × *leucantha* Jacq. and *I.* × *grandifolia* (Dammer) O'Donell), most of which are currently accepted as belonging to the section (Table 4.1).

In 1990, Austin and McDonald described one new species and one variety in the group *Batatas*: *Ipomoea tabascana* McDonald & Austin from Tabasco, Mexico, and *I. batatas* var. *apiculata* (Martens & Galeotti) McDonald & Austin from Veracruz, also in Mexico. Although no new collections of these entities have been reported since then, their taxonomic status has not been challenged. McDonald and Austin (1990) also cited *I. umbraticola* House, a species from Mexico and Central America, as a member of Section *Batatas*. This name was later placed as a synonym of *I. splendor-sylvae* House (Austin et al., 2012), but its position within *Batatas* remained unchanged.

A new entity belonging to the Batatas group was mentioned by Duncan and Rausher (2013b) in a study of several North American species. They provisionally named it *Ipomoea austinii* but never described the species formally[32].

Finally, Wood and colleagues (2015b) assigned the newly described Bolivian endemic *Ipomoea lactifera* to the sweet potato group.

As currently accepted and supported by our molecular studies (see previous chapter), the sweet potato group, formally *Ipomoea* series *Batatas*, consists of 15 accepted species (Table 4.1). The delimitation of these species and the relationship between them is discussed in the results section of this chapter.

Table 4.1. List of species currently recognized in the *Batatas* group, arranged by year of publication (in brackets, year of first mention as closely related to the sweet potato if different from publication date).

1753 (1908)	*Ipomoea lacunosa* L.
1753 (1845)	*Ipomoea triloba* L.
1788 (1988)	*Ipomoea leucantha* Jacq.
1797 (1845)	*Ipomoea batatas* (L.) Lam.
1810 (1988)	*Ipomoea cordatotriloba* Dennst.
1825 (1845)	*Ipomoea littoralis* Blume
1838 (1908)	*Ipomoea trifida* (Kunth) G.Don
1845 (1978)	*Ipomoea ramosissima* (Poir.) Choisy
1845 (1978)	*Ipomoea tenuissima* Choisy
1845 (1864)	*Ipomoea tiliacea* (Willd.) Choisy (= *I. fastigiata* Sweet)
1869 (1978)	*Ipomoea cynanchifolia* Meisn.
1907 (1908)	*Ipomoea splendor-sylvae* House (= *I. umbraticola* House)
1952 (1978)	*Ipomoea grandifolia* (Dammer) O'Donell
1990	*Ipomoea tabascana* J.A. McDonald & D.F. Austin
2013	*Ipomoea austinii**
2015	*Ipomoea lactifera* J. R. I. Wood & R.W. Scotland

* This species has not been formally described.

[32] The name *Ipomoea austiniii* was used one year later by Infante-Betancour (2014) to name a new Colombian species not related to the sweet potato. Although this new species was soon found to be a synonym of another species (Wood and Scotland, 2017c), the damage was already done and, according to the principles that guide the nomenclature of plants, *I. austinii* is no longer a valid name and a new one is needed for the entity identified by Duncan and Rausher.

4.1.2 | Morphological diagnosis of the group

A complete set of diagnostic characters for the species in the Batatas group, compiled from the works of Choisy (1833, 1845), Grisebach's (1864), Van Ooststroom (1953), Austin (1978) and Bohac *et al*. (1993), is as follows: prostrate or twinning herbs to subshrubs; flowers in peduncled umbellate cymes; calyces with sepals membranaceous to subcoriaceous, often oblong or lanceolate, apex acute to obtuse, margin ciliate or glabrous; corolla campanulate with 5 sepals: stamens included; one style with a capitate bilobate stigma; ovary 2–4 loculate, hirsute; seeds glabrous to slightly pubescent. It is relatively easy to recognise a specimen as a member of this group.

In contrast, morphological differentiation between species is difficult. Species identification relies mainly on the size and shape of the sepals of the flowers (Austin, 1978) (Figures 4.2 and 4.3). Other morphological characters, and specially leaf shape, are not reliable given the enormous variability within species and the frequent overlap (Martin and Jones, 1972; Nishiyama et al., 1961a; Wood et al., 2015b). For these reasons, identification of the species in this group is challenging and misidentification of herbarium specimens is abundant. In line with this, phylogenetic studies relying on pre-existing identifications without further testing are also likely to be affected by specimen misidentification.

Figure 4.2. Four species of *Ipomoea* Series *Batatas*. (A) *Ipomoea batatas*. (B) *Ipomoea trifida*. (C) *Ipomoea triloba*. (D) *Ipomoea grandifolia*.

Figure 4.3. Four species of *Ipomoea* series *Batatas*. (A) *Ipomoea leucantha*. (B) *Ipomoea cordatotriloba* (South American specimen). (C) *Ipomoea cynanchifolia*. (D) *Ipomoea ramosissima*.

4.1.3 | Chromosome numbers in the group

The basic chromosome number in *Ipomoea* is 15 (King and Bamford, 1937; Nishiyama et al., 1961a; Ozias-Akins and Jarret, 1994; Ting and Kehr, 1953; Ting et al., 1957). *Ipomoea batatas* is the only known hexaploid species in the genus, with all other species being diploid or tetraploid; it has been argued that these different ploidy levels

could act as a reproductive barrier in crosses between the sweet potato and its wild relatives (Diaz et al., 1996; Martin, 1965; Orjeda et al., 1991) (see discussion below). Ploidy levels of the species closely related to the sweet potato are presented in Table 4.2.

Table 4.2. Ploidy levels estimated for the species in *Ipomoea* Series *Batatas* (n = 15).

Species	Ploidy	Reference
Ipomoea batatas	4X, 6X	Ozias-Akins and Jarret, 1994
Ipomoea cordatotriloba	2X, 4X	Ozias-Akins and Jarret, 1994
Ipomoea cynanchifolia	2X	Ozias-Akins and Jarret, 1994
Ipomoea grandifolia	2X†	Jarret and Austin, 1993
Ipomoea lactifera	Unknown	–
Ipomoea lacunosa	2X	Jones, 1964; King and Bamford, 1937; Ozias-Akins and Jarret, 1994
Ipomoea leucantha	2X	Ozias-Akins and Jarret, 1994
Ipomoea littoralis	2X	Ozias-Akins and Jarret, 1994
Ipomoea ramosissima	2X	Ozias-Akins and Jarret, 1994
Ipomoea splendor-sylvae	2X	Ozias-Akins and Jarret, 1994
Ipomoea tabascana	4X	Ozias-Akins and Jarret, 1994
Ipomoea tenuissima	2X	Ozias-Akins and Jarret, 1994
Ipomoea tiliacea	2X†, 4X	Jones, 1964; Nishiyama et al., 1961a; Ozias-Akins and Jarret, 1994; Ting and Kehr, 1953
Ipomoea trifida	2X, 4X, 6X†	Jones, 1964, 1968; Ozias-Akins and Jarret, 1994
Ipomoea triloba	2X	Jones, 1964; Nishiyama et al., 1961a; Ozias-Akins and Jarret, 1994

† Experimental confirmation needed.

4.1.4 | Geographical distribution of species

Most species in Batatas have extensive distribution areas that frequently overlap with other species in the group. This, together with the vague delimitation and identification of species boundaries, mean that it is difficult to draw accurate maps of species in the group.

Figure 4.4 shows a recent attempt by Colin Khoury and collaborators (2015) to infer species distribution maps through modelling, showing that species with overlapping distribution patterns are prevalent in the group. This approach is useful to delimit the distribution area of the entire group globally. However, the maps rely upon data from GBIF and other online databases, which in many cases have not been properly curated and likely

contain a considerable number of misidentifications (Goodwin et al., 2015). For that reason, the individual species distributions presented in Khoury's paper are likely to be inaccurate and should be treated with caution.

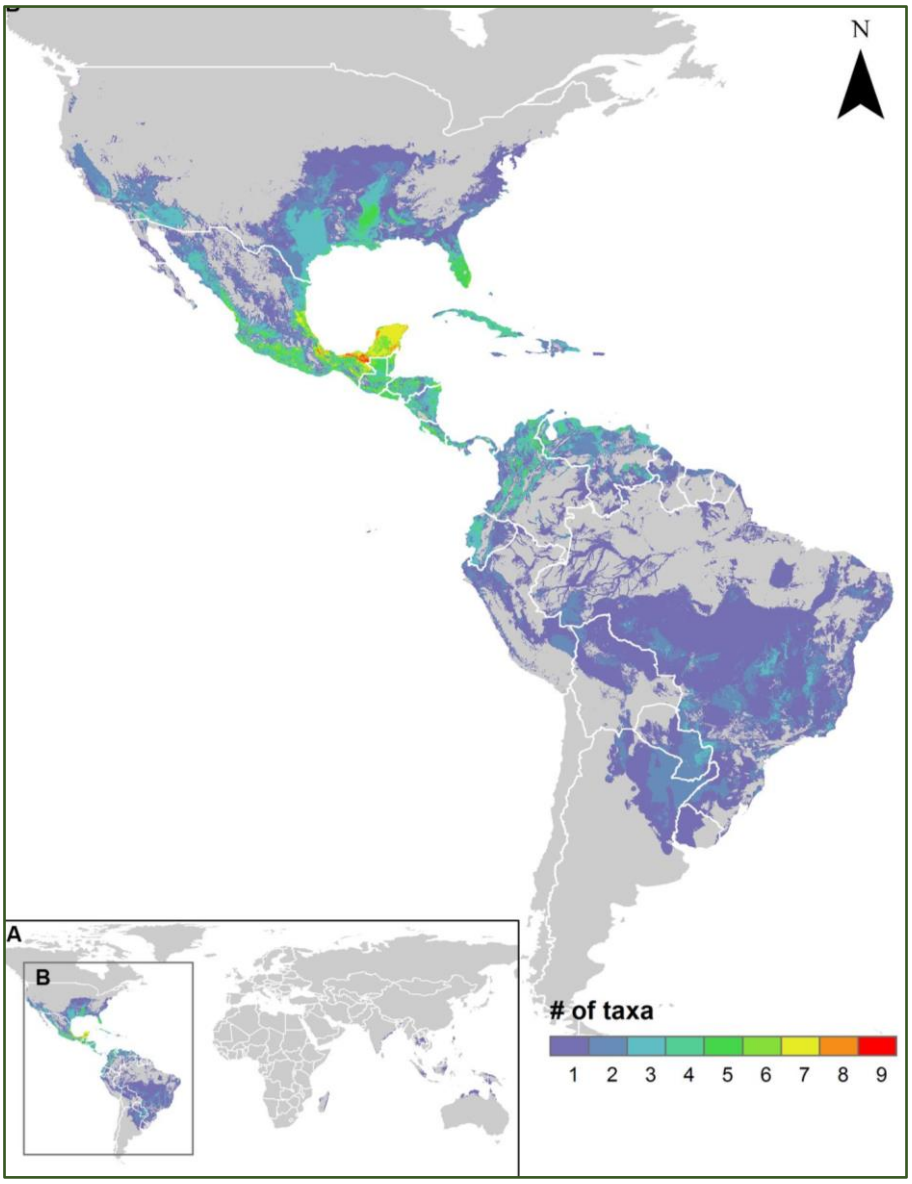


Figure 4.4. Distribution of the species in the Batatas group inferred through modelling. The Circum-Caribbean region is the richest region in number of species of the group. Figure reproduced from Khoury et al. (2015).

In their natural range, all species in the group are restricted to the Americas except *Ipomoea littoralis*, a coastal species widespread from Polynesia to Madagascar but absent from the American continent (Austin, 1991). In America, the group ranges from southern United States (~ 34ºN) to northern Argentina (~ -27ºS) and from sea level to around 3,500

meters a.s.l. I provide approximate distribution areas of several species, based on the specimens that we have studied, in the results section of this chapter.

To date, only *Ipomoea triloba* is confirmed as an introduced species in other regions —sweet potato aside (Holm, 1997). Online databases and some manuals include records of *I. tiliacea* and *I. trifida* in the Old World, but I was unable to confirm or corroborate those records from other sources.

4.1.5 | Evolutionary relationships between species

Studies aiming to clarify the relationship between species in the Batatas group and the origin of the sweet potato have been abundant, although unsuccessful, in the last decades. Both the relationship between species and their role in the origin of the sweet potato remain largely unresolved: almost all species in the group have been proposed as progenitor of the crop, but all these claims lack evidence.

At the beginning of the 20th century, House (1908) suggested that *Ipomoea batatas* had originated by cultivation from *I. tiliacea*, which he described as "doubtfully distinct" morphologically. This hypothesis was *repudiated* by Ting and collaborators (1953; 1957) due to the different ploidy level of the two species (*I. batatas* is hexaploid and *I. tiliacea* tetraploid[33]) and the lack of viable hybrids between them. Ting and collaborators further suggested that, based on a different behaviour of chromosomes in sweet potato meiotic studies, the sweet potato must be of hybrid origin (allopolyploid); although they provided no names, they considered that a tetraploid and a diploid species had to be involved in the origin of the crop.

[33] Nishiyama's (Nishiyama et al., 1961a) observations indicate that *Ipomoea tiliacea* may include diploid and tetraploid specimens, but this point has not been investigated later.

In a series of papers published in the 1960-70s, Nishiyama and collaborators claimed they had found a wild hexaploid plant in Mexico (K-123) that represented the progenitor of the cultivated sweet potato (Jones, 1967; Nishiyama, 1959, 1971; Nishiyama and Teramura, 1962; Nishiyama et al., 1961a, 1961b, 1975)[34]. According to them, hybrids between this hexaploid ancestor (identified as *I. trifida*) and the sweet potato were fertile and produced well-developed storage roots. Furthermore, building on the hypothesis of a hybrid origin of the sweet potato, Nishiyama suggested that diploid *Ipomoea leucantha* and tetraploid *I. littoralis* were the progenitors of the sweet potato. At the same time and making use of a certainly confusing narrative, he also argued that *I. batatas*, *I. trifida*, *I. tiliacea* and *I. littoralis* are all autopolyploids in a series derived from diploid *I. leucantha* (Nishiyama, 1971; Nishiyama et al., 1975). To top it all, this species —*I. leucantha*— has also been suggested to be a hybrid of *I. lacunosa* and *I. trichocarpa*[35] (Austin, 1978, 1988b; Jarret et al., 1992) and in our phylogenies it is clearly polyphyletic (see Results). In summary, Nishiyama's theory is likely to be wrong and, if *Ipomoea batatas* is a hybrid, it remains unknown what species are involved.

Nevertheless, Nishiyama expanded his approach by generating (and naming) two artificial hybrids between different species in the group —*I. littocantha* (*I. leucantha* X *I. littoralis*) and *I. lacunocilis* (*I. lacunosa* X *I. gracilis*) —, aiming to *demonstrate* that the sweet potato was the result of a hybridisation. He also attempted to reclassify *I. littoralis* and *I. leucantha* as varieties of *I. batatas* (Nishiyama, 1971), but subsequent authors did not take his nomenclatural proposal into account.

In contrast to Nishiyama, other contemporary authors had different opinions. Jones (1970) for example thought that *Ipomoea gracilis* had a role in the origin of the sweet

[34] I was unable to find several other papers published around those years by Nishiyama's collaborator T. Teramura in local Japanese journals, although I suspect they follow on the hypothesis built by Nishiyama.

[35] This species is no longer accepted and the name has been cited as a synonym of several others.

potato, whereas Heiser Jr. (1965) and Martin and colleagues (1974) accepted Nishiyama's theory but thought *Ipomoea trifida* was in reality a feral form of sweet potato escaped from cultivation. On the other hand, Kobayashi (1984) not only accepted *I. trifida* as a good species but also described a so-called "*Ipomoea trifida* complex" that would include all wild species in the *Batatas* group that can cross with the cultivated plant.

Finally, other authors studied the hexaploid Mexican plant (K-123) on which Nishiyama's hypothesis was largely based and agreed that K-123 was a feral form of the sweet potato (Austin, 1977; Jones, 1967; Martin and Jones, 1972; Yen, 1971).

A careful reading of the papers presented above clearly shows in part the reasons for so much confusion: authors did not have a good knowledge of the species discussed and constantly misidentified the specimens under study, even contradicting themselves in subsequent publications. The ideas of Nishiyama and of subsequent authors were little by little replaced by other hypotheses and suggested origins of the sweet potato, formulated by researchers with a better understanding of the diversity within the Batatas group. What remained, however, was the idea that the species more closely related to the sweet potato would be able to hybridise with it. As a consequence, the following years saw an increase in the search and study of hybrids, not only involving the sweet potato but also between other species in the group (Abel and Austin, 1981; Austin, 1978).

Austin (1977) suggested that the tetraploid plants collected by Nishiyama and others (e.g. K-134, K-222, K-233) in Mexico and elsewhere in the Americas were modern hybrids between *Ipomoea batatas* and *I. trifida*. Viable hybrids between *I. batatas* and *I. trifida* were produced in different experiments in the 1990s (Oracion et al., 1990; Orjeda

et al., 1991), demonstrating that the isolation due to different ploidy levels is not complete and thus a degree of introgression between *I. batatas* and *I. trifida* is possible[36].

Two years later, Austin and collaborators (Bohac et al., 1993) identified as *Ipomoea batatas* some more Mexican tetraploids maintained at the USDA germplasm collection, contradicting Austin's (Austin, 1977) previous assertion that all tetraploid individuals were hybrids between *I. batatas* and *I. trifida*. Rather, they seem to have found tetraploid *I. batatas* quite easily and even described a Mexican tetraploid as *I. batatas* var. *apiculata* (M. Martens & Galeotti) J.A.McDonald & D.F.Austin. They associated the tetraploid individuals mentioned above (K-134 and K-233) with this new taxon and explained that tetraploid sweet potato was distributed "from Southern Mexico to Western Ecuador and Northern Peru" (Austin et al., 1992), but that entity has never been studied again.

Several studies approached the Batatas group through morphological and karyological research. In 1972, Martin and Jones presented an extensive comparative study of six entities closely related to the sweet potato using 41 morphological plus 3 reproductive traits. They did not clarify the origin of the sweet potato but provided useful information about some species —and less useful about others. They confirmed for example that *I. lacunosa* is restricted in its distribution to the United States and this circumscription is currently accepted, whereas they proposed a broadly defined "*Ipomoea triloba complex*" that is not currently recognised.

[36] It is important to note that most crosses between species attempted in those studies were unsuccessful. For example, Orjeda and collaborators attempted 28,000 crosses between sweet potatoes and *Ipomoea trifida*. In principle each flower must produce between 3 and 4 seeds; however, they only obtained 730 seeds in total. As they explained in the paper, only «*one flower fructified from every 114 flowers pollinated*». They obtained, on average, 2.59 seeds per 100 pollinations. From those, less than 50% seeds germinated, and only 34% seedlings grew properly. In summary, the reproductive success of this cross was notably low. Also, those seedlings had different ploidy levels: 52 individuals were hexaploids, 190 were tetraploids and 5 were pentaploids (and therefore sterile).

Ipomoea triloba is another species that has been suggested as putative progenitor of the crop based on morphological similarities: Austin presented the sweet potato as a hybrid between *I. triloba* and *I. trifida* (Austin, 1988b) and, although evidence from different sources indicated early on that *I. triloba* might not be so closely related to the crop, its possible role as progenitor has been subsequently recorded by later authors as equally plausible with other hypotheses.

Through comparison of chromosome shape and organisation, Srisuwan and collaborators (2006) suggested that *I. triloba* is only distantly related to the sweet potato (and to all other species in their study) and that *I. tabascana* could be a modern hybrid between *I. batatas* and *I. trifida*.

In 1992 and 1993, Jarret and collaborators published two papers on the relationship between species in the Batatas group using, for the first time, DNA data (Jarret and Austin, 1993; Jarret et al., 1992). Specifically, they used restriction fragment length polymorphisms (RFLP) and RAPD markers and suggested that the sweet potato was more closely related to *Ipomoea trifida*, *I. tabascana*, *I. triloba* and some other annual species, with all other species in the group, notably *I. tiliacea* and *I. littoralis*, more distantly related. The authors also argued that *I. grandifolia*, previously presented by Austin (1978) as of hybrid origin, was a distinct species. The results of Jarret and collaborators are based on very small data sets: 21 accessions of 13 species in 1992 and 35 sweet potato clones and 10 specimens of 7 other species in 1993; they did not include all species in the series and, for most species, included only one specimen.

Subsequent analyses using DNA included RAPD markers (Dhillon and Ishiki, 1999); ITS and Waxy sequences (Miller et al., 1999); inter-simple sequence repeats (ISSR) and chloroplast DNA fragments (Huang and Sun, 2000); highly polymorphic AFLP markers and ITS sequences (Huang et al., 2002); nuclear *β-amylase* gene sequences (Rajapakse et

al., 2004); and chloroplast and nuclear microsatellite markers (Roullier et al., 2011, 2013a). These studies included representatives of more species in the series and a better representation of each of them, but still had insufficient data to provide robust conclusions. The resulting phylogenies provided further support for the delimitation of the group and suggested certain species relationships that have been subsequently confirmed in our studies, as well as a close relationship between the sweet potato and *I. trifida* (see Results). However, the small amount of data in those studies did not fully resolve species relationships nor species monophyly and failed to produce results that were strongly supported. Even more, conflicting results have been presented. For example, recent evidence provides support for both a single and a multiple origin of the sweet potato. The analysis of AFLP fragments (Zhang et al., 2004) and the identification of an *Agrobacterium* T-DNA in the nuclear genome of cultivated sweet potato plants but not in the wild relatives (Kyndt et al., 2015) point to a single origin. In contrast, the identification of two sweet potato lineages from chloroplast DNA data has been interpreted as evidence of a multiple origin (Roullier et al., 2011, 2013a, 2013b).

In conclusion, at the beginning of our project and after 100 years of phylogenetic studies, the situation was as follows: 1) as understood, the Batatas group was formed by some 14 species, including 2 hybrids; 2) most species were not well known and needed further investigation; 3) the relationship between the species in the group was unresolved; 4) *I. trifida* and *I. triloba* were presented as the most probable progenitors of the sweet potato and an origin from *I. tiliacea* or *I. littoralis* was no longer accepted, although definitive confirmation for any one theory was missing; 5) the sweet potato was thought to be an allopolyploid; and 6) it was not known whether sweet potato had a single or multiple origin. The results of our studies, presented in the next section, provide answers to all these questions.

4.2 | RESULTS AND DISCUSSION

Ipomoea batatas and fourteen other putative species form a monophyletic group in our phylogenies (**CLADE A3**, hereinafter the Batatas group). This group is always monophyletic with 100% support regardless of the data set and the method of phylogenetic inference (Figures 3.2–3.4). As in previous studies (Miller et al., 1999; Roullier et al., 2011), our *ITS* phylogeny of the Batatas group results in a polytomy in which the relationship between species cannot be assessed (Figure 3.7). In contrast, the nuclear and chloroplast phylogenies inferred using Hyb-Seq data provide phylogenetic resolution to investigate the relationship between species in the Batatas group, and the time-calibrated phylogenies facilitate discussion of the different evolutionary events in a temporal scale. I inferred the phylogenies presented in this chapter using genomic-scale data of 174 herbarium and germplasm collections of sweet potato and its 14 CWRs, which were identified as such during our global study of the genus. All species are represented by multiple specimens except three: *I. lactifera*, recently described and known from very few populations at the time of sampling (Wood et al., 2015b); *I. tabascana*, known from a single population in Mexico; and *I. tenuissima*, a poorly known species with very few herbarium collections. These three species are represented by only one specimen each in the Hyb-Seq phylogenies.

4.2.1 | Phylogeny of the Batatas group and species boundaries

Batatas is a clade that includes fifteen putative species of *Ipomoea*, not all of them monophyletic. We obtained consistent resolution within the group of sweet potato CWRs, irrespective of the kind of analyses (Figures 4.5–4.7).

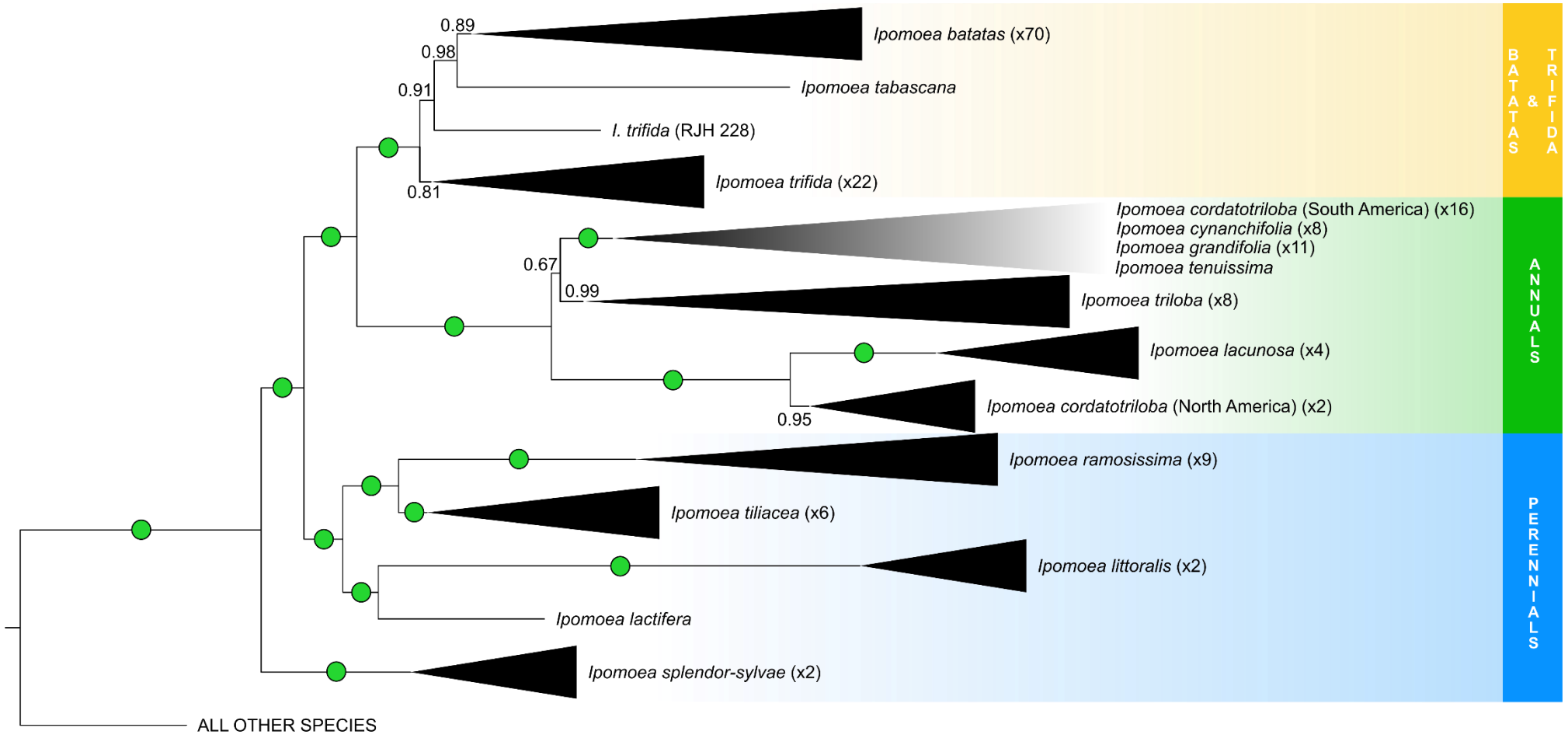


Figure 4.5. Phylogeny of the Batatas group inferred from nuclear coding regions using Coalescence (Astral-II) and Approximate Maximum Likelihood (FastTree). Numbers in the branches indicate Astral-II support values (first number) and Approximate ML support (second number; Shimodaira test, 1,000 replicates). Green dots indicate 100% support in both analyses. Specimens identified as *Ipomoea leucantha* are scattered in various clades and not shown in this phylogeny.

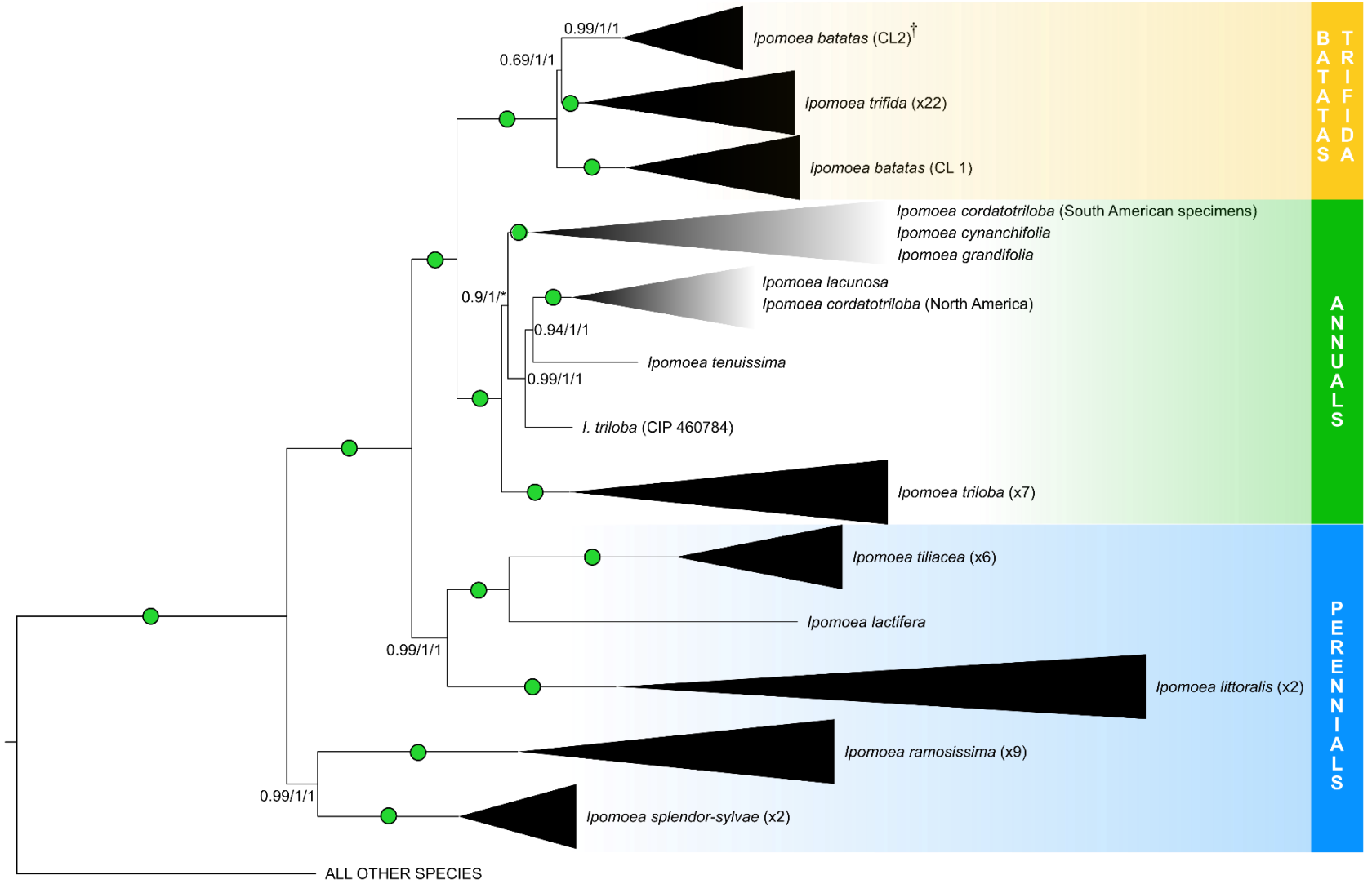


Figure 4.6. Phylogeny of the Batatas group inferred from whole chloroplast genomes using Maximum Likelihood (RAxML), Bayesian Inference (MrBayes) and Parsimony Analysis (PAUP). First number in the branches indicates ML support values (bootstrap support, 1000 replicates), second number Bayesian posterior probability and third number bootstrap support in the Parsimony analysis. Green dots indicate 100% support in all analyses. Specimens identified as *Ipomoea leucantha* are scattered in various clades and not shown in this phylogeny.

† This clade includes *I. tabascana* and tetraploid *I. trifida* (RJH 228). * This node is an unresolved polytomy in the Parsimony topology (Majority Rule).

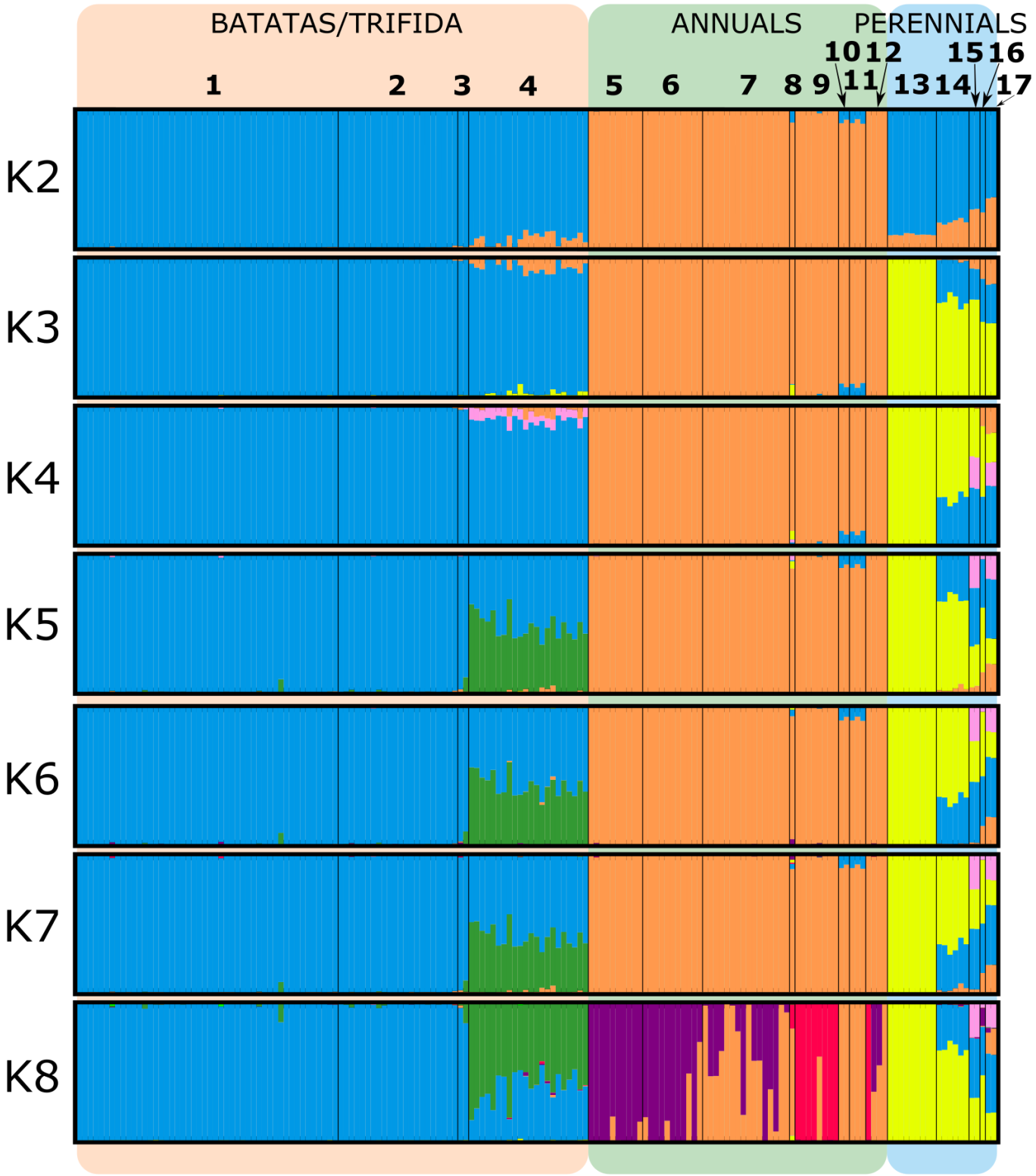


Figure 4.7. Population structure analysis of the species in the Batatas group inferred using 3,000 nuclear variable positions (150,000 MCMC replications and 100,000 burn-in repetitions. Admixture model assuming independent allele frequencies among population [λ = 0.4469; K = 1–5; 3 runs for each K value]. 1, *I.* batatas (chloroplast lineage 2); 2, *I. batatas* (chloroplast lineage 2); 3, *I. tabascana* and tetraploid *I. trifida* (RJH 228); 4, *I. trifida*; 5, *I. cynanchifolia*; 6, *I. grandifolia*; 7, *I. cordatotriloba* (South America); 8, *I. tenuissima*; 9, *I. triloba*; 10, *I. cordatotriloba* (North America); 11, *I. lacunosa*; 12, *I. leucantha*; 13, *I. ramosissima*; 14, *I. tiliacea*; 15, *I. littoralis*; 16, *I. lactifera*; 17, *I. splendor-sylvae*.

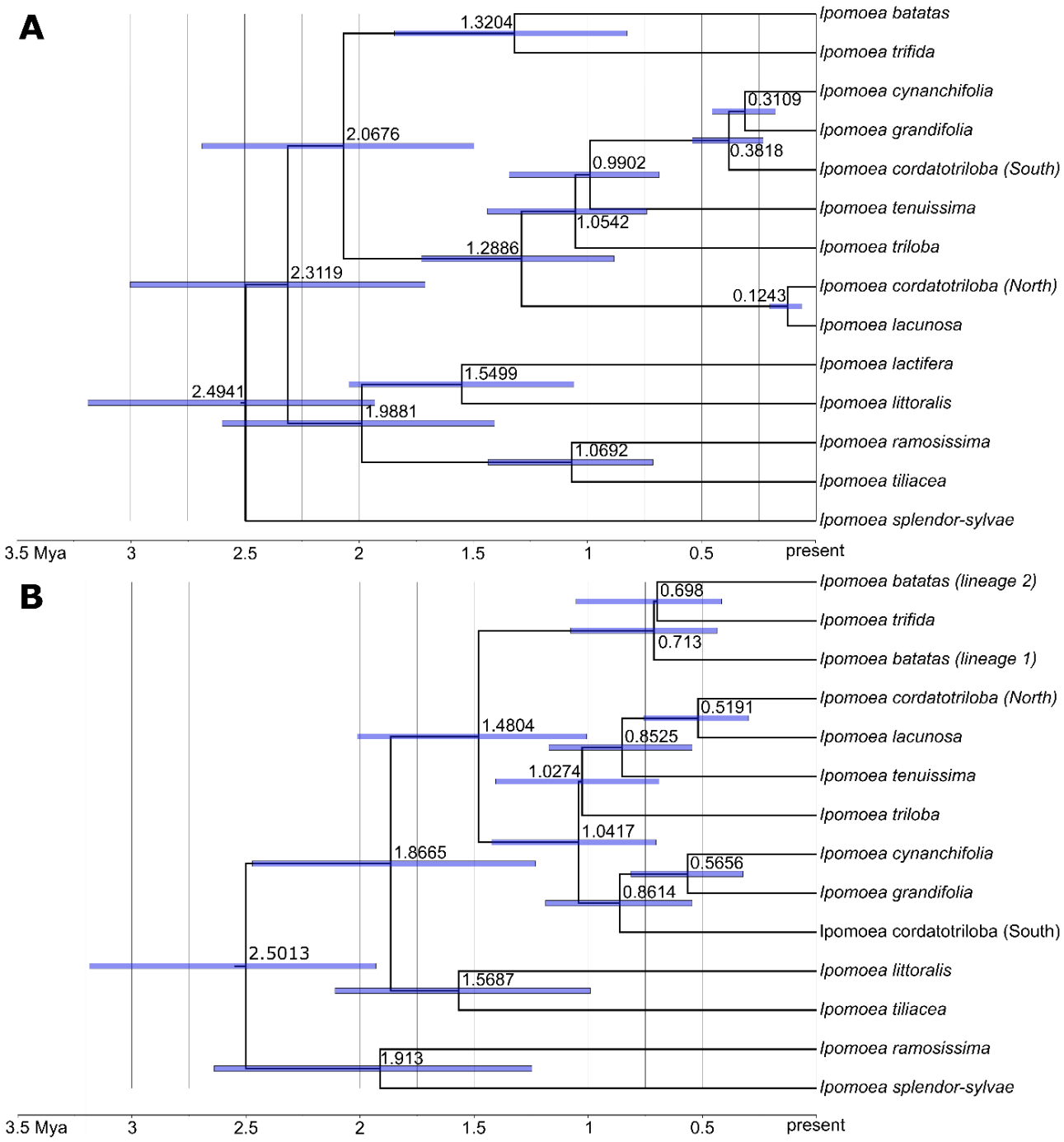


Figure 4.8. Time-calibrated phylogenies of the sweet potato and its CWRs inferred using (A) 21 nuclear regions and (B) whole chloroplasts, respectively. The 95% high posterior density (HPD) for the temporal duration of the branch ancestral to *I. trifida* and *I. batatas* CL2 in (B) is 4–56,000 years. Node bars represent 95% HPD intervals for node ages. The root age for these phylogenies is determined by the ages sampled for the clade in our graphical model constructed in RevBayes. *Ipomoea lactifera* was excluded from the chloroplast phylogeny because its unique structure (15,000 bp shorter than all other specimens with multiple unique indels) led to difficulty in estimating a molecular evolutionary rate

According to the nuclear phylogeny (Figure 4.5), ***Ipomoea splendor-sylvae* House**, a perennial species distributed in Mexico and Central America, is sister to the rest of the group and all other species split in two clades. The chloroplast phylogeny (Figure 4.6) recognises the same two clades but places ***I. ramosissima* (Poir.) Choisy**, another perennial species, together with *I. splendor-sylvae* as sister to the rest. One clade includes the four other perennial species (three in the chloroplast tree), and this clade forms a grade of perennials with *I. splendor-sylvae* (blue in Figures 4.5 and 4.6). The other clade is divided in two sub-clades: one including the annual species in the section (green in Figures 4.5 and 4.6) and the other one formed by *Ipomoea batatas*, *I. trifida* and *I. tabascana* (orange in Figures 4.5 and 4.6).

4.2.1.1 | Perennial species

Apart from *Ipomoea splendor-sylvae*, the group of perennial species includes the Old-World species ***Ipomoea littoralis***; two widespread tropical American species: ***I. tiliacea* (Willd.) Choisy in DC.** and *I. ramosissima*; and one species restricted to Bolivia, ***Ipomoea lactifera***. These species are monophyletic in all our phylogenies with 100% support[37].

4.2.1.1.1 | *Ipomoea littoralis*

Ipomoea littoralis is the only sweet potato CWR restricted to the Palaeotropics (Figure 4.9). It is widespread from Polynesia to the Indian coasts of Asia and Madagascar. It has been frequently confounded with *I. gracilis*, a rare Australian endemic to which it resembles. However, *I. gracilis* belongs to a different clade (CLADE C) and *I. littoralis* is not recorded from Australia, so misidentified specimens are easy to recognise. Interestingly, in the nuclear phylogeny this species is more closely related to *I. lactifera*

[37] As explained before, *Ipomoea lactifera* is represented by a single specimen in our Hyb-Seq phylogenies, but three in the *ITS* phylogeny that form a clade with 100% support.

than to any other species. According to our time-calibrated nuclear phylogeny (Figure 4.8 in page 108), *Ipomoea littoralis* diverged at least 1.5 million years ago, possibly resulting of the dispersal across the Pacific of its common ancestor with *I. lactifera*. I discuss the distribution of *I. littoralis* further in Chapter 5.

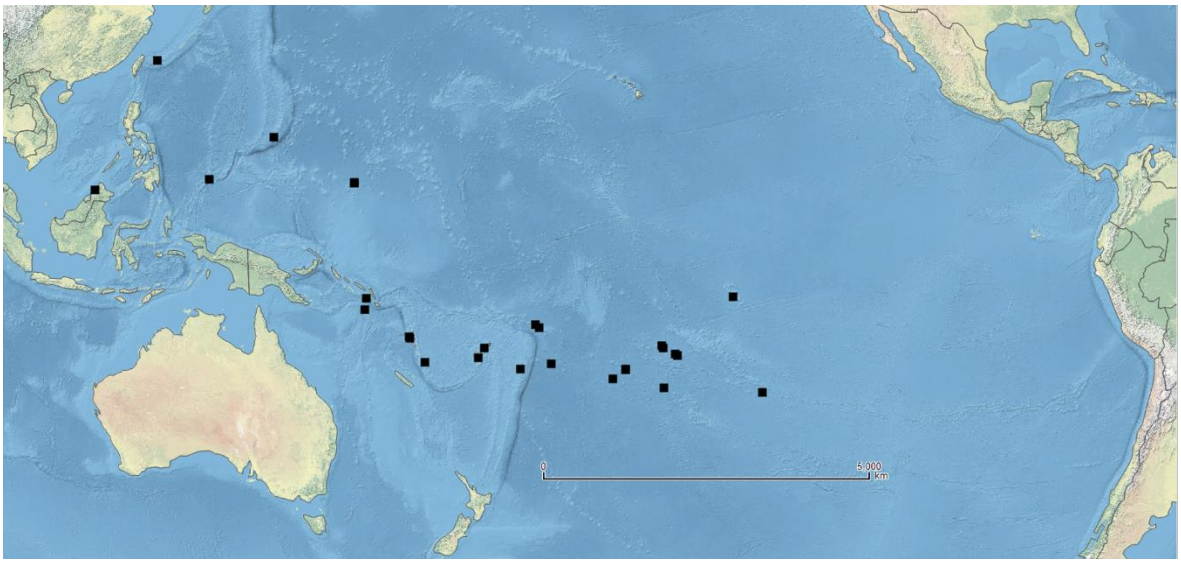


Figure 4.9. Distribution of the *Ipomoea littoralis* specimens included in our study (black squares). Distributed from the Pacific Islands to Asia and the Indian Ocean, this is the only species closely related to the sweet potato absent from the Americas.

4.2.1.1.2 | *Ipomoea ramosissima*

Ipomoea ramosissima, another perennial species, is almost identical morphologically to *I. cynanchifolia*, which belongs to the clade of annuals. These species have overlapping distribution areas and can only be safely distinguished when in fruit: the capsule of *I. ramosissima* is compressed vertically, whereas in *I. cynanchifolia* is not (Figure 4.10).

Also in the case of *Ipomoea ramosissima*, specimens from South America and Central America form two clades (Figure 4.11), which suggests the species is undergoing further diversification driven by geographical patterns (Figure H in Appendix 3).

Figure 4.10. *Ipomoea cynanchifolia* (left) and *I. ramosissima* (right) can only be distinguished by the shape of the fruit, *ovate* in *I. cynanchifolia* and flattened in *I. ramosissima.*

Figure 4.11. Distribution of *Ipomoea ramosissima* specimens included in this study (brown squares). This is a monophyletic species and Central and South American specimens form reciprocally monophyletic subclades.

4.2.1.2 | Annual species

Among the annual species, specimens from North and South America form reciprocally monophyletic groups, which suggests an evolutionary divergence reflecting geography. The only well-defined species in both nuclear and chloroplast phylogenies is ***Ipomoea triloba***. The rest of species previously recognised are not well defined and their specific status is questionable (Figure 4.5 and 4.7 and Figure H in Appendix 3). The monophyly of ***I. tenuissima***, represented by a single specimen, could not be assessed.

4.2.1.2.1 | *Ipomoea triloba*

Ipomoea triloba is a monophyletic species, sister to the group of species restricted to South America. In its natural range, this species has a Circum-Caribbean distribution, although it has naturalised in other tropical and temperate regions of the world. It has become a problematic weed in several countries in Africa, Asia and Oceania (Holm, 1997) and has been recently recorded as naturalised in the Mediterranean Basin (Joel and Listone, 1986; Silvestre Domingo, 2004). Although our sampling size in the Hyb-Seq analyses is small, it is interesting to note that all specimens of *I. triloba* from the Old World, in principle assumed to be introduced, form a clade (Figure H in Appendix 3). This could indicate that specimens outside the natural distribution range of the species come from a single introduction, although our sampling is small and more specimens could confirm or reject this hypothesis.

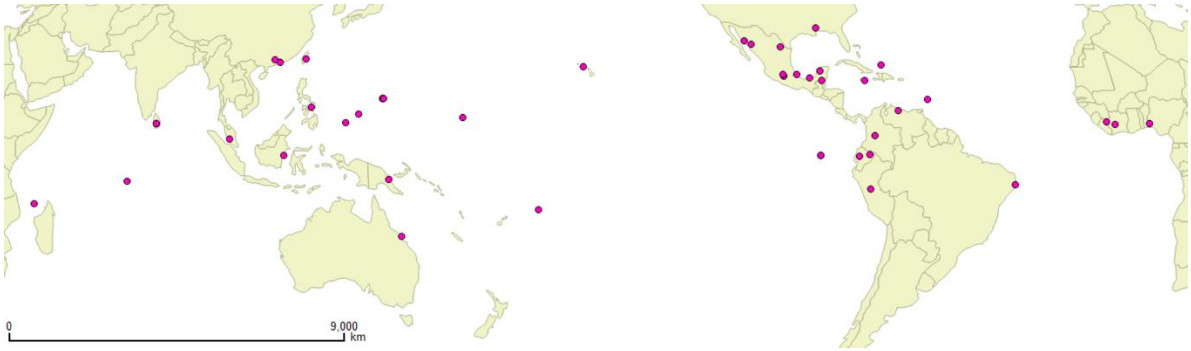


Figure 4.12. Distribution of *Ipomoea triloba* specimens included in this study (pink circles). This Circum-Caribbean species has naturalised in other tropical and temperate regions of the world.

4.2.1.2.2 | *Ipomoea cordatotriloba, I. cynanchifolia, I. grandifolia & I. lacunosa*

***Ipomoea cordatotriloba* Dennst.** has been traditionally recognised as a species with a North-South American amphitropic disjunct distribution (Austin, 1988a; Wood et al., 2015b) (Figure 4.13). However, our results suggest that the entities from both regions represent distinct lineages with different evolutionary histories (Figures 4.5–4.7). Additional studies using *ITS* sequences show that specimens labelled *"I. cordatotriloba*" form a clade of North American species and a grade of South American specimens (Figure 4.14). This provides additional support to the fact that the two entities represent different evolutionary histories and thus should be recognised as two independent entities.

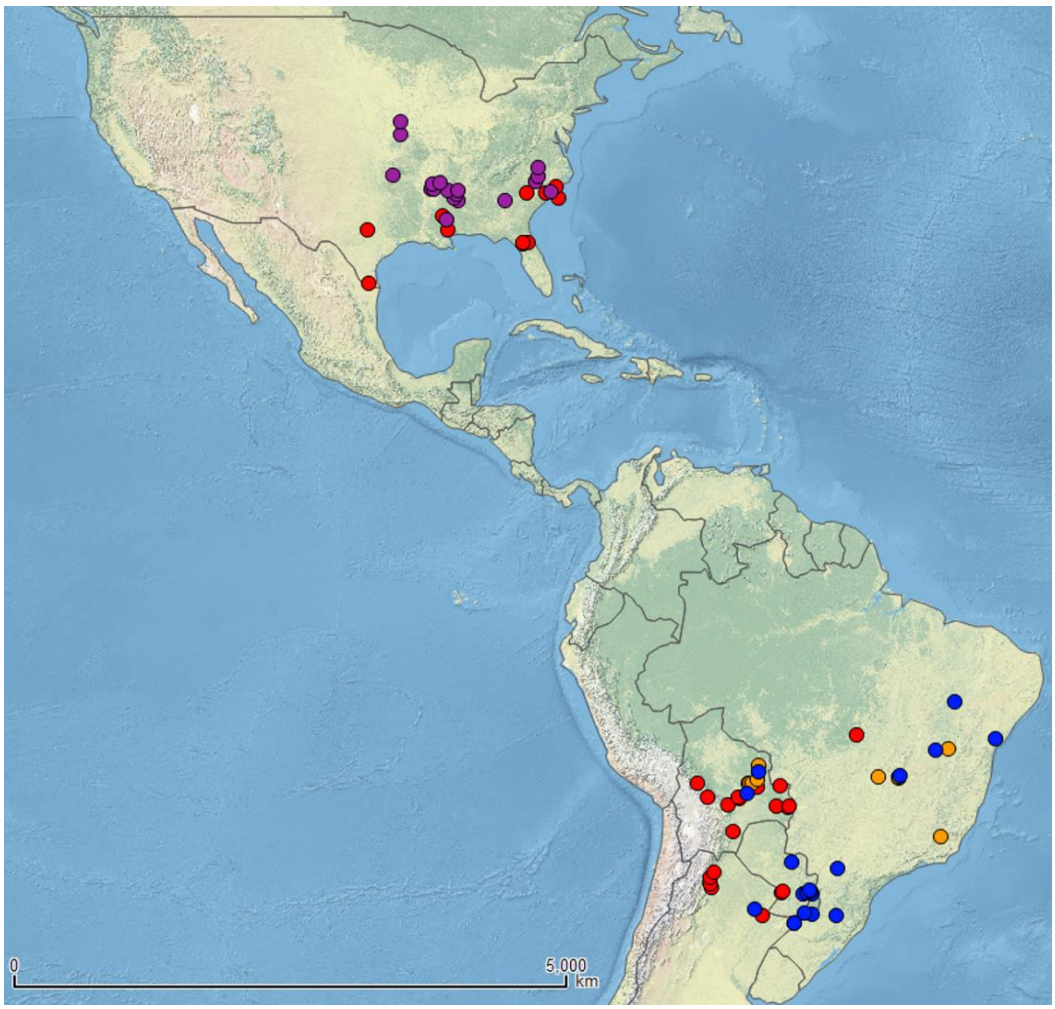


Figure 4.13. Distribution of *Ipomoea cordatotriloba* (red circles), *I. cynanchifolia* (blue), *I. grandifolia* (orange) and *I. lacunosa* (purple) specimens in this study. Our results indicate that *I.* cordatotriloba specimens from North and South America belong to different evolutionary lineages.

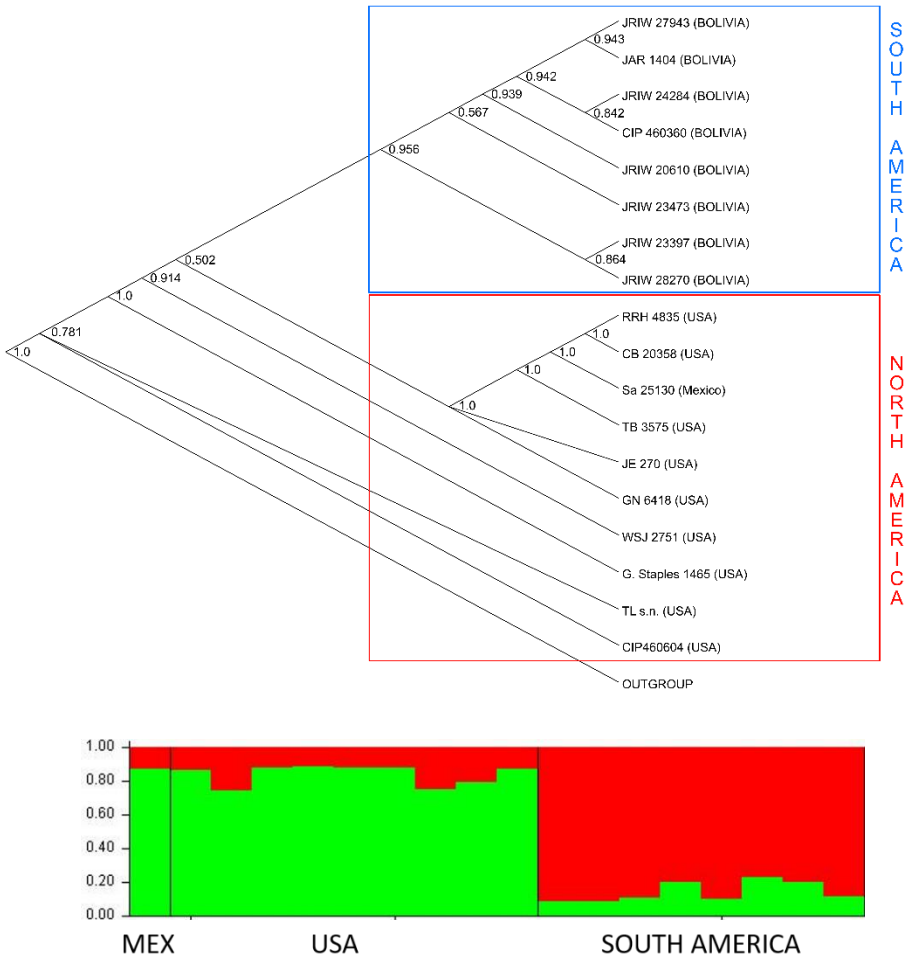


Figure 4.14. Specimens identified as *Ipomoea cordatotriloba* divide in two groups with distinct genetic signatures. A) Maximum Likelihood phylogeny (1,000 bootstrap replicates) inferred from 20 *I. cordatotriloba* samples using *ITS* sequences. B) Population structure analysis of the same 20 samples (16 variable positions; 150,000 MCMC replications and 100,000 burn-in repetitions. Admixture model assuming independent allele frequencies among population [λ = 0.4605; K = 1–4; 3 runs].

The specimens from North America form a clade with ***I. lacunosa***, from which it diverged recently, approximately 250,000 years ago (Figure 4.8). This may explain why the nuclear tree identifies both *I. lacunosa* and *I. cordatotriloba* as monophyletic whereas the tree based on chloroplast genomes, which evolve at slower rates (Drouin et al., 2008), does not.

In contrast, the *I. cordatotriloba* specimens from South America form a clade with *I. cynanchifolia* and *I. grandifolia*. In this clade, most specimens of each species are grouped in the same subclade, but there are several specimens that group with those of other species. This topology, together with the results of the population structure analysis

(Figure 4.7) and the time-calibrated phylogenies (Figure 4.8), could reflect ongoing diversification within this group of taxa, possibly from an ancestral polymorphism. At this point, however, we cannot reject the possibility that the outsiders are simply misidentified specimens. Further studies are needed to clarify the relationships between these species.

4.2.1.2.3 | *Ipomoea leucantha*

Our Hyb-Seq phylogenies include four *Ipomoea leucantha* Jacq. samples. Our analyses resolve *I. leucantha* is polyphyletic, with each sample in different parts of the tree (Figure H in Appendix 3). Specimens from different places in the Americas and quite distinct morphologically have been tentatively named *I. leucantha* by different authors (Figure 4.15). Furthermore, this species was suggested to be of hybrid origin by Austin (1978). Our results show that this name actually includes very different evolutionary entities and thus a global re-evaluation is needed. It could for example be the case that specimens identified as *I. leucantha* were various hybrids of different species.

The current knowledge of this species does not allow any further interpretation of its alleged hybrid origin. Additional taxonomic and phylogenetic studies, including as many specimens as possible recorded under this name, are required.

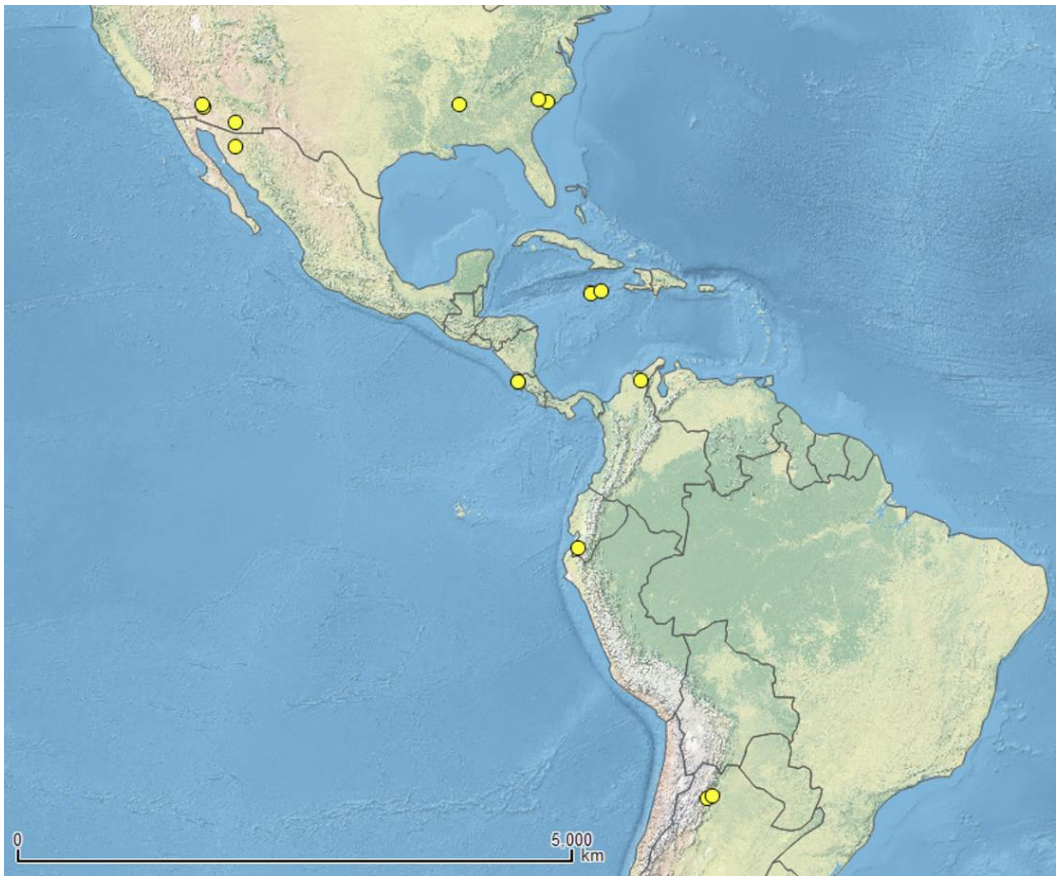


Figure 4.15. Distribution of *Ipomoea leucantha* specimens included in this study (yellow circles). Our results indicate that this species is polyphyletic, with specimens more closely related to different species in the phylogenetic trees.

4.2.1.3 | *Ipomoea tabascana*, a modern hybrid

Ipomoea tabascana was described by Austin and McDonald (1990) from a single location in Tabasco, Mexico. The position of the species in the nuclear phylogeny (Figure 4.5), sister to *I. batatas*; in the chloroplast phylogeny (Figure 4.6), nested within a sweet potato lineage; and the results of the population structure analyses (Figure 4.7) are congruent with this species being a hybrid between *I. batatas* and *I. trifida*, as previously suggested (Srisuwan et al., 2006).

The same intermediate position between *I. batatas* and *I. trifida* is retrieved for another tetraploid specimen identified as *I. trifida* (RJH 228). This specimen was collected in Panama and, according to our results, possibly represents another hybrid between both species (Bohac et al., 1993; Kobayashi, 1984).

4.2.3 | Origin and evolution of *Ipomoea batatas*

Ipomoea batatas is a well-defined monophyletic species and *I. trifida* is its closest relative. Both species diverged from their common ancestor at least 800,000 years ago according to the time-calibrated nuclear phylogeny, or at least 400,000 years ago according to the time-calibrated chloroplast phylogeny (Figure 4.8). The lineage including these two species diverged from the clade of annual species at least 1.3 million years ago (Figure 4.8). Importantly, the analysis using allelic variation shows no evidence that any other species in the group was involved in the origin of the sweet potato (Figure 4.16), which suggests that hexaploid *Ipomoea batatas* originated by autopolyploidization from the common ancestor with *I. trifida*.

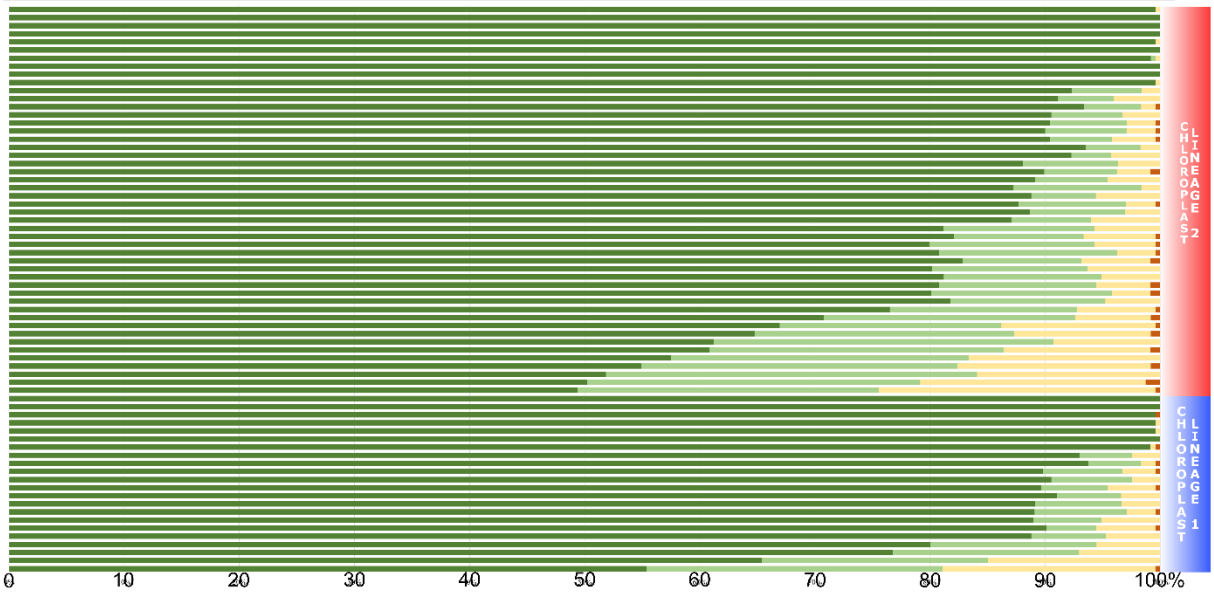


Figure 4.16. Summary of the topology of the sweet potato samples in the gene tree phylogenies inferred considering allelic variation. Each bar represents a sweet potato specimen and its length is the total number of gene trees in which that specimen appears. Colours indicate percentage: dark green = gene trees in which only one haplotype was assembled (no or minimum allelic variation); light green = the six copies form a clade; yellow = at least one allele groups with alleles from other sweet potatoes; red = at least one allele groups with alleles from other species.

In contrast to the nuclear phylogeny, which retrieves a monophyletic *Ipomoea batatas*, the chloroplast phylogeny revealed the existence of two distinct *I. batatas* lineages (Figure 4.6). All the statistical tests and additional analyses that I conducted to challenge this result (see description in section 2.5.2) support the existence of two differentiated *I.*

batatas groups. These two chloroplast lineages were previously identified by Roullier and collaborators, who presented their finding as evidence of a multiple origin of the sweet potato (Roullier et al., 2011). However, their study did not resolve the relationship between the two *I. batatas* lineages and *I. trifida*. Our results show that one of the *I. batatas* lineages (chloroplast lineage 2, hereinafter CL2) is more closely related to *I. trifida* than to the other *I. batatas* lineage (CL1) (Figure 4.17).

To further confirm our results, I visually examined the chloroplast alignment (provided in Supplementary File 10) in detail and discovered that both chloroplast lineages have unique indels and both also share indels with *I. trifida*, with no indels shared exclusively by the two sweet potato chloroplast lineages. This suggests that both *I. batatas* lineages could have inherited their chloroplast haplotypes from *I. trifida* at different times.

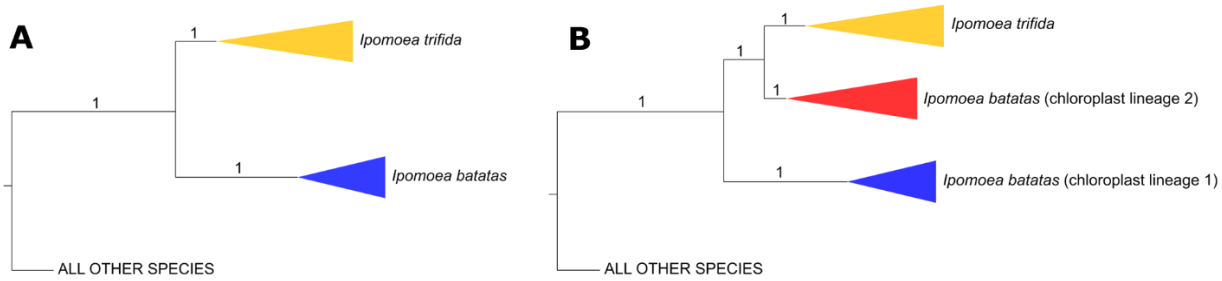


Figure 4.17. Conflicting phylogenies in the origin of *Ipomoea batatas*. The nuclear phylogeny (A) retrieves a monophyletic species, with all specimens descending from a common ancestor, in a clade sister to the wild species *Ipomoea trifida*. In contrast, in the chloroplast phylogeny (B) the *I. batatas* specimens split in two reciprocally monophyletic groups, one of them (chloroplast lineage 2) more closely related to the wild species *Ipomoea trifida* than to the other *I. batatas* lineage (chloroplast lineage 1).

In addition to the chloroplast signature, Roullier et al. (2011) reported weak signature for two *Ipomoea batatas* lineages from non-coding nuclear DNA. For that reason, I also investigated the nuclear coding regions and *ITS* sequences of all *I. batatas* specimens in our study, looking for traces of the two lineages. Although I found high levels of admixture within the crop, similar to those reported for other crops (Cuevas et al., 2017; Sardos et al., 2016), I did not find any evidence of two *I. batatas* lineages in the nuclear genome (Figure 4.18).

The two *I. batatas* chloroplast lineages are dominated by specimens collected from South America (CL1) and Central America (CL2) respectively, but also include some specimens from the other region (Figure 4.19). These results are similar to Roullier's, who interpreted them as evidence of multiple *domestications* of the sweet potato in Central and South America (see Chapter 5 for further discussion on the geographical origin of the crop).

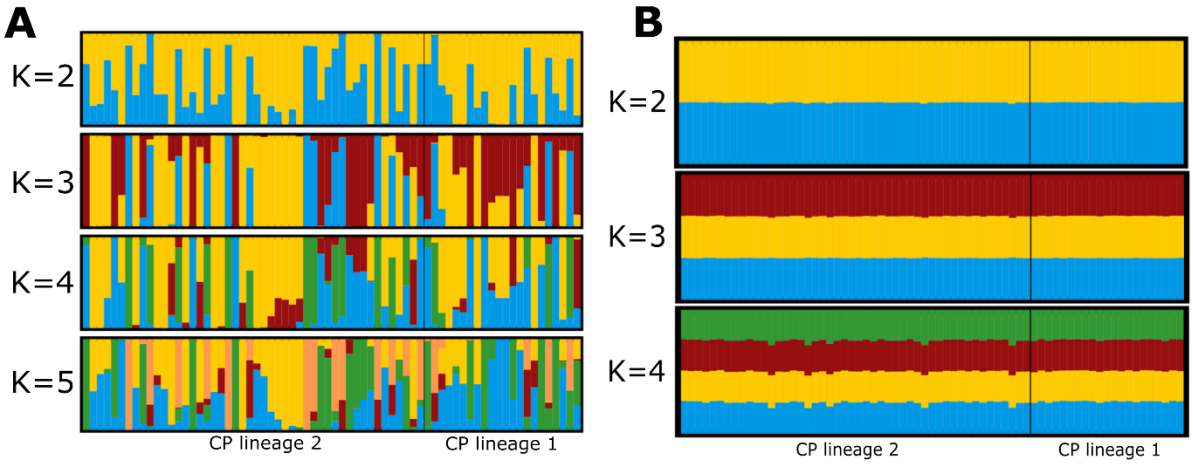


Figure 4.18. Population structure analyses of *Ipomoea batatas* inferred using A) 3,000 variable positions in the alignment of nuclear regions (λ = 0.4469; K = 1–5; 3 runs for each K value) and B) 16 variable positions in the alignment of *ITS* sequences (λ = 0.4605; K = 1–4; 3 runs) failed to identify traces of the existence of two distinct *I. batatas* chloroplast lineages in the nuclear genome. Analyses conducted using 150,000 MCMC replications and 100,000 burn-in repetitions; admixture model assuming independent allele frequencies among populations.

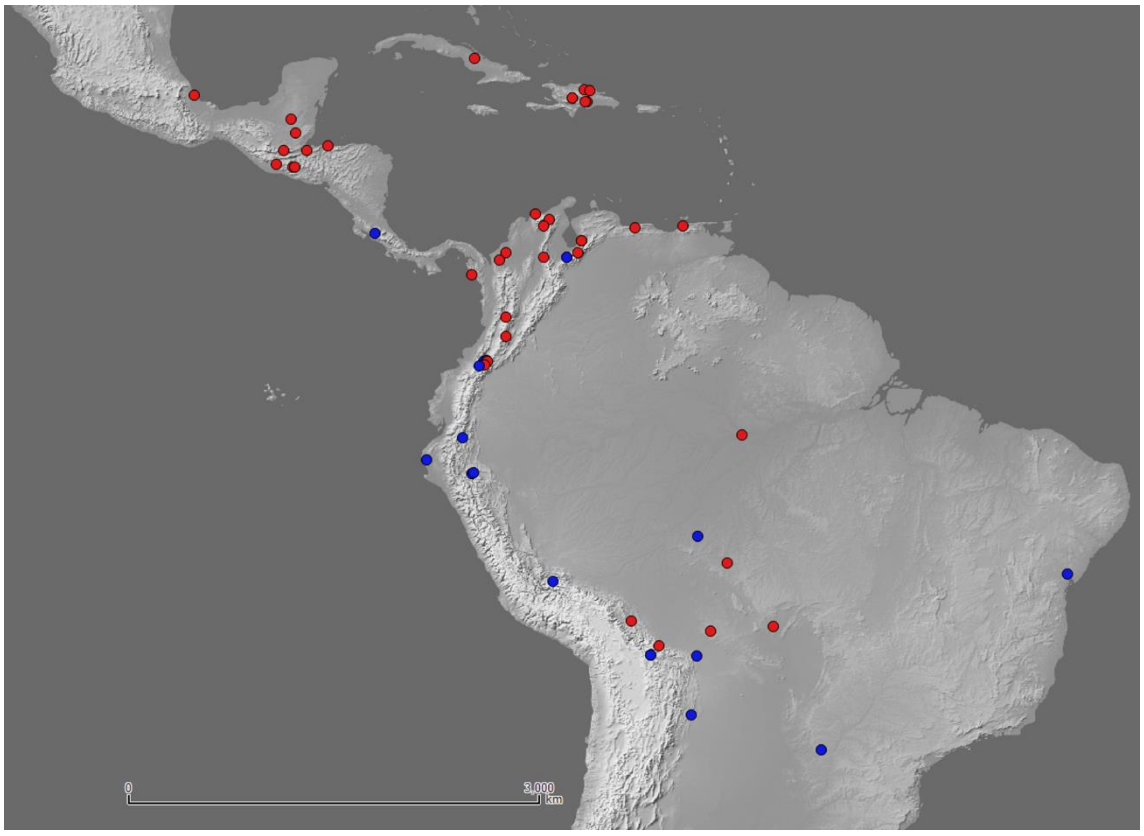


Figure 4.19. Place of collection of the American cultivated sweet potato specimens included in our Hyb-Seq study: blue circles, chloroplast lineage 1; red circles, chloroplast lineage 2.

4.2.3.1 | Hybridization with *Ipomoea trifida* and chloroplast capture

As explained before, the contrasting topologies obtained in the nuclear and chloroplast topologies are constant and strongly supported in all methods of phylogenetic inference that I used. These conflicting phylogenetic patterns could be the result of reticulated evolution from an ancestral chloroplast polymorphism or a multiple origin of the sweet potato. However, if the sweet potato had multiple origins, or if it had diversified from an ancestral polymorphism in *I. trifida* —which is unlikely, given that *I. trifida* is monophyletic in the chloroplast tree—, we would expect to identify traces of this pattern in the nuclear genome. However, I did not find such traces nor evidence of incomplete lineage sorting or recombination affecting the nuclear phylogeny.

The evidence presented here strongly suggests that the two distinct *Ipomoea batatas* chloroplast lineages are the result of an ancient hybridization between *I. batatas* and *I.*

trifida following species divergence. The result of that hybridization would be a sweet potato lineage carrying a chloroplast captured from *I. trifida*. (Figure 4.20). **Chloroplast capture** is the introgression of a chloroplast genome from one plant species into another, sometimes with no evidence of nuclear gene flow (Reiseberg and Soltis, 1991), and is commonly proposed to explain inconsistencies between phylogenetic trees inferred from different genomes (Folk et al., 2017; Reiseberg and Soltis, 1991).

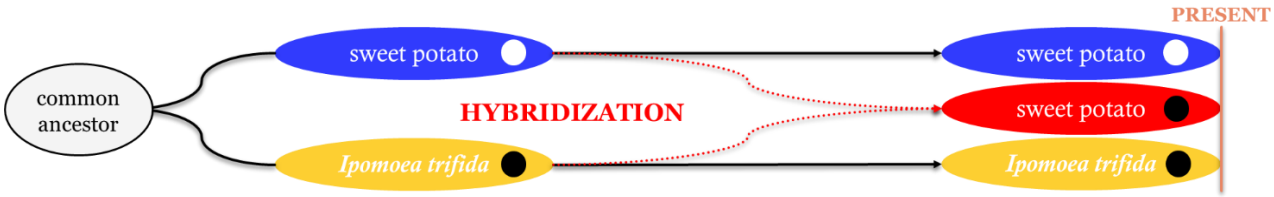


Figure 4.20. A hybridization event between the sweet potato (*Ipomoea batatas*) and *I. trifida* following species divergence generated a second sweet potato lineage that carries the chloroplast captured from *I. trifida* (black circle), hence explaining the conflicting phylogenies inferred from nuclear or chloroplast sequence data.

In the context of our results, we consider two possible mechanisms of chloroplast capture (Figure 4.21). In the first mechanism, the result of the hybridization between a female *Ipomoea trifida* (diploid) and a male *I. batatas* (hexaploid) would be an allotetraploid entity carrying a *trifida*-like chloroplast. This entity would later give rise to a new hexaploid form by one of three mechanisms:

- A cross between a normal gamete (2n) and an unreduced gamete (4n) of the tetraploid form, which would produce a hexaploid.
- A new hybridization event with diploid *I. trifida*, generating a triploid entity that subsequently doubled to yield a hexaploid.
- Additional autopolyploidization from the tetraploid intermediate and subsequent genome reduction.

Regardless the mechanism by which the new hexaploid formed, this entity would coexist with the "original" hexaploid *Ipomoea batatas*. Repeated crosses between both

entities would lead to the progressive loss of the *trifida* component of the nuclear genome in the new lineage, while maintaining a *trifida*-like chloroplast (Figure 4.21A).

In the light of our results, a possible alternative mechanism is **asymmetrical hybridization**, for which multiple examples have been described in recent years (Hedtke and Hillis, 2011). In this situation (Figure 4.21B), the entire nuclear genome would have been provided by an unreduced *Ipomoea batatas* (6n) sweet potato male gamete, whereas the chloroplast would have been provided by an *I. trifida* maternal progenitor. This mechanism would not only explain our conflicting genomic phylogenies, but also the lack of any evidence of the two lineages in the nuclear genome because it would be identical to that of the ancestral *I. batatas*. Furthermore, if this mechanism is correct, it would explain the capture of the chloroplast by *I. batatas* without the need of a second polyploidization event.

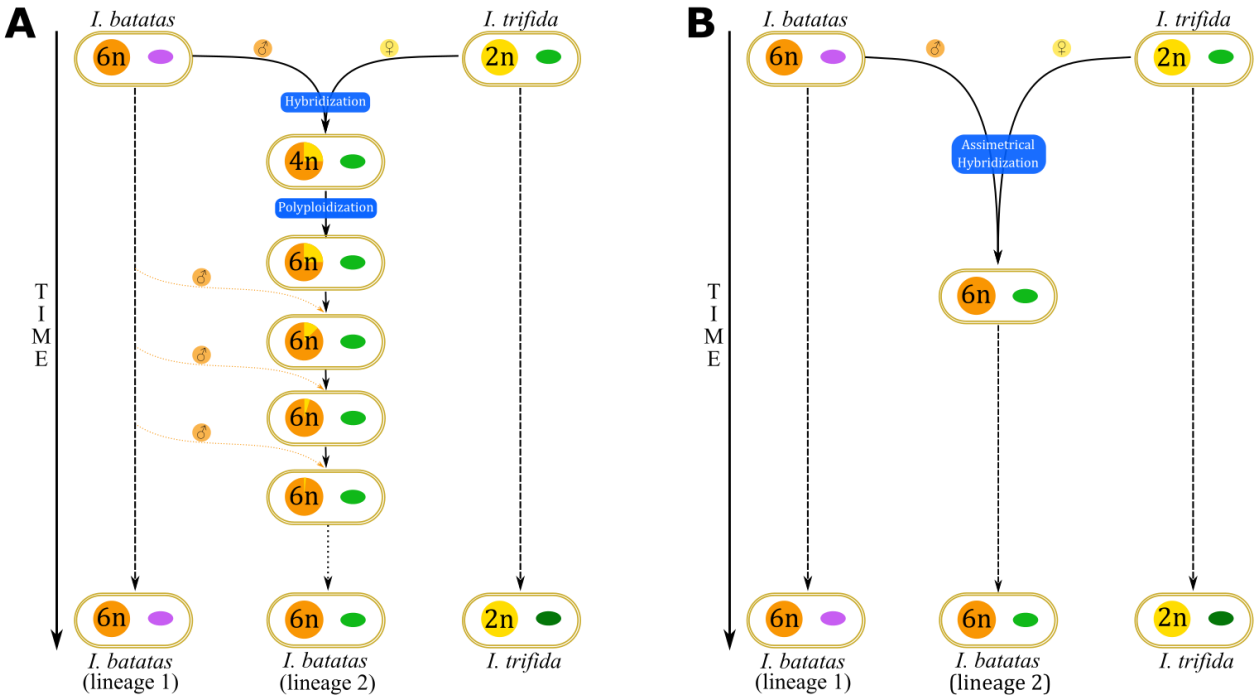


Figure 4.21. We consider two mechanisms of chloroplast capture by *Ipomoea batatas* from *I. trifida* that would explain the conflicting topologies retrieved from nuclear and chloroplast genomes. Orange and yellow = nuclear genomes of *I. batatas* and *I. trifida* respectively; purple = original *I. batatas* chloroplast; light green = captured *I. trifida* chloroplast, dark green = current *I. trifida* chloroplast. (A) The hybridization between *I. batatas* and *I. trifida* produced an allotetraploid

entity and, subsequently, a new hexaploid form by further polyploidization. Subsequently, the nuclear component of *I. trifida* in the newly formed hexaploid would be lost after several generations of introgression with ancestral *I. batatas*. Alternatively, (B) the new hexaploid entity could be the result of asymmetrical hybridization, in which the result of the hybridization between *I. batatas* and *I. trifida* would inherit its chloroplast from *I. trifida* and its entire nuclear genome from *I. batatas*.

4.2.3.2 | Implications of two distinct sweet potato lineages for sweet potato breeding

During my visit to CIP in 2017, I learnt that two sweet potato varieties widely used in breeding programmes, namely ***Beauregard*** and ***Tanzania***, differ in several traits: storage roots of *Beauregard* have orange flesh and low content of dry matter, whereas those of *Tanzania* have white flesh and high content of dry matter (Dorcus Gemenet, pers. comm.). Other studies have detected, for example, different levels of resistance to root-knot nematodes, with *Tanzania* highly resistant and *Beauregard* highly susceptible (Cervantes-Flores et al., 2008).

Considering our results about the origin of the sweet potato, I wanted to investigate whether the phenotypic diversity in these varieties had any relationship with the existence of two sweet potato chloroplast lineages. Previously, during the course of this project, I realised that a phylogeny inferred using the non-coding chloroplast region *rpl32-trnL*, approximately 1,200 bp long, allows the identification of various groups in the Batatas group with high confidence: the group of perennial species, the group of annual species and, importantly, the two sweet potato chloroplast lineages and *Ipomoea trifida*. Therefore, I decided to use this region to investigate the position of the two sweet potato varieties in relation to the two chloroplast lineages.

I obtained the *rpl32-trnL* region of a *Beauregard* and a *Tanzania* specimens using Sanger sequencing (DNA extracted and provided by CIP) and combined them with the regions extracted from all chloroplast genomes of Batatas to infer the position of the two commercial varieties in a phylogeny of Batatas reconstructed using this marker. Both

varieties belong to *Ipomoea batatas* CL1, the lineage that carries the "ancestral" *I. batatas* chloroplast (Figure 4.22).

Subsequently, I used the complete set of sweet potato chloroplast and nuclear data to test whether levels of genetic diversity are significantly different between lineages. The results of the statistical analyses confirmed that sweet potato CL1 holds higher values of genetic diversity than CL2 (see Methods section 2.7.1 and Appendix 1). This result confirms that sweet potato CL1 contains much phenotypic and genetic diversity, which may be of interest for sweet potato breeders because other varieties in CL1 could contain other traits of interest for crop improvement.

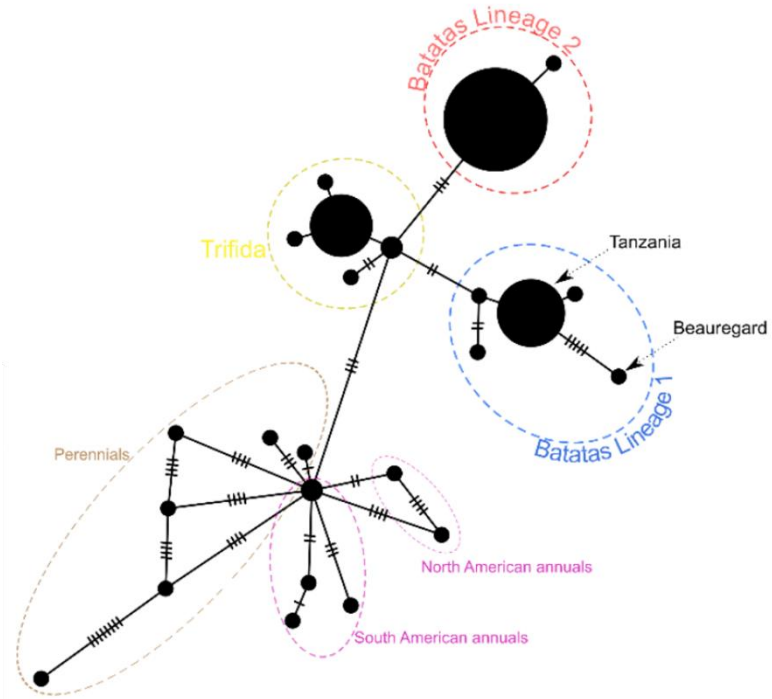


Figure 4.22. Median-joining phylogenetic network of the species closely related to the sweet potato, inferred using the *rpl32-trnL* chloroplast fragment. Two sweet potato varieties widely used in breeding programmes, *Beauregard* and *Tanzania*, belong to *Ipomoea batatas* chloroplast lineage 1 (blue ellipse), the lineage carrying the "ancestral" sweet potato chloroplast. Size of the dots equals number of samples (i.e., bigger circles group samples with identical sequences for this fragment).

4.2.3.3 | Timeframe of sweet potato evolution

We estimated divergence times in the Batatas group to investigate whether the sweet potato originated before or after humans; in other words, we wanted to test whether the storage root of the sweet potato was a consequence of human domestication. For this reason, it is important to mention that we forced our analyses to provide dates as young as possible to challenge this hypothesis, so it is likely that the dates presented here are younger than real divergence times.

According to our time-calibrated nuclear phylogeny, *Ipomoea batatas* diverged from *I. trifida* at least 800,000 years ago (Figure 4.8A), whereas according to the time-calibrated chloroplast phylogeny the minimum age of this divergence is more recent, at least 400,000 years ago (Figure 4.8B). These different estimates are likely to reflect the different evolutionary histories of the nuclear and chloroplast genomes, as well as the different patterns of molecular evolutionary rate variation between lineages. In addition, our dates are congruent with those obtained in a recent study of the sweet potato genome by Yang and collaborators (Yang et al., 2017) and clearly show that the origin of *Ipomoea batatas* predates modern humans.

Subsequently, the hybridization between *I. batatas* and *I. trifida* that led to chloroplast capture occurred within 56,000 years of the divergence of the two species (Figure 4.8B).

4.3 | SUMMARY AND CONCLUSIONS

In this chapter, I have presented a comprehensive phylogenetic study of the sweet potato and its wild relatives. The dense taxon sampling and large amount of data in our analyses allowed us to produce a robust phylogenetic framework with which to investigate the origin of the crop and the relationship between species in the Batatas group.

The Batatas group consists of fifteen putative species; *Ipomoea splendor-sylvae*, a perennial species, is sister to all other taxa, which in turn split in two strongly supported clades: one formed by all other perennial species and another formed by all annual species, the sweet potato and its closest relative *Ipomoea trifida*. All perennial species, three annual species, *I. batatas* and *I. trifida* are monophyletic, whereas the other species are not. All species have an American distribution except *Ipomoea littoralis*, which is restricted to the Palaeotropics but absent from America.

Ipomoea cordatotriloba, previously thought to have a disjunct distribution, splits in two groups: the North American specimens form a monophyletic group sister to *I. lacunosa*, whereas the South American specimens form a clade with two other South American species, *I. cynanchifolia* and *I. grandifolia*, none of them monophyletic. The North American lineage and the South American specimens have different evolutionary histories and should be considered independent species.

Among the non-monophyletic species, three South American entities (*I. cynanchifolia*, *I. grandifolia*, and South American *I. cordatotriloba*) form a clade where none of them is monophyletic, whereas *I. leucantha* is polyphyletic, with all specimens in our phylogenies scattered in different clades. A re-evaluation of these species and possibly a re-definition of their boundaries are needed.

Ipomoea batatas, the sweet potato, is a hexaploid species that evolved by autopolyploidy from the common ancestor with *I. trifida*, a diploid species with a Circum-Caribbean distribution. Our results refute the hypothesis of a hybrid origin of the crop and strongly suggest that the other extant species in Batatas, more distantly related, had no role in the origin of the crop.

Ipomoea batatas originated at least 400,000 years ago, well before modern humans. This, together with the fact that storage roots have evolved multiple times independently

in *Ipomoea*, as shown in the previous chapter, strongly suggests that the sweet potato storage root is not a consequence of domestication.

We identified two sweet potato chloroplast lineages, one of which captured its chloroplast from *Ipomoea trifida* in a hybridization event between both species shortly after species divergence. Therefore, although *I. batatas* evolved from *I. trifida* by autopolyploidy, the chloroplast capture provides evidence that there was a subsequent hybridization between the two species, and so the sweet potato contains two elements: one that is an autopolyploid and another that is technically an auto-allopolyploid. The ancestral sweet potato lineage has higher genetic diversity than the lineage that captured the chloroplast from *I. trifida*, and it could possibly hold hidden genetic resources of interest for sweet potato breeding.

5 | GEOGRAPHICAL ORIGIN OF THE SWEET POTATO AND ITS DISPERSAL TO POLYNESIA

«The early history of the sweet potato in the Pacific is important, for its large subsistence role. It is also peculiar, from its American origin, and it is also recalcitrant, because vegetables with tubers are so much less visible archaeologically than are grains and seeds.»

J. Hather & P.V. Kirch, 1991

5.1 | INTRODUCTION

Few questions in crop research have motivated as heated discussions, even between researchers from different disciplines, as those related to the origin and early dispersal of the sweet potato. Dozens of papers, and consequently multiple different hypotheses, have been put forward to answer the two same questions: where did the sweet potato originate and how did it come to be widespread across the Pacific islands by the time Europeans first arrived? As I will explain, the American origin of the plant is now generally accepted, although the specific region of America is not yet known. In contrast, the second question remains unanswered and studies that addressed it, often prolific in details (Yen, 1971), are also shrouded in speculation. Furthermore, as pointed by Pickersgill and Heiser Jr. (1978), the problem of sweet potato presence in both sides of the Pacific has somehow obscured the investigation of its actual geographical origin and ancestral distribution in America.

Three possible explanations for the presence of sweet potato in Polynesia have been proposed: dispersal by natural means, transportation by indigenous inhabitants of either side of the Pacific in pre-European times, or transportation by Europeans. Currently, the predominant theory combines the two latter explanations and is called the "tri-partite hypothesis" (Barrau, 1957; Denham, 2013; Roullier et al., 2013c; Yen, 1971). This hypothesis suggests the independent introductions of three sweet potato lineages from America into the Pacific: two introductions by Spanish and Portuguese travellers, from Mexico and the Caribbean respectively, in the 16th century and another, earlier introduction from South America by pre-European travellers around the 11th century. The two European introductions are well documented (Barrau, 1957; Yen, 1960, 1974). In contrast, the earliest hypothesised introduction has been assumed by many authors to be consequence of

human transportation (Denham, 2013; Green, 2005) but evidence for this claim is fragmentary and consists mainly of indirect evidence. Alternatively, a pre-historic arrival as the result of long-distance dispersal by natural means has also been proposed (Montenegro et al., 2008; Zhang et al., 2004) but has received little attention.

What follows is a detailed account of the body of evidence in relation to the origin and diffusion of the sweet potato, identified from different sources and mainly in the Americas and in the Pacific. My aim is to separate evidence from assertion and conjecture, which, as mentioned before, is abundant. Literature on these topics is extensive, although a lot of references merely repeat information previously published and confusion is frequent. For that reason, in this review I have attempted to highlight documents that provide original information, but also cite some more general studies and reviews for the relevance they had for subsequent studies (e.g. Heyerdahl, 1952; Yen, 1974). In regard to the Pacific islands, it has been shown by many authors that the sweet potato has been a staple in several Polynesian societies for centuries (Coil and Kirch, 2005; Handy, 1940; Kirch et al., 2005, 1995, 2004; Ladefoged et al., 2005) and it is not my intention to deny its importance in historic times. However, I aim to draw attention to the fact that the theory of a human-mediated pre-European introduction of the crop into Polynesia, accepted by many scholars without further discussion and in some way predominant nowadays, is reliant on a fragmentary set of evidences, which calls into question the validity of this theory.

Certain disciplines have had a predominant role in understanding the distribution of the sweet potato in ancient times, for example archaeology and linguistics. In contrast, the *botanical* commentary has been generally neglected:

> *«It is rather surprising that relatively few botanists have concerned themselves with the problem, when many of the basic facts pertaining to it are botanical».*
>
> (Yen, 1960)

My aim with this chapter is to help fill that gap. Following the literature review, I present the results of our genomic analyses in relation to the geographical origin of the sweet potato and its diffusion to Polynesia. I first explore several limitations of previous studies on the geographical origin of the crop and provide new insights based on the distribution of the sweet potato and its closest relative, *Ipomoea trifida*. Regarding its diffusion to Polynesia, I present data compatible with a trans-oceanic diffusion of the plant without human mediation. This possibility is strongly supported by several patterns identified in the group of species closely related to the crop and across the whole of *Ipomoea*. Our results help to understand the origin and early diffusion of the sweet potato, not only within America but also throughout the world and, consequently, how it arrived in Polynesia.

5.1.1 | Geographical origin of the sweet potato

In 1753, Linnaeus described *Convolvulus batatas* as a plant present in both Eastern and Western Indies (Asia and America), without reference to a specific place of origin —*«Habitat in India utraque»*[38]. Subsequently, Lamarck (1793) cited it as an American plant also cultivated in Asia when he transferred the name to *Ipomoea batatas —Ex America; culta in utrisque Indiis*[39]. In addition, Choisy (1833) described *Batatas edulis* (based on Thunberg's (1784) *Convolvulus edulis*) as a plant cultivated throughout the world but of Asian origin —*«Ex India orientali nata, ubique in tropicis regionibus culta»*[40]. It was soon agreed that both names, *Ipomoea batatas* and *Batatas edulis*, referred to the same

[38] Linnaeus (1753), p. 154.
[39] Lamarck (1793), p. 465.
[40] Choisy (1833), p. 435.

entity (Meyen, 1836) and therefore the plant must have originated in one region or the other. Thereafter, several authors noted the similarities between both entities (e.g. Candolle, 1883; Clarke, 1885; Hooker, 1867; Miquel, 1856; Reinhardt, 1911) and considered a single origin in Asia (Bojer, 1837: 225; Choisy, 1845: 338) or in America (Humboldt, 1825: 470, Boissier, 1839).

It was Alphonse De Candolle who, in 1883[41], first discussed the question in detail and proposed that different lines of evidence support an American origin of the crop. The arguments in favour of this hypothesis (and against the hypothesis of an Asian origin), offered by De Candolle and supported by subsequent authors until the present, are fourfold:

1. **Most species closely related to the sweet potato are found in America** (Austin, 1978; Bohac et al., 1993; Candolle, 1883; Khoury et al., 2015). As I have shown in the previous chapter, all but one species in the Batatas group, including the sweet potato closest wild relative *Ipomoea trifida*, have a natural distribution restricted to the Americas.
2. **There is no historical record of African or Near East sweet potato cultivars** before European travellers brought it from America (see below), something to be expected if its origin was in Asia (Candolle, 1883; Conklin, 1963; Yen, 1971).
3. **The oldest fossil remains found in America** and assigned to the sweet potato date back more than 8,000 years (Engel, 1970a) or at least 3,500 years (Ugent et al., 1981), whereas the oldest fossil found in Oceania dates back to the 13-15th century only (Horrocks et al., 2004a).

[41] De Candolle (1883), p. 53.

4. **The etymology of the names** found in Asia and the Pacific —*umara*, *gumarra*, *gumalla*— resemble the American names —*cumar*, *kuala*, *umala* (Austin, 1988b; Buck, 1938; Candolle, 1883; Conklin, 1963; Hooker, 1867; Humboldt, 1825).

In summary, the evidence indicates that the sweet potato is a species of American origin. However, where in the Americas did sweet potato originate is a question that remains unanswered. Most authors vaguely indicate Mesoamerica, Tropical America or South America (e.g. Austin, 1988b; Dressler, 1953; Pickersgill and Heiser Jr., 1978; Roullier et al., 2013c; Yen, 1971). Other authors are more specific and point for example to the Mexican region of Veracruz and Oaxaca (Nishiyama, 1971) or to the Amazonian periphery (Clement et al., 2010, 2015), but evidence to support these claims is still lacking.

5.1.2 | Early dispersal, transportation and cultivation of the sweet potato around the world

5.1.2.1 | Prehistoric evidence of sweet potato cultivation

Archaeological evidence associated with the sweet potato is, in most cases, fragmentary and dubious and often it is not possible to determine whether the remains come from cultivated or wild plants. Macro-remains found in coastal Peru, inland Peru and Polynesia, the accuracy of which is discussed in this section, have been identified as *Ipomoea batatas* (Figure 5.1).

In contrast and as far as I am aware, no macro-remains associated with *I. batatas* have been found in the Amazonian region, Central America nor the Caribbean, where remains of other root crops such as cassava (*Manihot esculenta* Crantz) are abundant (Watling et al., 2018). Micro-remains —mainly starch grains, because *I. batatas* does not seem to produce phytoliths (Horrocks et al., 2004a; Pearsall, 1993)— have been tentatively assigned to the sweet potato across the continent (Pezo-Lanfranco et al., 2018;

Piperno and Pearsall, 1998; Wesolowski et al., 2010). However, analyses by Linda Perry (Perry, 2002) found that the size of starch grains in sweet potato, as well as in cassava, varies notably, even between grains produced by the same root. In addition, Perry found that starch grain size and date of deposition do not correlate, and therefore the value of starch grains to differentiate between cultivated sweet potato, non-cultivated sweet potato and other species of *Ipomoea* must be treated with caution[42].

Despite the lack of reliable archaeological evidence, the tropical American region has been postulated by many authors as the origin of the crop:

> *«Unfortunately, it is grown principally in areas of abundant rainfall, and the archaeological record for the sweet potato is, consequently, very unsatisfactory.* […] *While sweet potato has probably been widely grown around the Caribbean for thousands of years, no archaeological evidence confirms this probability, nor is there any archaeological evidence by which the rate of distribution of the sweet potato may be checked. »*
>
> (Smith Jr., 1968)

Finally, sweet potato is recognised as a post-European introduction and only marginally cultivated in different parts of the Pacific (Burrows, 1936, 1937; Lepofsky, 2003; Whistler, 1988). Its introduction in Papua New Guinea and Melanesia is well-documented to be of European origin (Barrau, 1957; Yen, 1960, 1974) and provoked intensive changes in those places over a short period of time (Denham, 2013; Watson, 1977).

---

[42] I was only able to find two experimental studies of starch granules in *Ipomoea* species other than *I. batatas*. The first study is by Reichert (1913), who studied starch granules in many species of plants, including several Convolvulaceae such as *I. jalapa*, *I. littoralis* or *I. purga*, all with overlapping size ranges. The second study is by Asante and colleagues (1993), who studied starch granules in pencil-like roots of diploid *I. trifida*, the closest relative of the sweet potato. Although their studies showed a bigger average size of starch grains in the cultivated plant (Table 2 in Asante et al., 1993), the size ranges between both species overlap; that is, large grains produced by *I. trifida* plants are bigger than small grains produced by the cultivated *I. batatas*. They also found that amylose content is similar or even bigger in *I. trifida*.

In addition, Horrocks and collaborators (2004b) comment: «*Similarly, starch grains and xylem of sweet potato are not easily differentiated from those of the single indigenous species* [in New Zealand] *(Ipomoea cairica)* ».

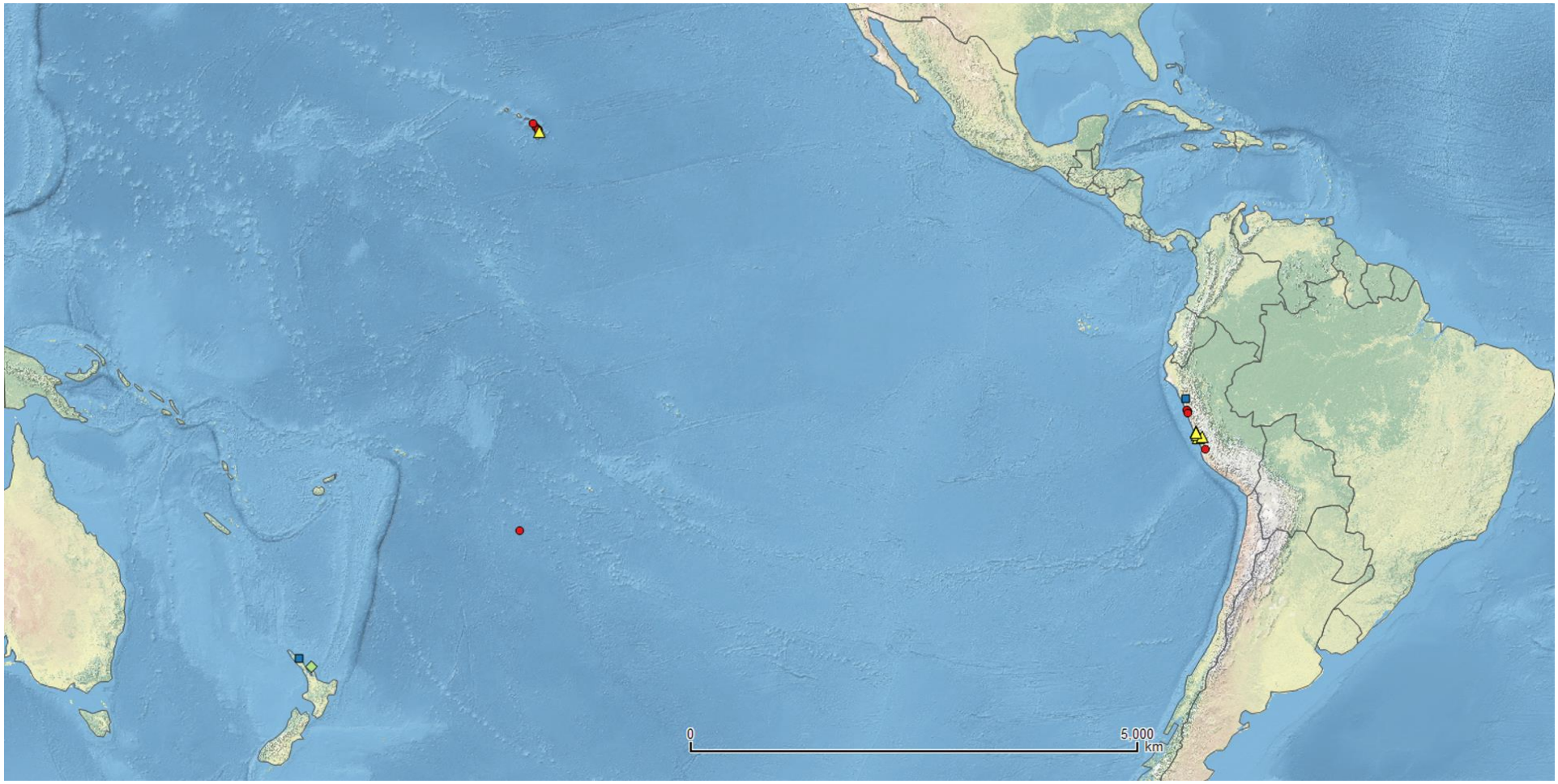


Figure 5.1. Archaeological remains associated to the sweet potato in the American continent and the Pacific. Only remains discussed in the text are presented here. Other micro-remains (pollen grains, starch grains and xylem fragments) have been associated to the sweet potato in the American continent (see for example Pezo-Lanfranco et al., 2018; Piperno and Pearsall, 1998; Wesolowski et al., 2010). Red circles: root fragments; green rhombus: starch grains and xylem fragments; blue squares: indirect evidence; yellow triangles: uncertain evidence.

5.1.2.1.1 | Sweet potato in coastal Peru

Wilhelm Reiss and Alphons Strübel were the first archaeologists to identify sweet potato remains in coastal Peru. They found these remains by the end of the 19th century at the Acropolis of Ancón, some 40 kilometres north of Lima and inhabited approximately since 8000 B.C. and until the Inca conquest (Wittmack, 1887)[43]. However, the authors did not follow the methods of modern archaeology and did not date the remains, which makes it impossible to ascribe them to any specific era. Thirty years later, William E. Safford presented different remains from the same site at Ancón at the 19th International Congress of Americanists (Safford, 1917). Safford included sweet potato remains among his findings but did not provide a geographical nor temporal context for the remains — only cited them as "prehistoric" in the proceedings of the congress— and, considering that the place was inhabited until the Inca empire, these remains could well belong to a more recent epoch. It is unfortunate, as noted by Towle (1958) that these and other collections not discussed here were carried out before the development of modern stratigraphy, and therefore discoveries presented by those authors cannot be associated to a specific culture.

Perhaps adding more doubts about the cultivation of the sweet potato in coastal Peru in ancient times, the plant was absent in two studies of plant remains identified at Ancón site, coetaneous to those aforementioned. De Rochebrune (1879) and Dorsey (1893) recorded root crops such as the "olluco" (*Ullucus tuberosus* Caldas) or the "achira" (*Canna indica* L.) but did not mention sweet potato. In addition, one other report, published twenty years later and including not only Ancón but also the Chorrillos, Pachacamac and Rinconada sites (all of them in the vicinity of Lima), recorded two other

[43] This reference is a mere sketch of what is labelled as "*Convolvulus Batatas* L. Tuber of the batata", but the authors do not provide a description nor explanation of the finding.

root crops: "aro" (*Xanthosoma sagittifolium* (L.) Schott) and "manioc" (*Manihot esculenta*[44] Crantz), but did not mention any sweet potato remains (Costantin and Bois, 1910).

Later, between the 1960s and the 1980s, a team of archaeologists from the United States presented a series of new discoveries from Peruvian archaeological sites. First, Patterson and Moseley (1968) described some tuberous remains c. 1300 B.C. found at Ancón that "could be" sweet potato, but Ugent later suggested that the only existing sweet potato remains at Ancón are very recent, from the Inca period (1000 – 1500 A.D.) (Ugent et al., 1981).

Perhaps the only case in which a comprehensive context was provided was the identification of eighteen sweet potato roots from three adjacent sites in coastal north-central Peru, in the Casma valley: nine at Pampa de las Llamas (1800 – 1500 B.C.), one at Huaynuma (2000 B.C.) and eight at Tortugas (1800 – 1500 B.C.) (Ugent et al., 1981). In this publication, Ugent and collaborators provided the most complete identification of any sweet potato remains in America so far, describing their physical appearance and their starch grains. Assuming, therefore, that these remains represent true sweet potato, we should also assume that sweet potato was cultivated in the region at least 3500 years ago. It is worth noting, nevertheless, that the remains found by Ugent are notably small: the entire roots described are 18, 31 and 32 mm long, respectively (see figures in Ugent, 1968). As indicated by Ugent (1981), these remains are smaller than those from the Chilca Canyon (discussed below) and from others at Pachacamac (60 mm in length according to the author), which are also small compared to modern sweet potatoes[45].

[44] Recorded as *Manihot utilissima* Pohl.

[45] As explained by Ugent, the different size must be due, at least in part, to the total loss of water, but it could also indicate that a human-mediated selection of sweet potato size occurred in more recent times.

Also in the second half of the 20th century, several archaeologists mention new records found by themselves and by other authors from coastal Peru, but these references invariably are repetitions of previous works by the same authors[46] or, when original, provided no information or direct evidence with which to assess the value of their discoveries. As an example, Lanning (1965) cited sweet potato remains dated in 1900–1750 B.C. from Punta Grande, in the Chillón valley, but the publication lacks any description of what type of remains these are or where they were deposited, as Lanning only mentioned them in passing during a description of early human settlements in coastal Peru. One other example is the identification of a sweet potato root at Pachacamac in 1973: Ugent and Peterson (1988) cited this discovery but did not provide any further details. Along similar lines, another previous study mentions the possibility of sweet potato remains existing in another site 60 km north of Lima (Lauri, Chancay valley)[47], but neither images nor further analyses are provided (Friedberg, 1959).

Ultimately, we must agree with Bonavia (1984) who warned that, in the end, factual evidence of sweet potato cultivation in coastal Peru in ancient times is not so abundant.

5.1.2.1.2 | Sweet potato in inland Peru

The situation is slightly different in more mountainous regions of Peru, although sweet potato remains are not abundant there either. If the stratigraphy presented is accurate, the oldest remains identified as sweet potato roots in Peru were found by Bernardino Ojeda and Frédéric Engel at the *Tres Ventanas* cave complex, located in the Chilca Canyon at an altitude of around 2,800 meters a.s.l. (Engel, 1970a). Engel claimed to have

[46] Mainly Mark N. Cohen and Thomas C. Patterson or their mentor, Margaret A. Towle (Cohen, 1978a; MacNeish et al., 1975; Patterson, 1971; Patterson and Moseley, 1968). According to the author, (Cohen, 1978) was published elsewhere at least two more times, in 1975 and 1977.

[47] *«Seul parait pouvoir etre identifié de façon certaine (morphologie de la racine et des grains d'amidon) le nº 16, qui doit etre sans aucun doute* Ipomoea batatas.*»* [Only number 16 can be identified by certain root morphology and starch grains, which must be without doubts *Ipomoea batatas*].

collected tuberous roots at different levels in Cave II (there are three main caves), the oldest stratum being carbon-dated in 8080 ± 170 years B.C. However, besides lacking a specific date for the roots, other aspects indicate that Engel's discovery must be taken with caution: first, Engel contradicted himself in subsequent works by assigning the same sweet potato remains to different caves and different strata each time (Engel, 1970b, 1970c, 1973). Even more worrying, as confirmed by Bonavia (1984) in conversation with one of Engel's collaborators, it is likely that the remains from different strata were mixed during the cleaning process.

A few years after Engel's discovery at Chilca canyon, Douglas E. Yen re-examined the sweet potato remains found by Engel and provided a detailed description of the roots, indicating that they are *«approximately 3 inches* [~7.6 cm] *in size»* (Yen, 1974: 25). However, although Yen identified the remains as sweet potato roots, he also considered the possibility that these roots were of wild provenance. In line with Yen's hesitation to confirm the cultivated status of the sweet potato at Chilca, it is worth noting at this point the discovery presented by Ugent and Peterson (1988) that ancient Peruvian cultures used other wild species in prehistory, for example the wild potato species *Solanum maglia* Schltdl., and their suggestion that they probably used other sweet potato wild relatives as well.

There is strong evidence from pottery that the Mochica civilization (100–700 A.D.) in northern Peru —current La Libertad department— cultivated sweet potatoes (Towle, 1958). Pottery is even more abundant as we get closer to the Inca period (Tello, 1924; Towle, 1958; Yen, 1974), which logically indicates that inhabitants from that region were already familiar with the crop by that time —around 500 A.D. according to several sources (Yen, 1974). There are, however, no records of sweet potato remains in the region

in earlier times, which could be probably expected considering it is not far from the Casma valley previously mentioned.

Finally, with regard to the southernmost part of Peru, sweet potato has been identified in a Chincha settlement 200 km south of Lima, but the dates provided, ca. 1000 – 1500 A.D., are just pre-Inca (Ugent and Peterson, 1988). Further south, although Ugent and Peterson (1988) mention sweet potato records from Paracas[48], the crop is absent from archaeological remains and from pottery recovered in southern Peru and northern Bolivia from the Nazca (100–800 A.D.) and Tihuanaco cultures (1500 B.C. – 1200 A.D.), which likely indicates that it was not cultivated in that region until more recent, historic times.

In summary and despite many claims, the oldest confirmed pre-historical sweet potato or sweet potato-*like* archaeological remains seem to be those found at three locations in the Casma Valley, dated at between 2000–1500 B.C. This is the only clear evidence of sweet potato presence in coastal Peru[49] in ancient times and, together with the lack of other remains elsewhere in Peru, perhaps suggests that its cultivation in ancient coastal Peruvian settlements remains a matter of conjecture. Sweet potatoes, in the same way as most other crops that later became staples in the region (e.g. squash, maize, common-bean) was likely introduced here from other regions, since there is no evidence of on-site domestication (Cohen, 1978b) and its closest wild relative¸ *Ipomoea trifida*, does not reach such southern latitudes in its distribution range.

[48] Ugent and Peterson (1988) refer to a study by Tello and Nordenskiold published in 1931, but in that reference (Nordenskiöld, 1931) there is no mention to the sweet potato. Nevertheless, the Paracas culture developed between 700 B.C. and 200 A.D., which is of little relevance in terms of ancient distribution of the crop.

[49] And as far as I am aware also in the rest of the country, although archaeological activity elsewhere is not so extensive.

5.1.2.1.3 | Sweet potato in the Pacific: evidence from archaeology

Literature is littered with mentions of the presence of sweet potato in the Pacific region in ancient times. Records that are supported by physical evidence are discussed below, whereas evidence collected from oral traditions and legends that have been thoroughly studied by previous authors (Best, 1925; Heyerdahl, 1952) is beyond the scope of this thesis. However, if we are to accept oral tradition as a source of evidence, a brief mention of two specific events collected from native traditions may be relevant. These events have been recorded by Turei and Kapiti (1912), Best (1925), Heyerdahl (1952) and many others. As the story goes, sweet potatoes arrived in New Zealand from Tahiti (some 4,000 km away) on-board a vessel called "Horouta". Later, at least one other vessel (named "Te Aratawhao") would have sailed from New Zealand to Polynesia in search of more sweet potatoes. Although this does not affect the question of how sweet potato arrived in Polynesia, it shows that islanders could be responsible for the dispersal of the crop to western Pacific islands and New Zealand. It seems indisputable, not only from oral traditions but also from other lines of evidence, that contacts between different islands in ancient times existed (Best, 1925; Buck, 1938; Cox and Banack, 1991; Hornell, 1946). Thus, if the sweet potato was present anywhere inhabited in the Pacific in ancient times, Polynesians could have had a role in its distribution between islands.

Archaeological remains assigned to the sweet potato have been found in Mangaia Island (Cook Islands), Hawai'i and New Zealand. These remains are more recent and, in many cases, less compelling than those in the American continent. Also, some authors suggest that sweet potato was present in Easter Island in "prehistoric times" (see for example Coil and Kirch, 2005; Wallin et al., 2005), but I was unable to find records of

archaeological sweet potato remains supporting this claim[50, 51]. Several findings in the Pacific region provide indirect evidence interpreted by the authors *«as having been for sweet potato cultivation»* (Coil and Kirch, 2005), for example the storage pits and mounds in New Zealand mentioned below. Regarding direct evidence, in the shape of carbonised fragments, two issues must be considered in the discussion:

1) Most remains could not be undoubtedly assigned to a pre-European era, as admitted even by some of the most vehement proponents of the existence of transoceanic contacts (see for example Jett, 2017).

2) It is not clear to me to what degree of confidence authors are able to differentiate between the sweet potato and other *Ipomoea* or Convolvulaceae species with storage roots (see also section 5.1.4). The similarity of the sweet potato roots with those from other closely related taxa (McCormick, 1916) seems to generate doubts among other authors:

> *«On the whole, there are insufficient modern reference studies of these plants to produce solid evaluations of the degree to which tuber anatomy or microfossil characteristics can truly allow an archaeobotanical separation of sweet potato from other related, naturally occurring plants».*
>
> (Ladefoged et al., 2005: 363)[52]

Perhaps as a consequence of this, some remains assigned to sweet potato have been later re-identified as wood charcoal or as belonging to other species (Coil and

[50] Wallin and collaborators (2005) indicate that Roggeveen reported sweet potato being cultivated in Easter Island when he discovered the island in 1722. The transcription of his log book by Corney (1903: 21), which is the document cited by Wallin *et al*., records *«bananas, potatoes, sugar-canes* [...] *and many other kinds of the fruits of the earth»* but not sweet potatoes.

[51] Wallin and collaborators (2005) reported archaeological findings by Skjolsvold and Orliac. I was unable to access these publications, but Wallin *et al*. seem to doubt about the prehistoric status of those remains and rely on indirect evidence from ancient landscape to infer the presence of sweet potato in the island in the 13th century.

[52] See also Horrocks et al., 2004b: 153.

Kirch, 2005; Patterson and Lanning, 1964, 1966). Except for the remains from Mangaia Island and some in Hawai'i, authors do not mention the criteria upon which they identified their findings as sweet potato, which complicates discussion of their results. For example, it does not rule out the possibility of those remains belonging to another species in the genus *Ipomoea* as stated above.

Compelling evidence of sweet potatoes or sweet potato-*like* remains in the Polynesian archaeological record comes from Tangatatau site in **Mangaia**, a small volcanic island in the **Cook Islands archipelago** (Hather and Kirch, 1991). According to this and a subsequent study by Kirch and colleagues (1995), the site was inhabited between 1000 and 1600 A.D. and sweet potato remains were dated from 1000 A.D to 1400 A.D[53]. Furthermore, the level at which *Ipomoea* remains were found is also the level where plant remains are most abundant, which according to the authors indicates that the site hosted permanent settlers during that period. Among the remains there were no whole roots: they consist of vesicular fragments of different size, some of them including bits of root epidermis (Hather and Kirch, 1991).

Multiple macro and micro-remains associated with the sweet potato can be accounted for in the **Hawai'ian archipelago**, although most of them are recent and only one has an oldest estimated age that predates European contact (Ladefoged et al., 2005). Rosendahl and Yen (1971), for example, identified three rooteous remains associated with ancient fire pits, at least one of which they identified as sweet potato. Two of the remains, carbonised but almost entire, are very small (23 and 25 mm in length respectively) and the authors acknowledge that, given their small size, they could also be immature storage organs of other tuberous crops; the shape and size of the third piece,

[53] These ages were estimated for the layers above and below the one containing sweet potato remains, which was not dated.

around 80 mm in length, resembles sweet potato more than the others. Radiometric dating of this last piece suggested an origin between 1425 and 1725 A.D. which, as noted by the authors, does not serve to discard the possible transportation of the sweet potato into Hawai'i by Spanish ships in 1527 (see section 5.1.2.2), a theory which on the other hand has been heatedly contested by Dahlgren (1917) and Stokes (1932). Other remains were found at Kohala, in north western Hawai'i (Coil and Kirch, 2005), and Kahikinui, in Maui (Kirch et al., 2004), but these are dated to a post-European period. A thorough review and analysis of all sweet potato remains found in Hawai'i was published by Ladefoged and collaborators (2005), together with the description of more fragments identified as charred sweet potato in two trenches excavated at Kohala, north-western Hawai'i (trenches labelled 50 and 12). These remains were dated to 1290–1430 A.D. (trench 50) and 1640–1960 A.D (trench 12). Regardless of the exact date of the second case, the remains found at trench 50, assuming their identity and date are correct, provide the strongest support for the presence of the sweet potato in Hawai'i in pre-European times.

Finally, different micro-remains and storage structures have been presented as both direct and indirect evidence of sweet potato cultivation in **New Zealand** as early as 1300 A.D. For example, Horrocks and collaborators identified multiple starch grains and xylems at different coastal locations spanning 600 km on the northern island, frequently associated with *Colocasia* sp. remains, that they identified as *Ipomoea batatas* or, less likely, *I. cairica* (Horrocks, 2004; Horrocks and Lawlor, 2006; Horrocks et al., 2004a, 2004b). What they consider most relevant are the starch grains found in three human or dog coprolites at the Great Barrier Island (a small island north of the Northern Island), since these indicate ingestion of "partially cooked sweet potatoes" —although they do not explain how they know that these were cooked—, as well as the remains found in two mounds at the Pouerua region and in another site in southern Auckland, associated to

sweet potato cultivation. Although the authors first explain that these remains are clearly pre-European (indirectly dated around between 490 and 550 years BP), in subsequent papers they admit that the remains cannot be assigned to a pre-historic or historic period with certainty, because they could come from a more recent deposit that has been displaced downwards (Horrocks and Lawlor, 2006). In summary, the authors inferred that *Ipomoea* and *Colocasia* were dominant crops in the island in the past, but the remains cannot be undoubtedly assigned to a pre-historic nor a historic context.

In conclusion, archaeological evidence of sweet potato presence in the Pacific is restricted to two of its extremes (Hawai'i and New Zealand) and Mangaia Island. The most compelling evidence comes from Hawai'i and Mangaia, each presenting remains dated well before European travellers arrived in the region. If the identification of these remains is correct, they would entail unquestionable evidence of sweet potato presence in Polynesia in ancient times.

5.1.5.2.4 | Sweet potato in the Pacific: evidence from linguistics

Different terms have been used, throughout history and in different places, to refer to the sweet potato. In America for example, the sweet potato is mainly known as *batata* (in the Caribbean region) and *camote* (in Mexico and South America), with variants and specific names in different parts of the continent. Linguistics has had a central role in the disputes about the origin and early dispersal of the crop into Polynesia because one of those variants, *cumar*, used in a Quechuan dialect from Ecuador highlands, resembles various terms used across the Pacific (*kumara*, *umara*, *gumalla*) in reference to the sweet potato. The similar terminology in distant places has been presented as evidence of human transport across the Pacific in pre-European times (Ballard et al., 2005; Heyerdahl, 1952; Yen, 1971), although other authors suggest that the profusion of terms created more confusion than it helped to clarify (Moreno Gómez, 2010). Linguists and archaeologists have

conducted thorough literature reviews and monographic studies of this question (Buck, 1938; Yen, 1974; Rensch, 1991), so here I summarise some of this literature.

In 1865, Berthold C. Seemann noted that the Polynesia term for the sweet potato, *kumara*, was identical to the term used by the Maoris from New Zealand and suggested that this was indicative of an introduction from those islands into New Zealand. Furthermore, Seemann (1865) and subsequent authors[54] noted the similarity between that name and the one used for the sweet potato in the Ecuadorian highlands, *cumar*, and Seemann was the first author to suggest that this resemblance (which he considered "identical") could indicate a transportation of the crop from South America by Pacific islanders[55]. It is important to note that Seemann also recorded a second vernacular name for the sweet potato in Fiji, *Kawai ni papalagi*, the translation of which (according to himself) would be "foreign *Dioscorea*". A similar observation was made by Ivens (1918: 48), who recorded *kumara* but also *uhi ni haka* ("foreign yam") in the Solomon Islands. However, these alternative vernacular names seem to have been systematically ignored by all subsequent authors except Dixon (1932: 58).

Seemann's observations were assimilated and repeated by Gray and Trumbull (1883: 248), Hillebrand (1888: 314), Friederici (1929: 477), Hornell (1946: 59), Carter (1950: 162) and others without further re-consideration. Later, Heyerdahl (1952: 429) presented the term *cumar* as used in other dialects in Ecuador, despite Seemann's observation that it was restricted to a local dialect. However, it seems clear that the use of the word *cumar* to refer to the sweet potato in America, both in pre-European and in colonial times, was restricted to Chinchasuyo, a Quechua dialect spoken by some inhabitants from

[54] Yen (1971) mentions several etymological studies in the 1920s by Imbelloni and Palavecino, but unfortunately I was not able to access those studies.

[55] Yen (1974) indicates that Markham (1864) also reported the word *cumar* as a vernacular name for the sweet potato, but I could only find the term *apichu* for the sweet potato in Markham's work.

the Ecuadorian highlands and was not used in any other Quechua region (Brand, 1971). As pointed out by Brand, this would make it highly unlikely for Polynesian people to *receive* this term in an exceptional visit to South America.

In summary, this linguistic similarity has received great attention in all studies about the sweet potato in the Pacific. Nowadays, scholars still accept the linguistic connection between South America and Polynesia, based on ambiguous sources, as strong evidence of sweet potato transportation into the Pacific by indigenous peoples (Clarke, 2009).

5.1.2.2 | Historic evidence of sweet potato cultivation and transportation

Christopher Columbus reached what now is America on the 12th of October of 1492. Historical records indicate that, only one year later, in 1493, sweet potatoes were already being cultivated in Malaga, southern Spain (Moreno Gómez, 2010). There is, therefore, little doubt that Columbus found sweet potatoes in cultivation in the Caribbean[56] and brought some roots with him back to Europe under the name "batatas"[57]. The crop quickly spread throughout the world transported in the vessels of European travellers. Portuguese merchants carried it to Africa, India and their territories in the Indian Ocean, whereas Spanish travellers exported it to northern Europe —it was known in England in 1560 (Simon, 1914)— and the Far East (O'Brien, 1972).

In 1521, when Portuguese explorer Ferdinand Magellan, in the service of the Spanish Crown, discovered the natural strait between the Atlantic and the Pacific oceans, a new world was revealed to Europeans. Importantly, Antonio Pigafetta, chronicler of Magellan's trip around the world, specifically cited the sweet potato (*batate*) among the

[56] The caravels did not reach the American mainland on Columbus' first voyage (de las Casas, 1552).

[57] This term seems to be of Haytian origin, which is congruent with the crop being first found by Columbus in the Caribbean region.

goods that they stocked up on during their stop in Brazil (*Terra del Verzino*) before setting sail southbound:

> *«Navigammo fra'l mezzogiorno e'l libeccio, finchè guignemmo in una terra, detta la Terra del Verzino* […] *Utilissima traffico noi facemmo con gli abitatori di quel paese. Per un amo da pescare, o per un coltello, ci davano cinque o sei galline; per un pettine un paio d'oche; per uno specchio, o per una cesoia, tanto pesce che avrebbe bastato a faziare dieci uomini; per un fonaglio o una stringa, una cesta di batate, que son certe radici lunghe come i navoni ed hanno il gusto della castagna.»*[58]
>
> (Pigafetta, 1524)

This suggests, firstly, that the sweet potato was already an important crop in south eastern Brazil, somewhere near present-day Rio de Janeiro, by that time (Sauer, 1950); and secondly, that there were sweet potatoes on-board the Spanish vessels when they sailed the Pacific for the first time. In subsequent years, the number of voyages commanded by Spanish travellers grew so quickly that the Pacific Ocean became known, in the 16th century, as "the Spanish Lake" (Spate, 2004). Even more, the discovery of the *tornaviaje* (literally, "return trip") by Andrés de Urdaneta, in 1565, allowed the establishment of a route across the Pacific between the Philippines and the American continent. This route was used for commercial purposes by the Manila galleon for over 250 years.

Beyond the Manila galleon, Spanish explorers travelled frequently across the South Pacific[59] and records exist that many islands and archipelagos were discovered and visited during the 16th and early 17th centuries, for example Guam, the Marianas and the Philippines (Magellan, in 1521); the Marshall Islands (Alonso de Salazar, in 1526); Hawai'i

[58] *We navigated between midday and the southwest, until we reached a land called Terra del Verzino* [Brazil] […] *A very useful trade we made with the inhabitants of this country. For a hook of fish, or a knife, they gave us five or six hens; for a comb, a pair of geese; for a mirror, or a shear, so much fish that it would have sufficed for ten men; for a bell or a string, a basket of sweet potatoes* [batate]*, these are some roots as long as the radish and have the taste of chestnut».*

[59] Some famous 16th century voyages are those of García Jofré de Loisa in 1525, Ruy López de Villalobos in 1542, Miguel López de Legazpi in 1565 and Andrés de Urdaneta in 1565.

(Álvaro de Saavedra, 1527 and perhaps Juan Gaetano in 1555)[60]; the Caroline Islands (Álvaro de Saavedra, in 1528 and later explorers in 1543, 1545 and 1565); the Galapagos Islands (Fray Tomás de Berlanga in 1536); the Solomon Islands (Álvaro de Mendaña, in 1567 and 1595 and Hernan Gallego in 1568[61]); New Guinea and the Society Islands (Pedro Fernández de Quirós, in 1606), etc. Finally, López de Legazpi founded the first Spanish settlement in the Philippines in 1565.

Sweet potatoes were certainly known in South East China by the end of the 16th century, when it had a role in preventing famine, although there is evidence that it could have been introduced to the western provinces decades before (Ho, 1955) [62]. In any case, authors agree that the sweet potato arrived in China and Japan only after Spanish travellers got to the Philippines in 1521[63] (Pigafetta, 1524). The absence of sweet potato in commercial exchanges between China, Japan (where the sweet potato arrived in 1627 (Simon, 1914)) and the Philippines before that date further supports an introduction of the sweet potato to the region by Spanish travellers. Even more, Rumphius (1747) mentions the term *ubi kastela* (Castilian yam) to refer to the sweet potato in Ambon Island (currently Indonesia) and Ho (1955) also mentions several variations of "foreign tuber" across South East China: linguistics, in this case, reflects an early introduction of the sweet potato to the region by the Spanish.

During Captain Cook's trip on the *Endeavour*, Joseph Banks and Daniel Solander collected multiple plant specimens. One of them, collected in the Society Islands in 1769,

[60] Álvaro de Saavedra sighted Hawai'i on November 28th, 1527. One day later, two ships under his command went lost, and several authors, from different sources of evidence, suggest that they arrived in Hawai'i (see for example Dixon, 1932). Gaetano's contact with Hawai'i is still debated.

[61] With Gallego's log book lost, the only reference available is the translated version by Guppy (1887).

[62] «*As early as 1563 the western prefecture Ta-li, near Burma, already recorded the sweet potato*» (Ho, 1955).

[63] Magellan arrived in Homonhon on March 16th, 1521, during the first circumnavigation of the globe.

was a sweet potato. This collection, together with the log books of Cook's trip, led naturalists to propose a pre-European transportation of the crop to Polynesia. Although presented by many as the first contact between Europeans and Polynesians (see for example Handy, 1940; Hather and Kirch, 1991; Roullier et al., 2013c), the timeline presented here makes clear that contacts between Europeans (mainly Spanish and Portuguese) and Polynesians started a long time before Captain Cook's voyage. Those early contacts by Spanish and Portuguese travellers probably had a role in the spread of the sweet potato across the Pacific and between islands, and therefore presenting information collected in the eighteen century as if no previous contacts existed is likely to be misleading. Furthermore, the chronicles written by those early travellers are profuse in details and include explicit mentions to root crops such as taro, but systematically neglect the sweet potato (see detailed reports in Dixon, 1932; Hackney and Thomson, 1901). The lack of mention *per se* does not allow the rejection of the hypothesis of a sweet potato widespread in Polynesia; however, and considering the degree of detail provided in the accounts aforementioned, the lack of mention would indicate that the sweet potato, if present at all, was not widespread in the Polynesia at the time of the first contacts with Europeans.

5.1.3 | Seed production in sweet potato and genetic diversity in Pacific varieties

Almost 300 sweet potato varieties have been recorded in Polynesia (Yen, 1974: 41). This apparently large diversity has been presented by some authors as evidence of an ancient arrival of the crop in Polynesia, assuming that the plants do not produce seeds and are vegetatively propagated (Heyerdahl, 1952; Sauer, 1950), thus requiring a long time in the islands to accumulate the diversity observed. However, despite the uncertainty generated in some publications (Clarke, 2009; Heyerdahl, 1952), the sweet potato does produce seeds —CIP for example hosts a large collection of sweet potato seeds. With reports

of natural seed production in America and also in the Pacific (see e.g. Austin, 1977; Guppy, 1906; Yen, 1960), there is no reason to believe it was different in the past. That, taken together with the fact that sweet potato varieties are cross-fertilized and produce highly heterozygous progeny (Katayama et al., 2017) could explain the existence of a large number of varieties in the Pacific that could have generated in a short timeframe. This explanation has been proposed for the genetic diversity existing in New Guinea sweet potato cultivars (almost 500 varieties recorded), a region where the crop was introduced by Europeans (Roullier et al., 2013b).

5.1.4 | Morphology of the storage root in sweet potato and related species

When describing sweet potato remains from archaeological sites, several authors explain that the remains they found can be assigned more or less confidently to the genus *Ipomoea* (e.g. Coil and Kirch, 2005; Yen, 1974). However, most authors take the identifications for granted and do not question the identity of the findings (see for example Ballard et al., 2005; Clarke, 2009). In contrast, it is not clear to me to what extent those fragments can be confidently assigned to a specific species; among other reasons because, as explained in the introduction to this chapter, most authors do not explain the criteria based upon which they identify their findings as sweet potato and even some of them admit the remains could well belong to other closely related taxa.

In the case of the remains found in Mangaia Island and Hawai'i (described in section 5.1.2.1.3), the most compelling among those from Polynesia, identification relied on the extensive expertise of the authors on the identification of root and tuber remains (Hather, 1988, 2000) and there is no reason to doubt that these remains belong, at least, to an *Ipomoea* species. However, the absence of any work comparing the roots of *Ipomoea batatas* to those of other species in *Ipomoea* is notable. In this sense, it is worth

mentioning a recent work by Eserman and colleagues (2018) in which they present a detailed study of the development of storage roots in *Ipomoea batatas*, *I. lindheimeri* and *Distimake dissectus* (Jacq.) A.R. Simões & Staples (a species belonging to Merremiae, the other tribe in the family Convolvulaceae). Although root anatomy is not our field of expertise, Figure 1 in Eserman's paper (reproduced in Figure 5.2 in this dissertation) shows a remarkable anatomical similarity between the roots of all three species; it would be interesting to test how similar these and other storage roots in *Ipomoea* are after carbonization and whether the charred remains of different species could be distinguished.

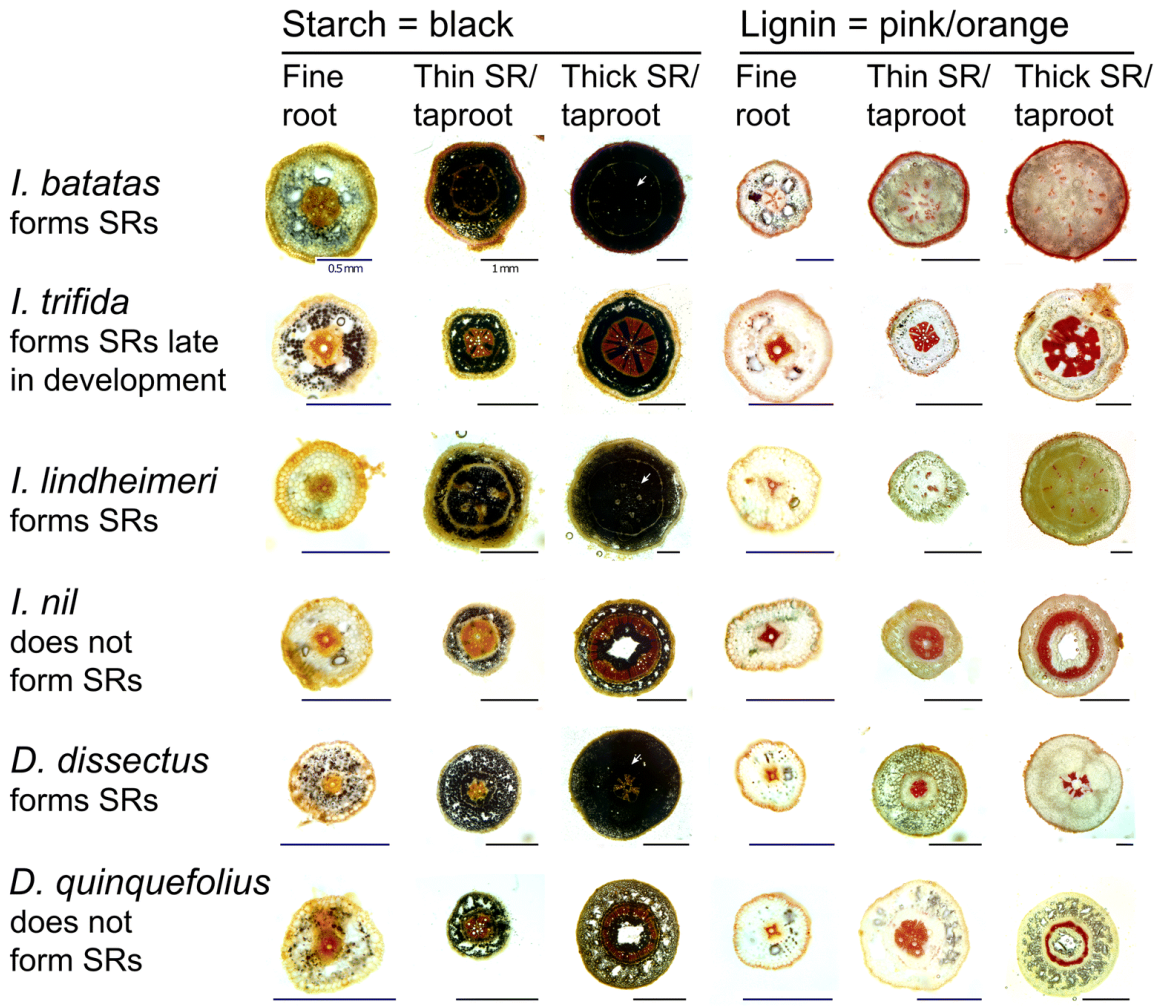


Figure 5.2. Root cross sections from three pairs of species of Convolvulaceae, where one member of the species pair forms storage roots and the other does not: *Ipomoea batatas*—*I. trifida*, *I. lindheimeri*—*I. nil* and *Distimake dissectus*—*D. quinquefolius*. Black scale bars are 1 mm and blue bars are 0.5 mm in length. White arrows indicate starch-storage tissue. Figure reproduced from Eserman et al. (2018).

5.1.5 | Other biological evidence of pre-European human contacts between America and Polynesia

Human contacts in pre-European times have been suggested by different authors as the mechanism by which several plants and animals made their way into the Pacific islands from America, or vice versa. In addition to sweet potatoes and cultural evidence, human genetics, bottle gourds, *mahoes*, chickens, pigs and even pathogens have been presented as biological evidence of these contacts with different degree of success. Doubts raised by different naturalists during the early 20th century called into question most of this alleged evidence. For example, the *maho* or *mahoe* (*Hibiscus tiliaceus* L., family Malvaceae), first thought to be a plant of American origin transported into Polynesia by humans (Cook and Cook, 1918), was later agreed to have a naturally occurring pantropical distribution (Merrill, 1920).

In addition, the results of recent molecular studies found no traces of genetic material of American origin in Polynesian samples or, when found, could not be conclusively dated to a pre-European contact and generated heated discussions. For example, the results of an analysis of ancient **chicken DNA**, reported as evidence of trans-oceanic contact, has been heatedly debated by supporters and opponents of the hypothesis of trans-oceanic contacts (Beavan, 2014; Bryant, 2014; Storey and Matisoo-Smith, 2014; Storey et al., 2013; Thomson et al., 2014; *et cetera*).

In the case of **human DNA**, the debate is also far from coming to an end. Although most scholars nowadays agree that the first **human settlers** in Polynesia arrived from Southeast Asia through Melanesia (see e.g. Clarke, 2009; Friedlaender et al., 2008; Jones, 2011; Thorsby, 2012) some authors, for example Thor Heyerdahl (1952), have proposed that the eastern Polynesian islands were first colonised by Americans. Genetic studies of human populations from Polynesia and New Zealand have failed to identify traces of

American ancestry in all islands except Easter Island[64] (Gosling and Matisoo-Smith, 2018; Hudjashov et al., 2018; Knapp et al., 2012; Wollstein et al., 2010). In 2012, Thorsby (Thorsby, 2012) identified two Amerindian Human Leukocyte Antigens (HLA) haplotypes in Rapanui individuals[65] and further support was provided two years later, in a genome-wide study that identified an approximate 8% American ancestry in Rapanui people (Moreno-Mayar et al., 2014). However, a recent analysis of the mitochondrial and nuclear genomes of five human individuals dating pre- and post-European contact casts doubts on previous results (Fehren-Schmitz et al., 2017). Fehren-Schmitz and collaborators found no evidence of American ancestry and argued that introductions of already admixed individuals in historical times could lead to a misinterpretation of the results in previous studies[66].

Finally, the **bottle gourd** (*Lagenaria siceraria* (Molina) Standl., family Cucurbitaceae), for which an African origin is generally accepted (Candolle, 1883; Heiser, 1979; Richardson, 1972), was first thought to be a post-European introduction to the New World and Polynesia and later suggested to be the result of trans-oceanic contacts in pre-European times. This human-mediated transfer from America has been vehemently sustained

[64] A study published in 1995 identified two individuals (among a sample of 1,178 living people), one from Tahiti and one from the Cook Islands, carrying a mitochondrial haplotype associated in the first instance to an exclusively-American lineage (Sykes et al., 1995). However, that finding was soon contested by the argument that the close relationship of Polynesian and American haplotypes was due to a common Asian origin and not to admixture (Bonatto et al., 1996). Although Bonatto's reply was also criticised (Cann and Lum, 1996), seven years later Hurles and colleagues (2003) showed that the Amerindian traces in Polynesian individuals can be explained as the result of the modern slave trade. The same explanation has been given to the Amerindian and European ancestry found by Lie and colleagues (2007) in modern Rapanui individuals and dated to the early 19th century.

[65] Thorsby (2012) also looked for similar patterns in mitochondrial DNA and Y chromosome markers, but traces of American ancestry were absent. Negative results were also retrieved in previous studies using HLA and mtDNA samples (González-Pérez et al., 2006; Hagelberg et al., 1994; Lie et al., 2007).

[66] This seems to be the case also for the alleged identification of Polynesian ancestry in two ancient Brazilian genomes (Malaspinas et al., 2014).

by Green (2000, 2005), although evidence for such introduction relies mainly on linguistic grounds and on a hypothesised connection with a sweet potato transfer, a connection for which no evidence has been presented. In contrast, the possibility of a naturally-driven arrival of the plant to the Americas and to Polynesia received strong support from a practical essay in which Whitaker and Carter (1954) showed that bottle gourd fruits are capable of floating for over 200 days in oceanic water and that the seeds inside remain viable. Therefore, the worldwide distribution of the bottle gourd, including its presence in Polynesia, seems to be a consequence of dispersal by sea currents (Kistler et al., 2014; Whistler, 1990).

In summary, the possible occurrence of pre-European contacts between America and Polynesia generates heated discussions that are far from ending. In this context, and especially now that all other biological evidence has been called into question, it is easy to understand why the possible human transportation of the sweet potato to Polynesia has attracted broad attention in recent times, as it appears to be the only remaining biological evidence for these alleged pre-Columbian contacts.

5.2 | RESULTS AND DISCUSSION

Our approach to the questions about the geographical origin of the sweet potato and its diffusion to Polynesia is not restricted to analyses of the crop. Instead, we take a global approach in which these issues are illuminated by our broader studies of the genus *Ipomoea*. The results presented in this chapter raise further doubts about those alleged pre-European contacts across the Pacific and support a different scenario that takes into account that both storage roots and long-distance dispersal are commonplace in the group (sections 3.2.4 and 3.2.5). Specifically, our results suggest that the presence of sweet potato in Polynesia in ancient times, if confirmed, could be more convincingly explained as

the result of long-distance dispersal by natural means, rather than by human contacts in pre-European times.

5.2.1 | Geographical origin of the sweet potato

I have shown that *Ipomoea batatas* had a single origin from its common ancestor with *I. trifida*. The distribution of *I. trifida*, from southern Mexico to northern South America, is well defined (Figure 5.3). Also, all annual species in the Batatas group, which form a clade, are sister to the sweet potato clade and restricted to the Americas in their natural distribution. This, together with the other multiple lines of evidence presented before, confirms the American origin of *Ipomoea batatas*, most likely somewhere in the distribution area of *I. trifida*, between southern Mexico and northern South America.

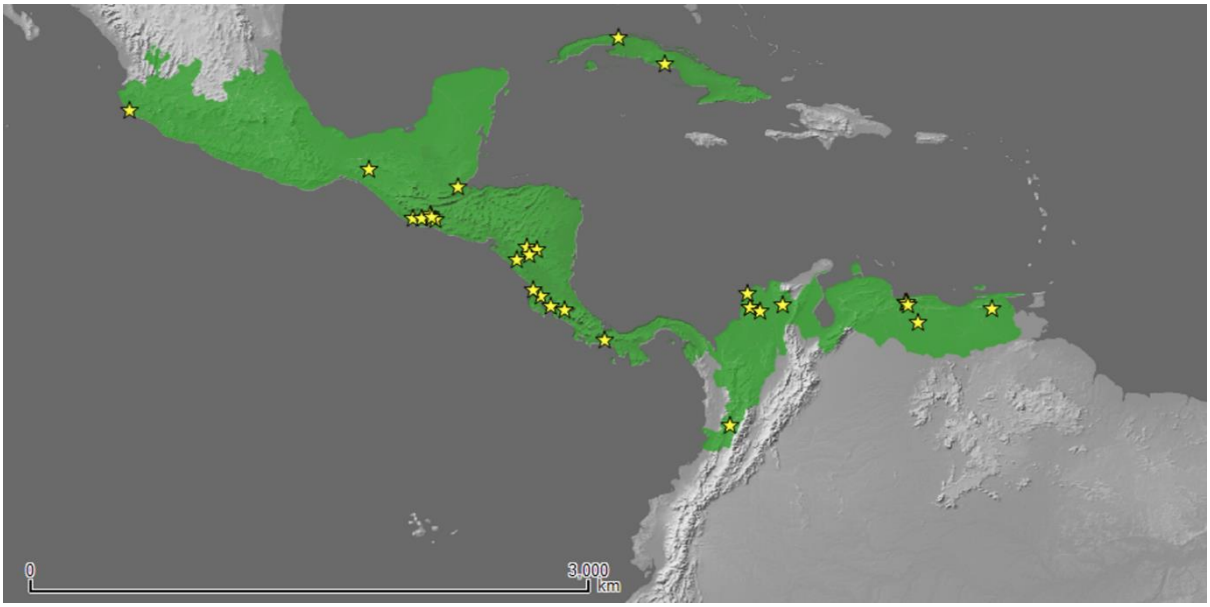


Figure 5.3. Distribution map of all *Ipomoea trifida* samples included in our molecular studies (yellow stars). The green area on the map represents the global distribution of the species, drawn from the full set of herbarium specimens that we have studied.

This is a large area and includes many different habitats, so the specific place of origin of the species remains unknown. One of the problems affecting the investigation of the geographical origin of the crop is that, despite many claims, wild sweet potatoes have never been collected. Several authors collected plants growing in the wild and tentatively identified them as wild sweet potatoes (Austin, 1988b; Bohac et al., 1993; Jones,

1967; Kobayashi, 1984), but those specimens have been subsequently re-identified either as one of its wild relatives or as feral sweet potato, given their proximity to human habitation.

In the absence of specimens that represent wild forms of the sweet potato, some authors attempted to infer the place of origin of the sweet potato by examining current levels of diversity within the crop. This approach, mainly taken in a series of papers published by Roullier and colleagues (2011, 2013a) and Clarke (2009) is inherently problematic. This is because we cannot be sure that the place of collection of the samples corresponds to the place of origin of the varieties or landraces that they represent, or whether they originated elsewhere and were later transported into the area. In addition, because we do not know their history, we do not know whether they are *pure* sweet potato lineages or the result of modern crosses between different varieties. Geographical patterns inferred from the place of collection of modern material, such as those inferred by Roullier and colleagues and Clarke, may be misleading. For example, Roullier and colleagues (2011, 2013a) argued that the two sweet potato chloroplast lineages are restricted geographically, whereas we showed that, although each of the clades is dominated by accessions from Central and South America respectively, there are also samples from Central America in the South American clade and vice versa (Figure 4.19). Furthermore, no clear geographical structure can be inferred within each chloroplast lineage.

In this context, and until truly wild sweet potatoes are found, studies to infer the geographical origin of *Ipomoea batatas* rely on indirect evidence. An alternative approach, although probably imperfect, is to consider the distribution of *Ipomoea trifida*, the closest relative of the sweet potato, to inform studies on the origin of the crop. Given that there are no records of *I. trifida* being used or transported by humans, it is reasonable

to assume that its current geographical distribution, certainly narrow compared to most other species in the Batatas group, is not affected by human transportation, but rather follows natural patterns of distribution.

As highlighted elsewhere in the text, one of the strengths of our project compared to previous studies is the large amount of molecular data generated. In the case of *Ipomoea trifida*, we have genomic-scale data for 23 specimens spanning its entire distribution except Mexico[67]. The phylogeny inferred using nuclear data resolved an *I. trifida* clade in which specimens collected in Colombia and Venezuela are nested within a clade of Central American and Caribbean specimens (Figure H in Appendix 3). A possible explanation, congruent with this topology and with the timeframe inferred for the split of these two species (Figure 4.8), would be that *Ipomoea trifida* originated in Central America and then spread towards the south. In this context, it is plausible to consider Central America as the most likely place of origin of *Ipomoea batatas*.

5.2.2 | Sweet potato in Polynesia

The arrival of sweet potato in Polynesia has been a matter of discussion for two centuries, although in recent times it has been generally assumed that its presence in Polynesia is the result of three human introductions, referred to as the "tri-partite hypothesis" (Barrau, 1957; Denham, 2013; Roullier et al., 2013c; Yen, 1971). As explained before, the two European introductions are well-documented, but evidence for a pre-European human-mediated introduction is not conclusive. In this section, I consider two alternative possibilities that have received little attention in previous studies but could explain the early presence of the sweet potato reported in the Pacific: the possibility that some or

[67] In a first approach, I tried to incorporate the genomic data of a Mexican specimen (PI 618966) identified as *Ipomoea trifida* by Eserman and collaborators (Eserman et al., 2014) and included in their phylogenetic studies of the morning glories. However, our results identified this specimen as a different species outside the *I. batatas*/*I. trifida* clade.

most *Ipomoea* remains associated with the sweet potato belonged to other species, and the possibility of a long-distance dispersal of *I. batatas* across the Pacific by natural means (Montenegro et al., 2008; Zhang et al., 2004).

5.2.2.1 —Ipomoea batatas *or another species?*

Ladefoged and colleagues (2005) suggested the possibility that Polynesian inhabitants consumed other species of *Ipomoea* with edible storage roots. As explained before, it is possible that remains of these other species have been subsequently identified by archaeologists as sweet potatoes, without further discussion. The possibility of the remains belonging to other species was also unintentionally suggested by Handy:

> *«In Kau, Hawaii, it is said that Hawaiians used to dig up in the forests, from great depths what was called the Paha or Koali uala (sweet potato morning-glory), which had a heart-shaped leaf like the wild morning glory but tubers like the sweet potato.* […] *E.H. Bryan, Jr., tells me that he has seen what was evidently a wild potato* [sic] *of this description growing on the island of Kahoolawe. On Easter Maui it is said that the variety named Kupa grows wild in the uplands; and vines, which may be Kupa, growing wild above Nahiku, have been described to me. In Kaupo I was told that the variety named Aehaukae is actually a wild potato* [sic]*, which was found in many localities before the days of ranching.»*
>
> (Handy, 1940)

I think this possibility deserves consideration, especially as more and more species of Ipomoea are found to develop storage roots and are recorded as food, as I showed in Chapter 3. At least 24 species of *Ipomoea* have been recorded for the Pacific Islands. I compiled the list of species in Table 5.1 from floras and floristic treatments of the Pacific Islands; this list is not comprehensive and a systematic search would possibly add more species to the list. Nevertheless, this list includes at least four species of *Ipomoea* that produce edible storage roots and are present and utilised in the region. To these, it would be necessary to add other elements in the family Convolvulaceae found in the islands and

recorded as edible elsewhere (Austin, 2007b; Mueller-Dombois and Fosberg, 1998; Staples, 2009, 2010). The existence of other species of morning glories with storage roots in the islands, together with the fact that storage root remains seem to be difficult, if not impossible, to identify at the species level (section 5.1.4), calls into question the identification as *I. batatas* of all rooteous remains found in the Pacific.

Table 5.1. Species of *Ipomoea* recorded from the Pacific Islands. Names in bold indicate species with storage roots used as food.

Species	Reference*
Ipomoea alba	1, 4, 5, 13
Ipomoea aquatica	7, 10, 14
Ipomoea batatas	2, 4, 5, 7, 9, 10, 11, 12, 13, 14
Ipomoea cairica	4, 6, 10, 13
Ipomoea carnea subsp. *fistulosa*	7, 10, 14
Ipomoea coccinea†	1
Ipomoea fimbriosepala	10
Ipomoea hederifolia	7, 8, 10
Ipomoea horsfalliae	7
Ipomoea imperati (= *I. stolonifera*)	13
Ipomoea indica (= *I. congesta*)	1, 4, 6, 10, 12, 13, 14
Ipomoea littoralis	1, 2, 5, 7, 13, 14
Ipomoea mauritiana	7, 10
Ipomoea obscura	5, 7, 10
Ipomoea pes-caprae subsp. *brasiliensis*	2, 4, 5, 8, 9, 12, 13, 14
Ipomoea quamoclit	5, 7, 10, 14
Ipomoea marginata (= *I. sepiaria*)	7
Ipomoea purpurea	4
Ipomoea sepiaria	7
Ipomoea tiliacea†	5
Ipomoea tiliifolia (= *S. tiliifolia*)	2, 5
Ipomoea triloba	7
Ipomoea violacea (= *I. macrantha*)	2, 3, 5, 7, 11, 13, 14

† Species cited from the region that need confirmation.
*1, Hemsley, 1894; 2, Cheeseman, 1903; 3, St. John and Mason, 1953; 4, Allan et al., 1961 5, Sachet, 1975; 6, John, 1976; 7, Fosberg and Sachet, 1977; 8, Sykes, 1981; 9, Zizka, 1991; 10, Smith, 1991; 11, Florence et al., 1995; 12, Florence and Lorence, 1997; 13, Mueller-Dombois and Fosberg, 1998; 14, Franklin et al., 2008.

5.2.2.2 — Long-distance dispersal to Polynesia

All species in the Batatas group, except *Ipomoea littoralis*, are restricted to the Americas. *Ipomoea littoralis* is a species distributed from Polynesia to Madagascar and absent from the American continent (Austin, 1991; Khoury et al., 2015). This Old-World species diverged from its sister species, *I. lactifera*, at least one million years ago (Figure

4.8), strongly suggesting that the distribution of *I. littoralis* is best explained by natural dispersal of an ancestor across the Pacific, followed by subsequent evolution into a distinct species. Although as far as I know the buoyancy of its seeds has not been tested, it has been shown that the seeds of several other *Ipomoea* species that live in similar coastal environments can survive after floating long distances (see section 3.2.5). In addition, the wide distribution of other coastal species, such as *I. pes-caprae*, throughout the tropics, has been explained as the result of long-distance dispersal mediated by sea currents (Miryeganeh et al., 2014). Taking this into account, it would be very difficult to explain the distribution of *I. littoralis* except in terms of long-distance, inter-continental dispersal across the Pacific Ocean.

Apart from *I. littoralis*, in section 3.2.5 I presented other examples of species with a disjunct distribution that would be best explained as a result of long-distance dispersal. Another example that I highlighted is *I. tuboides*, a Hawaiian endemic nested in a clade of Mexican species, more than 5,000 km away (Figure 5.4). This species represents another case of trans-continental diffusion and within the same geographical region as *I. littoralis*. The time-calibrated phylogeny of the Tuboides group (Figure 5.5) shows that *I. tuboides* diverged from its sister species at least 1.1 million years ago, and the most likely explanation for its presence in Hawai'i is also naturally occurring long-distance dispersal.

These two examples demonstrate that species closely related to the sweet potato and with similar seed, fruit, and dispersal biology (Figure 5.6) are readily dispersed over very long distances. Furthermore, the significance of this highly iterative process is that disjunct distribution patterns of closely related species are not uncommon in *Ipomoea*, and therefore a species native to America but also present in Polynesia, be this a crop or not, would not be that surprising. Long-distance dispersal should therefore be considered

a plausible explanation of how the sweet potato came to be distributed in pre-Columbian Polynesia.

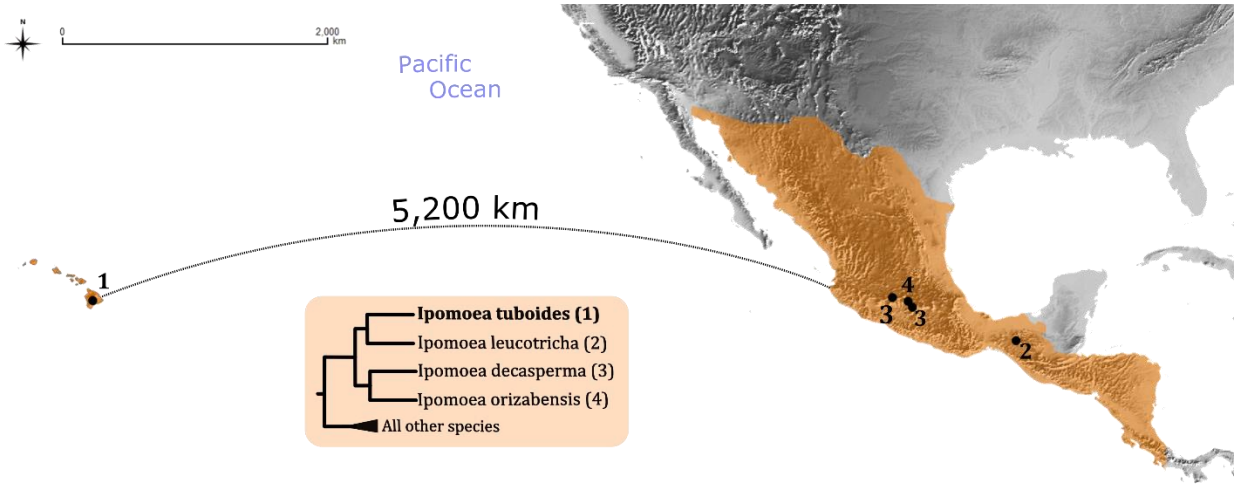


Figure 5.4. According to our phylogenetic studies, *Ipomoea tuboides* is a Hawaiian endemic nested in a clade of Mexican species, and its distribution is most likely the result of long-distance dispersal by natural means. Orange indicates the global distribution are of this group. Black dots indicate the place of collection of the specimens sequenced in this study, and numbers refer to the species in the nuclear phylogenetic tree (summarised in the orange box, all branches 100% support).

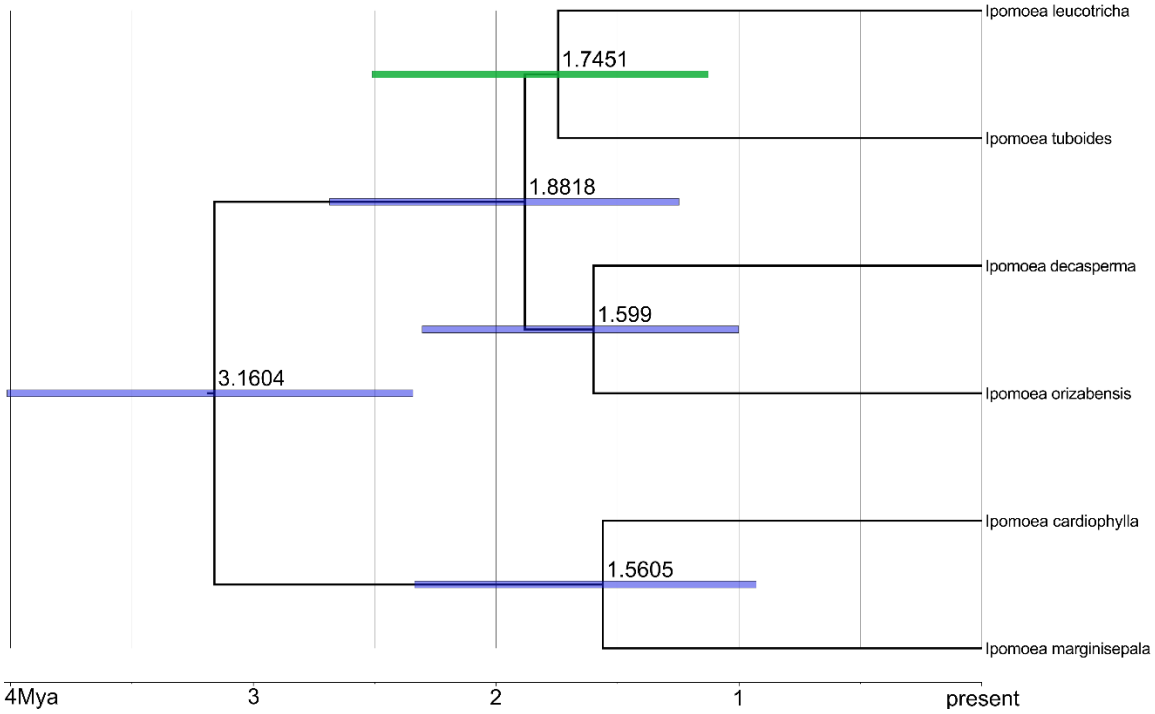


Figure 5.5. Time-calibrated phylogeny of the species in the Tuboides group inferred using 21 nuclear regions. Node bars represent 95% high posterior density intervals for node ages. The root age for these phylogenies is determined by the ages sampled for the clade in our graphical model constructed in RevBayes.

Figure 5.6. Seeds of *Ipomoea littoralis* (left; specimen *L.J. Brass 13940* [BM]) and *I. batatas* (right; *S9 55* [USDA]).

5.2.2.2.1 | Banks and Solander's collection

In addition to other sources of data, the study of specimens collected in Polynesia during the first European trips to the region, and in principle unaffected by modern exchanges of material, is of extraordinary interest (Roullier et al., 2013c). The most iconic of these ancient specimens was collected by Joseph Banks and Daniel Solander in the Society Islands, in 1769, during Captain Cook's expedition on the *Endeavour* (Figure 2.7). This specimen, the oldest sweet potato specimen collected in Polynesia, is often presented as proof of sweet potato presence in the region in ancient times (Roullier et al., 2013c). We successfully sequenced Banks and Solander's specimen using genome skimming and assembled its whole-chloroplast genome and fragments of the nuclear coding regions (see methodology in section 2.3.3.6). The assembled chloroplast genome of this specimen clearly places it within sweet potato CL1, the sweet potato chloroplast lineage that carries the ancestral chloroplast (Figure 5.7). This position for Banks and Solander's specimen was already inferred in a previous study using only a small chloroplast DNA

fragment (Roullier et al., 2013c). The agreement of our results with those in a previous study, together with the fact that other samples in our analyses have accumulated similar levels of mutations over time (Figure 5.7), and the lack of evidence of DNA damage (Supplementary Files 1 and 6), indicates that the probability of our results being affected by cross-contamination from other *Ipomoea* DNA is very small.

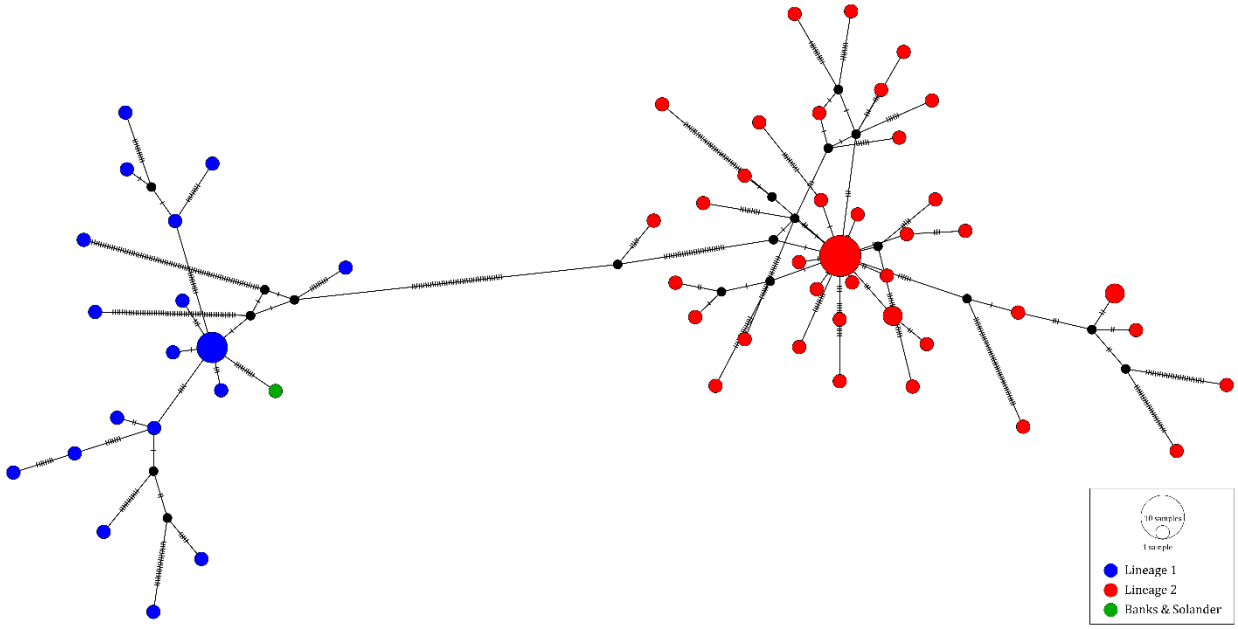


Figure 5.7. Median-Joining network inferred using whole chloroplast genomes of 72 *Ipomoea batatas* specimens. The specimen collected by Banks and Solander in the Society Islands (green circle) belongs to sweet potato chloroplast lineage 1 (blue circles), the lineage that carries the ancestral chloroplast. Red circles indicate specimens in sweet potato chloroplast lineage 2.

We subsequently aimed to investigate the time of divergence of the lineage represented by Banks and Solander's specimen from its closest relative in our phylogeny We designed a conventional time-calibrated phylogeny and a coalescent analysis including all whole chloroplast genomes of *I. batatas* and *I. trifida* (see section 2.8.1). Both analyses indicate that the lineage to which the Banks and Solander specimen belongs diverged from its closest relative in our phylogeny at least 111,500 years ago (at least 139,000 years ago in the coalescent analysis; Figure 5.8 and Figure I in Appendix 3). This does not mean that the sweet potato arrived in Polynesia more than 100,000 years ago, but simply that the Banks and Solander specimen and other specimens in our study show

levels of genetic variation that place them in a much older timeframe of evolution of the species than if the storage root was a product of human domestication and brought into Polynesia in the last 1,000 years.

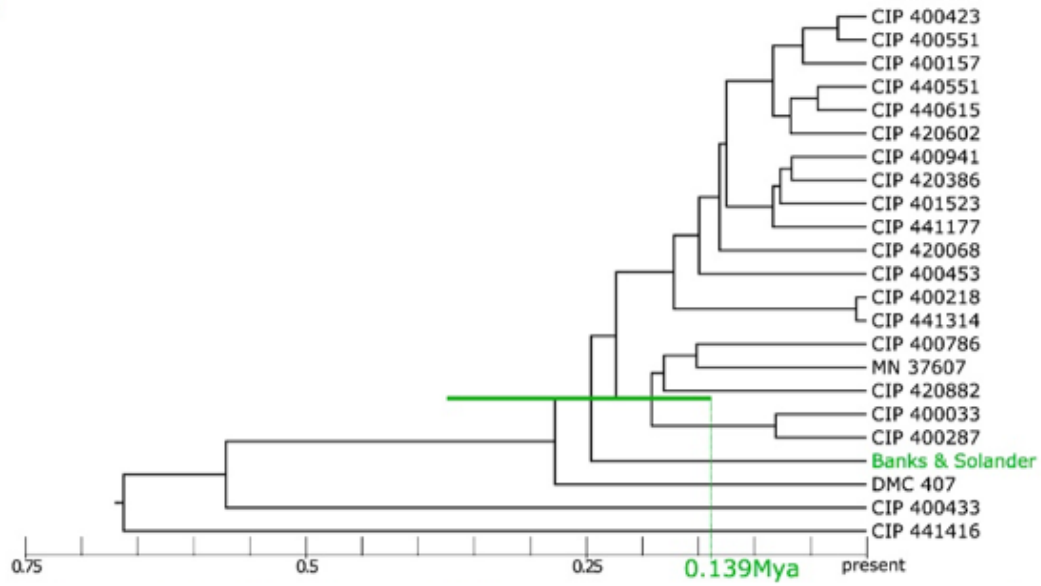


Figure 5.8. Time-calibrated phylogeny of *Ipomoea batatas* chloroplast lineage 1 using the coalescence method and showing that Banks and Solander specimen diverged from its closest relative at least 139,000 years ago (111,500 years ago using a conventional divergence time estimation method). The green bar represents 95% HPD intervals for node ages.

This result, together with the distinct admixture pattern inferred from the nuclear sequences (Figure 5.9), is congruent with the long-term isolation of this distinct variety in comparison to varieties from Central and South America. This long-term isolation is compatible with an arrival of the species to the Pacific by long-distance dispersal. Alternatively, the Banks and Solander specimen could represent an American lineage that was transported to the Pacific in more recent times but subsequently disappeared from the American continent. However, considering the high levels of genetic admixture in modern sweet potato varieties (Figure 4.18A) this possibility seems less likely.

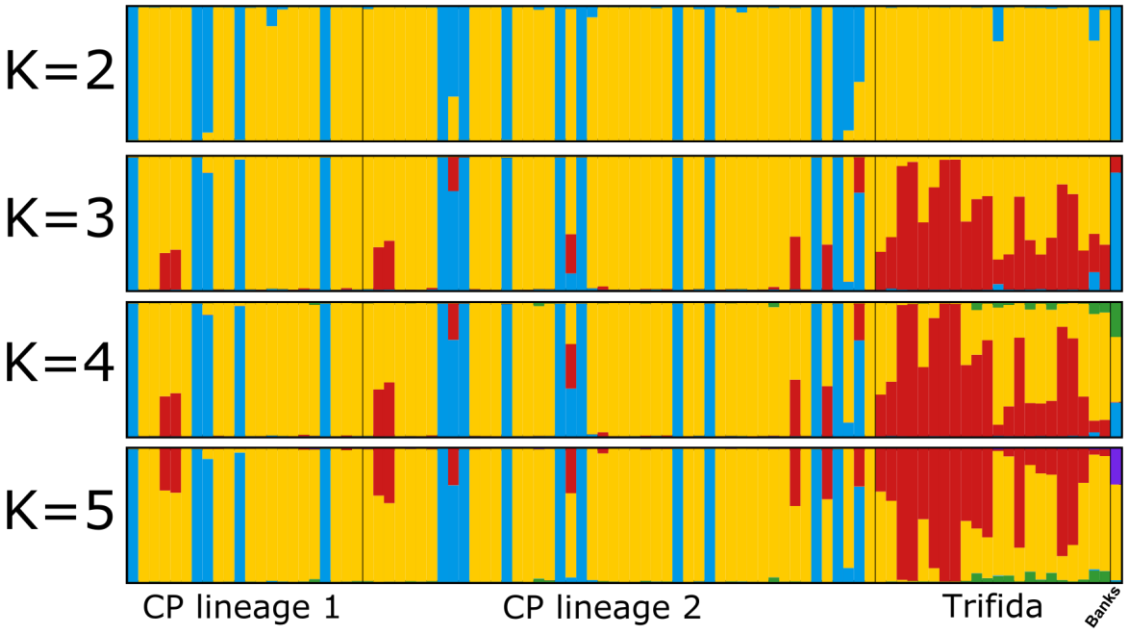


Figure 5.9. Population structure analysis of all *Ipomoea batatas* and *I. trifida*, including the specimen collected by Banks and Solander, inferred using 5,735 variable positions from the nuclear alignments (150,000 MCMC replications and 100,000 burn-in repetitions. Admixture model assuming independent allele frequencies among populations [λ = 0.3483; K = 1–5; 3 runs]). The distinct signature in Banks and Solander's specimen is congruent with the long-term isolation of that lineage in Polynesia.

5.3 | CONCLUSIONS

Existing evidence strongly supports an American origin of *Ipomoea batatas*, although which part in the Americas remains unknown. Two main problems hinder the identification of the geographical origin of the species: 1) the absence of *I. batatas* specimens of wild provenance and 2) the potential problems of inferring the origin of the species based on the distribution of modern cultivated specimens, the provenance of which is frequently unknown. Until wild *I. batatas* specimens are found, an alternative approach is to take into account the distribution range of *I. trifida*, the sweet potato's closest wild relative. Given that I have shown *I. trifida* most likely evolved in Central America, it is reasonable to think that *I. batatas*, which I have shown had a single origin, also originated in Central America. This conclusion is, of course, preliminary, and needs to be tested in future studies.

Studies on the early dispersal of the sweet potato must take into account several factors. First of all, there is no evidence that the sweet potato, if present in Polynesia in ancient times, was as widespread as frequently suggested. This allegedly widespread distribution was mostly based on reports by 18th and 19th century explorers who explored the Pacific and were erroneously considered, by many authors, as the first Europeans to contact Polynesia. From what has been explained here, it seems clear that Spanish voyagers discovered and contacted many of the islands two centuries before the great age of exploration. Those early Spanish voyagers (and possibly islanders travelling between islands too) had an important role in the diffusion of the plant between islands from as soon as the early 16th century, hence explaining —at least partially— the numerous mentions by 18th and 19th century explorers.

Secondly, rooteous remains found in the Pacific have almost always been identified as sweet potatoes, but the possibility that some or most of these remains belong to other species exists. A re-evaluation of sweet potato archaeological remains, illuminated by comparative studies on the anatomy of *I. batatas* roots and those of other species of *Ipomoea*, may provide important information in this regard.

Finally, if sweet potato truly arrived in Polynesia in pre-European times, there is compelling evidence to suggest that it did so by natural dispersal, rather than by deliberate human transport. This possibility is supported by multiple lines of evidence:

1) Multiple examples of disjunct distributions in *Ipomoea* that are most easily explained as the result of long-distance dispersal by natural means, including two Pacific species nested in clades of American species, one of them a sweet potato CWR.

2) Genetic signature observed in the oldest specimen collected in Polynesia is congruent with this lineage being isolated for a long time.

3) The fact that *Ipomoea batatas* originated more than 800,000 years ago, which provides a much older timeframe within which to postulate a dispersal across the Pacific, compared with that of sweet potato domestication within the last 8,000 years and it being taken to Polynesia by humans in the last millennium.

4) Modelling experiments showing that dispersal of the sweet potato across the Pacific is possible (Montenegro et al., 2008).

Whether pre-European contacts between Americans and Polynesians, deliberate or not, existed, is a question that may never be satisfactorily answered. The possibility that the sweet potato was transported to Polynesia by pre-European travellers is, however, more difficult to defend. Although the results presented here do not preclude the arrival of the sweet potato in Polynesia via human-transport in the last thousand years, they do present long-distance dispersal by natural means as a more likely explanation. Our results call into question the predominant theories and show that there is no need to claim a human-mediated transport of the crop to Polynesia in pre-European times.

6 | SUMMARY AND FINAL CONCLUSIONS

Despite being one of the most important crops in the world, there was much uncertainty surrounding the origin of the sweet potato when I started my D.Phil. Some unanswered questions were: Did the sweet potato have single or multiple origins? Did it evolve from a single ancestor or was it of hybrid origin? What wild species were involved in its origin? When did sweet potato originate? A prerequisite for answering these questions is to understand the evolutionary relationship between the sweet potato and its most closely related species, the Crop Wild Relatives (CWR). Furthermore, understanding this relationship is essential for food security, as CWR constitute potential sources of genetic variation for crop improvement.

Multiple studies in the last decades have tried to answer the questions outlined above. However, they provided only incomplete, often contradictory results, and the relationship between the sweet potato and its CWR remained poorly understood. Ultimately, this lack of knowledge hindered the use of wild species in breeding programmes, and thus sweet potato improvement has been restricted to using the genetic diversity held within the crop.

The aim of my DPhil research was to conduct a comprehensive phylogenetic study of the sweet potato and its wild relatives and address the unresolved questions pertaining to the origin of the crop. In addition, I also aimed to explore the evolutionary relationships within the megadiverse genus *Ipomoea* —to which the sweet potato belongs— which, at the beginning of this project, were also poorly known.

To overcome previous limitations, we designed our sampling to be as complete as possible: we obtained the whole chloroplast genome and 605 nuclear genes from 384 specimens representing around two hundred species in *Ipomoea*. Our sampling aimed to cover as much diversity within the genus as possible, while at the same time to include a

good coverage of the sweet potato and all its CWR. The phylogenies inferred using genomic data were complemented by other phylogenies inferred using DNA barcodes with a much larger taxon sampling (1,560 sequences in total).

This D.Phil. study contributes to knowledge of the genus *Ipomoea*, especially of the American species and the species closely related to the sweet potato. The phylogenies presented here, inferred using extensive taxon and character sampling, provide a robust phylogenetic framework with which to explore the evolutionary relationships between taxa in the group, as well as to provide answers to most questions outlined at the beginning of this project. The main conclusions of this study are:

1) *Ipomoea* is the largest genus in the tribe *Ipomoeeae*. All smaller genera in the tribe, previously considered as separate, are **nested within *Ipomoea*** and none except *Astripomoea* are monophyletic. Based on this and in conformity with the principles of monophyly and diagnosability, we propose that **all smaller genera should be sunk into *Ipomoea*, thus making *Ipomoea* monophyletic**. In addition, most previously defined infrageneric ranks (subgenera, series and sections) are not monophyletic and their use should be discontinued.

2) *Ipomoea* splits into **two main clades**: one clade including a third of the species in the genus is mostly restricted to the Old World, while a second clade comprising two thirds of the species consists of a grade of African species and a species-rich clade dominated by species from the Americas.

   Several clades are recognised and strongly supported in all phylogenies, regardless of the method of phylogenetic inference or of the data set. Some of these clades have diagnostic morphological characters (for example, species with coriaceous sepals or species with a tree habit).

3) *Ipomoea* probably **originated in the Palaeotropics**.

4) We identified multiple examples of disjunct distribution across very long distances in *Ipomoea*, even across oceans and between continents. These are most easily explained as the result of **long-distance dispersal** by natural means.

5) **Storage roots evolved multiple times independently** in *Ipomoea*. Species with storage roots are often most closely related to species without them, and species with edible storage roots are found in all continents.

Previous studies on the origin and evolution of the sweet potato have been affected by the continuous misinterpretation of two concepts, **origin** and **domestication**: many authors referred to the *origin* of the sweet potato as the time when it was taken into cultivation, both terms being used interchangeably (see Clarke, 2009; Denham, 2013; Dixon, 1932; Heyerdahl, 1952; Roullier et al., 2013a). In fact, sweet potato remains in the archaeological record have always been assumed to be evidence of active cultivation (see for example Horrocks et al., 2004a). For the first time in this thesis, I presented fully resolved phylogenies of the species closely related to the sweet potato. This permits a thorough investigation of the relationships within the group and of the origin of the sweet potato, as well as enabling us to differentiate between the origin and domestication of this important crop. The main conclusions are:

1) *Ipomoea batatas* forms a clade with fourteen other species that are the **crop wild relatives**. Only nine species in this group are confirmed as monophyletic: *I. batatas*, *I. lacunosa*, *I. lactifera*, *I. littoralis*, *I. ramosissima I. splendor-sylvae*, *I. tiliacea*, *I. triloba* and *I. trifida*.

The name *I. cordatotriloba* represents two entities with different evolutionary histories in North and South America. The North American entity is monophyletic

and should be recognised as a good species. In contrast, the South American entity forms a clade with *I. cynanchifolia* and *I. grandifolia* in which none of them is monophyletic. Also, *I. leucantha* is polyphyletic and the monophyly of *I. tenuissima* could not be assessed. Further studies are necessary to clarify the delimitation of these entities.

2) ***Ipomoea tabascana*** is a modern hybrid between *I. batatas* and *I. trifida*.

3) *Ipomoea batatas* had a **single origin by autopolyploidy**.

4) ***Ipomoea trifida***, a circum-Caribbean species, is sweet potato's closest wild relative and the only extant species that had a role in the evolution of the crop. All other species previously proposed as progenitors of the crop are more distantly related.

5) *Ipomoea batatas* **originated well before modern humans**, at least 800,000 years ago. This, together with the multiple origin of storage roots in *Ipomoea*, strongly suggests that **humans found and cultivated sweet potato plants that already had storage roots**. Therefore, the sweet potato storage root is not likely a trait of domestication, but rather a pre-adaptation that predisposed this taxon to cultivation.

6) The phylogeny inferred using whole chloroplast genome data identifies two distinct *I. batatas* lineages, one of which is more closely related to *I. trifida* than to the other sweet potato lineage. These distinct chloroplast lineages were previously interpreted as evidence of multiple origins of *I. batatas*. However, in the light of the results of our analyses using nuclear data, we suggest that the conflicting pattern observed in the chloroplast phylogeny is the result of an **ancient hybridization event** between *I. batatas* and *I. trifida* following species divergence.

The progeny of this hybridization event **captured the chloroplast** from *I. trifida*, hence explaining the conflicting topologies. This event occurred **within 56,000 years of species divergence** from their common ancestor, and additional studies are necessary to investigate the mechanism by which the hybrid progeny of *I. batatas* and *I. trifida* became hexaploid again.

7) **Beauregard** and **Tanzania**, two sweet potato commercial varieties used in modern breeding research, belong to the sweet potato chloroplast lineage that carries the "ancestral" sweet potato chloroplast. This lineage holds **more genetic diversity** than the lineage that captured the chloroplast from *Ipomoea trifida*.

8) All but one of the sweet potato wild relatives are restricted to the Americas in their natural distribution ranges. This, together with the fact that archaeological remains found in the Americas are significantly older than anywhere else, further supports the hypothesis that **the crop is of American origin**.

9) **Archaeological evidence** of sweet potato presence in the American continent is fragmentary and less abundant than previously assumed. Many references to sweet potato in archaeological remains lack supporting evidence or information that allows an objective assessment of the findings. It is possible, for example, that the oldest remains linked to the sweet potato, found by Engel in Peru and dated from 8,000 years ago, were wrongly dated.

In this situation, the oldest remains undoubtedly linked to the sweet potato in America date from around **3500 years ago**. The lack of reliable evidence could indicate that *Ipomoea batatas* was not as widespread in America in ancient times as previously thought, although future archaeological findings and additional interdisciplinary studies will help clarify this question.

10) *Ipomoea trifida*, the sweet potato's closest relative, is restricted to the region between southern Mexico and northern South America. Our phylogenies suggest that this species originated in Central America, which possibly indicates that Central America is also the place of origin of *Ipomoea batatas*. This conclusion is, however, preliminary, and future studies including wild forms of the sweet potato are necessary to confirm or reject it.

In summary, *Ipomoea batatas*, the sweet potato, *originated* well before humans became aware of its existence and therefore had to thrive by itself, as a wild plant, for a very long time before being taken into *cultivation*. Further, considering this ancient origin of the species and the fact that storage roots are commonplace in the genus *Ipomoea*, as I have demonstrated, it is necessary to call into question some aspects of sweet potato research that have been taken for granted in previous studies. The results presented here provide a new framework with which to address several questions pertaining to the origin and cultivation of *Ipomoea batatas*, for example, what changes did cultivation produce in the crop, if any? Is sweet potato a domesticated crop and, if so, was it domesticated once or multiple times? If sweet potato has been domesticated, did the domestication process cause a loss of genetic diversity —the "bottlenecks" of genetic diversity traditionally associated with plant cultivation (Dempewolf et al., 2017)— or does the genetic diversity within the crop remain intact? Answering these questions will only be possible by comparing cultivated sweet potatoes with wild forms of the species, whose identity remains to be confirmed.

Finally, as part of this study I also aimed to explore a question that has been of interest for over two centuries: how did the sweet potato, a crop of American origin, come to be widespread in Polynesia before the arrival of the Europeans? This question has been

a major source of uncertainty. The predominant hypothesis is that it was transported by humans in pre-European times, implying the existence of human contacts across the Pacific previous to the European Age of Exploration. Answering this question required investigation of some aspects that were taken for granted in the 19th and early 20th centuries and have never been revisited since, the main one being the assumption that the sweet potato was in fact widely distributed before the arrival of Europeans. The main findings in relation to this question were:

1) Based on existing archaeological and historic evidence, and contrary to what is frequently assumed, it can be concluded that **the sweet potato was not widespread in Polynesia, if present at all, by the time Europeans first arrived**. Only the remains from two archaeological sites, in Hawai'i and the Cook Islands, can be assigned with some confidence to a pre-European origin. If the identification of those remains is correct, it would certainly be difficult to deny that the sweet potato was present in Polynesia before Europeans arrived. However, even if this were the case there is no evidence to support a widespread distribution of the crop in Polynesia in the pre-European era.

2) Europeans, mainly Spanish travellers from the early 16th century, as well as islanders travelling between islands, likely had a role in the dispersal of the sweet potato throughout the Pacific in historic times, explaining most references to the crop by 18th and 19th century European travellers.

3) The roots of *Ipomoea batatas* are very similar morphologically to those of other species in the genus. This, together with the fact that at least a dozen species of *Ipomoea* have edible roots similar to those of *I. batatas* —and many more species produce storage roots—, raises doubts about the accuracy in the identification of

charred remains as sweet potatoes. Although it may seem unlikely in the case of *I. batatas*, it is possible that at least some of these ancient remains belonged to **a wild species other than the sweet potato, or to the sweet potato itself but before cultivation**. This possibility must be taken into account, especially considering that the roots of other wild species of *Ipomoea* are known to be consumed in times of famine.

Sweet potato is presented by many as strong evidence of human exchanges across the Pacific in pre-European times. Other lines of biological evidence for contacts across the Pacific, such as bottle gourds, chickens and human DNA, are now considered questionable. The hypothesis of a human-mediated transfer of the sweet potato from America is based on the identification of similar terms to refer to the sweet potato in both regions (*cumar* and *kumara* respectively). As I showed in this thesis, this evidence seems to be based on a 19th century observation that was taken for granted and repeated by subsequent authors. Furthermore, the term *cumar* is exclusive to a Quechuan Andean dialect, so indicating the need for contacts between Polynesians and speakers of this specific dialect from America.

Most authors disregarded the possibility of a long-distance dispersal of the sweet potato into the Pacific by natural means. However, I have shown that long-distance dispersal is not uncommon in *Ipomoea*. I presented evidence of at least two species distributed across the Pacific but nested within clades of American species: *Ipomoea tuboides* and *I. littoralis*. Furthermore, the latter is a sweet potato CWR with seed and fruit morphology and dispersal biology similar to the crop. The distribution of these species is most easily explained as the result of long-distance dispersal by natural means. In the light of our results, it can be concluded that:

1) The evidence presented in this study, together with previous modelling studies, shows that **long-distance dispersal of the sweet potato by natural means** is possible.

2) The position in our analyses of the oldest sweet potato specimen from Polynesia, collected by Banks and Solander in the Society Islands in 1769, is congruent with this lineage being isolated for a long time.

Whether pre-European contacts across the Pacific existed or not, is a question that may never be satisfactorily answered. The possibility that the sweet potato was transported to Polynesia by pre-European travellers is, however, difficult to demonstrate. Our results show that, if sweet potato was present in Polynesia by the time Europeans first arrived, this could be more easily explained as the result of long-distance dispersal by natural means, so there is no need to invoke a human-mediated transport of the crop into the Pacific in pre-European times.

SUPPLEMENTARY FILES

The CD attached to this dissertation contains the following files:

- **Supplementary File 1**: Excel file including:
  - Sheets 1 and 2: passport data of all specimens included in this study (Hyb-Seq and DNA barcodes).
  - Sheets 3 and 4: descriptors of nuclear and chloroplast data.
  - Sheet 5: percentage of nucleotides in the Batatas group.
  - Sheet 6: GenBank accessions used in time-calibrated phylogenies.
  - Sheet 7: herbarium specimens of *Ipomoea trifida* used to define the species distribution presented in Figure 5.3.
- **Supplementary File 2**: *ITS* sequences included in this study in FASTA format.
- **Supplementary File 3**: single-copy nuclear coding regions included in this study (605 FASTA files compressed in a ZIP file).
- **Supplementary File 4**: whole chloroplast genome sequences in FASTA format.
- **Supplementary File 5**: *trnL-rpl32* sequences in FASTA format.
- **Supplementary File 6**: results of the mapDamage analysis of the specimen collected by Banks and Solander in 1769.
- **Supplementary File 7**: *Ipomoea* Identification System script.
- **Supplementary File 8**: nuclear phylogenies in NEWICK format.
- **Supplementary File 9**: chloroplast phylogenies in NEWICK format.
- **Supplementary File 10**: chloroplast alignment of the Batatas group in FASTA format.

REFERENCES

Abdel Khalik, K. (2013). Systematic implications of seed coat diversity in some representatives of the genus *Ipomoea* (Convolvulaceae). *Turk. J. Bot. 37*, 811–824.

Abdel Khalik, K., Osman, G., and Al-Amoudi, W. (2012). Genetic diversity and taxonomic relationships of some *Ipomoea* species based on analysis of RAPD-PCR and SDS-PAGE of seed proteins. *Austral. J. Crop Sci. 6*, 1088–1093.

Abel, W.E., and Austin, D.F. (1981). Introgressive hybridization between *Ipomoea trichocarpa* and *Ipomoea lacunosa* (Convolvulaceae). *Bull. Torrey Bot. Club 108*, 231.

Aguiar, D., and Istrail, S. (2012). HapCompass: a fast cycle basis algorithm for accurate haplotype assembly of sequence data. *J. Comput. Biol. 19*, 577–590.

Allan, H.H., Moore, L.B., Edgar, E., Healy, A.J., Webb, C.J., Sykes, W.R., Garnock-Jones, P.J., Given, D.R., Connor, H.E., and Dawson, M.I. (1961). *Flora of New Zealand* (Wellington, N.Z.: Government Printer).

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol. 215*, 403–410.

Asante, S.A., Yamada, T., Hisamatsu, M., and Shiotani, I. (1993). Studies on the properties of starch of diploid *Ipomoea trifida* (H. B. K.) Don. strains. *Starch 45*, 299–306.

Austin, D.F. (1975a). Family 164. Convolvulaceae. *Ann. Missouri Bot. Gard. 62*, 157–224.

Austin, D.F. (1975b). Typification of the New World subdivisions of *Ipomoea* L. (Convolvulaceae). *Taxon 24*, 107–110.

Austin, D.F. (1977). Hybrid polyploids in *Ipomoea* section *Batatas*. *J. Hered. 68*, 259–260.

Austin, D.F. (1978). The *Ipomoea batatas* Complex-I. Taxonomy. *Bull. Torrey Bot. Club 105*, 114–129.

Austin, D.F. (1979). An infrageneric classification for *Ipomoea* (Convolvulaceae). *Taxon 28*, 359–361.

Austin, D.F. (1980). Additional comments on infrageneric taxa in *Ipomoea* (Convolvulaceae). *Taxon 29*, 501–502.

Austin, D.F. (1988a). Nomenclatural changes in the *Ipomoea batatas* complex (Convolvulaceae). *Taxon 37*, 184–185.

Austin, D.F. (1988b). The taxonomy, evolution, and genetic diversity of the sweet potato and its wild relatives. In *Exploration, Maintenance and Utilization of Sweet Potato Genetic Resources* (Lima: International Potato Center), pp. 27–60.

Austin, D.F. (1991). *Ipomoea littoralis* (Convolvulaceae) - Taxonomy, distribution, and ethnobotany. *Econ. Bot. 45*, 251–256.

Austin, D.F. (1992). Rare Convolvulaceae in the Southwestern United States. *Ann. Missouri Bot. Gard. 79*, 8.

Austin, D.F. (1997). Dissolution of *Ipomoea* Series *Anisomerae* (Convolvulaceae). *J. Torr. Bot. Soc. 124*, 140.

Austin, D.F. (1998). Convolvulaceae Morning glory family. *J. Ariz-Nev. Acad. Sci. 30*, 61–83.

Austin, D.F. (2006). Noteworthy distributions and additions in Southwestern Convolvulaceae. *Canotia 2*, 79–106.

Austin, D.F. (2007a). Water Spinach (*Ipomoea aquatica*, Convolvulaceae): a food gone wild. *Ethnobot. Res. App. 5*, 123.

Austin, D.F. (2007b). *Merremia dissecta* (Convolvulaceae): condiment, medicine, ornamental and weed. A review. *Econ. Bot. 61*, 109–120.

Austin, D.F. (2011). Indian potato (*Ipomoea pandurata*, Convolvulaceae)—A record of confusion. *Econ. Bot. 65*, 408–421.

Austin, D.F., McDonald, J.A. & Murguía-Sánchez, G. *Ipomoea* (Convolvulaceae) (2012). In *Flora Mesoamericana* (Mexico, St. Louis & London: Universidad Nacional Autónoma de México, Missouri Botanical Garden and The Natural History Museum), 4(2), pp. 318-352.

Austin, D.F. (2013b). Moon–flower (*Ipomoea alba*, Convolvulaceae)—medicine, rubber enabler, and ornamental: a review. *Econ. Bot. 67*, 244–262.

Austin, D.F., and Bianchini, R.S. (1998). Additions and corrections in American *Ipomoea* (Convolvulaceae). *Taxon 47*, 833.

Austin, D.F., and Huáman, Z. (1996). A synopsis of *Ipomoea* (Convolvulaceae) in the Americas. *Taxon 45*, 3–38.

Austin, D.F., and McDonald, J.A. (2014). Relationships and taxonomy of *Ipomoea variabilis* (Convolvulaceae). *Phytoneuron 67*, 1–4.

Austin, D.F., Jarret, R.L., Tapia, C., and de la Puente, F. (1992). Collecting tetraploid *I. batatas* (Linnaeus) Lamarck in Ecuador. *Pl. Genet. Resources Newlett.* 33–36.

Austin, D.F., Jarret, R.L., and Johnson, R.W. (1993). *Ipomoea gracilis* R.Brown (Convolvulaceae) and its allies. *Bull. Torrey Bot. Club 120*, 49–59.

Austin, D.F., Kitajima, K., Yoneda, Y., and Qian, L. (2001). A putative tropical American plant, *Ipomoea nil* (Convolvulaceae), in pre-Columbian Japanese art. *Econ. Bot. 55*, 515–527.

Austin, D.F., Staples, G.W., and Simão-Bianchini, R. (2015). A synopsis of *Ipomoea* (Convolvulaceae) in the Americas: further corrections, changes, and additions. *Taxon 64*, 625–633.

Baafi, E., Blay, E.T., Ofori, K., Gracen, V.E., Manu-Aduening, J., and Carey, E.E. (2016). Breeding superior orange-fleshed sweetpotato cultivars for West Africa. *J. Crop Improv. 30*, 293–310.

Baldwin, B.G. (1992). Phylogenetic utility of the *Internal Transcribed Spacers* of nuclear ribosomal DNA in plants: an example from the Compositae. *Mol. Phyl. Evol. 1*, 3–16.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., and Donoghue, M.J. (1995). The *ITS* region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann. Missouri Bot. Gard. 82*, 247–277.

Ballard, C., Brown, P., Bourke, R.M., and Harwood, T. (Eds.) (2005). *The sweet potato in Oceania: a reappraisal* (Sydney NSW, Australia: University of Sydney).

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol. 19*, 455–477.

Barrau, J. (1957). L'énigme de la patate douce en Oceanie. *Études d'Outre-Mer 40*.

Baucom, R.S., Chang, S.-M., Kniskern, J.M., Rausher, M.D., and Stinchcombe, J.R. (2011). Morning glory as a powerful model in ecological genomics: tracing adaptation through both natural and artificial selection. *Heredity 107*, 377–385.

Beavan, N. (2014). No evidence for sample contamination or diet offset for pre-Columbian chicken dates from El Arenal. *Proc. Natl. Acad. Sci. U.S.A. 111*, E3582–E3582.

Best, E. (1925). *The Maori canoe* (Wellington, N.Z.: A.R. Shearer).

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics 27*, 578–579.

Bohac, J.R., Austin, D.F., and Jones, A. (1993). Discovery of wild tetraploid sweetpotatoes. *Econ. Bot. 47*, 193–201.

Boissier, P.-E. (1839). *Voyage botanique dans le midi de l'Espagne* (Paris: Gide et companie, Libraires-Éditeurs).

Bojer, W. (1837). *Hortus Mauritianus: ou enumeration des plantes, exotiques et indigenes, qui croissent a l'Ile Maurice, disposees d'apres la methode naturelle* (Maurice: Imprimerie d'Aimé Mamarot et Compagnie).

Bonatto, S.L., Redd, A.J., Salzano, F.M., and Stoneking, M. (1996). Lack of ancient Polynesian-Amerindian contact. *Am. J. Hum. Genet. 59*, 253–258.

Bonavia, D. (1984). La importancia de los restos de papas y camotes de época precerámica hallados en el valle de Casma. *J. Soc. Am. 70*, 7–20.

Bovell-Benjamin, A.C. (2007). Sweet potato: a review of its past, present, and future role in human nutrition. In *Advances in Food and Nutrition Research*, S.L. Taylor, ed. (Nebraska: Academic Press), pp. 1–59.

Brand, D.D. (1971). The sweet potato: an exercise in methodology. In *Man across the Sea. Problems of Pre-Columbian contacts*, C.L. Riley, J.C. Kelley, C.W. Pennington, and R.L. Rands, eds. (Austin & London: University of Texas Press), pp. 343–365.

Brandegee, T.S. (1889). A collection of plants from Baja California, 1889. *Proc. Natl. Acad. Sci. U.S.A. 2*, 117–216.

Bruen, T.C. (2005). A simple and robust statistical test for detecting the presence of recombination. *Genetics 172*, 2665–2681.

Bryant, D. (2014). Statistical flaws undermine pre-Columbian chicken debate. *Proc. Natl. Acad. Sci. U.S.A. 111*, E3584–E3584.

Buck, P.H. (1938). *Vikings of the sunrise* (New York: Frederic A. Stokes).

Burri, B.J. (1997). Beta-carotene and human health: a review of current research. *Nutr. Res. 17*, 547–580.

Burrows, E.G. (1936). *Ethnology of Futuna* (Honolulu, Hawaii: Bernice P. Bishop Museum).

Burrows, E.G. (1937). *Ethnology of Uvea (Wallis Island)* (Honolulu, Hawaii: Bernice P. Bishop Museum).

Candolle, A. de (1855). *Géographie botanique raisonnée; ou, Exposition des faits principaux et des lois concernant la distribution géographique des plantes de l'epoque actuelle* (Paris: V. Masson).

Candolle, A. de (1883). *Origine des plantes cultivées* (Paris: Germer Baillière).

Cann, R.L., and Lum, J.K. (1996). Mitochondrial myopia: reply to Bonatto et al. *Am. J. Hum. Genet. 59*, 256–258.

Carine, M.A., and Scotland, R.W. (2002). Classification of Strobilanthinae (Acanthaceae): trying to classify the unclassifiable? *Taxon 51*, 259.

Carlquist, S. (1967). The biota of long-distance dispersal. V. Plant dispersal to Pacific Islands. *Bull. Torrey Bot. Club 94*, 129.

Carter, G.F. (1950). Plant evidence for early contacts with America. Southwest. *J. Anthropol. 6*, 161–182.

de las Casas, B. (1552). *Viajes de Cristóbal Colón* (Unpublished).

Castañeda-Álvarez, N.P., Khoury, C.K., Achicanoy, H.A., Bernau, V., Dempewolf, H., Eastwood, R.J., Guarino, L., Harker, R.H., Jarvis, A., Maxted, N., et al. (2016). Global conservation priorities for crop wild relatives. *Nat. Plants 2*, 16022.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol. 17*, 540–552.

Cervantes-Flores, J.C., Yencho, C.C., Pecota, K.V., Sosinski, B., and Mwanga, R.O.M. (2008). Detection of quantitative trait loci and inheritance of root-knot nematode resistance in sweetpotato. *J. Amer. Soc. Hort. Sci. 133*, 844–851.

Cheeseman, T.F. (1903). VI. The flora of Rarotonga, the chief island of the Cook group. *Trans. Linn. Soc. London, Bot. 6*, 261–313.

Chifman, J., and Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics 30*, 3317–3324.

Chifman, J., and Kubatko, L. (2015). Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *J. Theoret. Biol. 374*, 35–47.

Chitchak, N., Traiperm, P., Staples, G., Rattanakrajang, P., and Sumanon, P. (2018). Species delimitation of some *Argyreia* (Convolvulaceae) using phenetic analyses: insights from leaf anatomical data reveal a new species. *Botany 96*, 217–233.

Chock, A.K. (1968). Hawaiian ethnobotanical studies I. Native food and beverage plants. *Econ. Bot. 22*, 221–238.

Chodat, R., and Hassler, É. (1905). Plantae Hasslerianae soit énumération des plantes recoltées au Paraguay par le Dr Emile Hassle d'Aarau (Suisse), de 1885 á 1902. *Bull. Herb. Boissier, sér. 2 5*, 288–305.

Choisy, J.D. (1833). Convolvulaceae Orientales. *Mem. Soc. Phys. Hist. Nat. Geneve 6(1-2)*, 383–502.

Choisy, J.D. (1838). De Convolvulaceis. In *Memoires de la Société de Physique et d'Histoire Naturelle de Genève* (Geneve: Librairie d'Abraham Cherbuliez), pp. 43–86.

Choisy, J.D. (1845). Convolvulaceae. In *Prodromus Systematis Naturalis Regni Vegetabilis*, A.-L.-P.-P. de Candolle, ed. (Parisii: Sumptibus Sociorum Treuttel et Würtz), pp. 323–462.

Christoff, A.-P., Loreto, E.L.S., and Sepel, L.M.N. (2012). Evolutionary history of the *Tip100* transposon in the genus *Ipomoea*. *Gen. Mol. Biol. 35*, 460–465.

Clarke, A.C. (2009). *Origins and dispersal of the sweet potato and bottle gourd in Oceania: implications for prehistoric human mobility*. Massey University.

Clarke, C.B. (1885). Convolvulaceae. In *The Flora of British India*, J.D. Hooker, ed. (London: L. Reeve & Co., Ltd), pp. 179–228.

Clement, C., De Cristo-Araújo, M., Coppens D'Eeckenbrugge, G., Alves Pereira, A., and Picanço-Rodrigues, D. (2010). Origin and domestication of Native Amazonian crops. *Diversity 2*, 72–106.

Clement, C.R., Denevan, W.M., Heckenberger, M.J., Junqueira, A.B., Neves, E.G., Teixeira, W.G., and Woods, W.I. (2015). The domestication of Amazonia before European conquest. *Proc. R. Soc. Lond. B: Biol. Sci. 282*, 20150813.

Cohen, M.N. (1978a). Archaeological plant remains from the central coast of Peru. *Ñawpa Pacha 16*, 23–50.

Cohen, M.N. (1978b). Population pressure and the origins of agriculture: an archaeological example from the coast of Peru. In *Advances in Andean Archaeology*, D.L. Browman, ed. (The Hague: Mouton Publishers), pp. 91–132.

Coil, J.H., and Kirch, P.V. (2005). An Ipomoean landscape: archaeology and the sweet potato in Kahikinui, Maui, Hawaiian Islands. In *The sweet potato in Oceania: a re-appraisal*, C. Ballard, P. Brown, R.M. Bourke, and T. Harwood, eds. (Sidney: University of Sydney), pp. 71–84.

Conklin, H.C. (1963). The Oceanian-African hypothesis and the sweet potato. In *Plants and the migration of Pacific peoples: a symposium*, (Honolulu, Hawaii: Bishop Museum Press), pp. 129–136.

Cook, O.F., and Cook, R.C. (1918). The maho, or mahagua, as a trans-Pacific plant. *J. Wash. Acad. Sci. 8*, 153–170.

Corney, B.G. (1903). Extract from the official log of the voyage of Mynheer Jacob Roggeveen, in the ships Den Arend, Thienhoven and De Afrikaansche Galey, in 1721-2 in so far as it relates to the discovery of Easter Island. In *The Voyage of Captain Don Felipe González to Easter Island 1770-1* (Cambridge: Hakluyt Society), p. 21.

Costantin, and Bois (1910). Sur les graines et tubercules des tombeaux Péruviens de la période Incasique. *Rev. Gen. Bot. 22*, 242–265.

Cox, P.A., and Banack, S.A. (1991). Islands, plants, and Polynesians: an introduction to Polynesian ethnobotany: proceedings of a symposium (Portland, Oregon: Dioscorides Press).

Cuevas, H.E., Rosa-Valentín, G., Hayes, C.M., Rooney, W.L., and Hoffmann, L. (2017). Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: implications for germplasm conservation, evaluation, and utilization in crop improvement. *BMC Genom. 18*.

Dahlgren, E.W. (1917). Were the Hawaiian Islands visited by the Spaniards before their discovery by Captain Cook in 1778? *Kungl. Svenska Vetenskapsakad. Handl. 57*, 1–222.

Dammer, C.L.U. (1895). *Die Pflanzenwelt Ost-Afrikas und der Nachbargebiete* (Berlin: Dietrich Reimer).

Darling, A.C.E. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res. 14*, 1394–1403.

Darling, A.E., Mau, B., and Perna, N.T. (2010). ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE 5*, e11147.

Darlington, W. (1847). *Agricultural botany an enumeration and description of useful plants and weeds, which merit the notice, or require the attention, of American agriculturists* (New York: J.W. Moore).

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). JModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods 9*, 772–772.

Das, S. (2011). Congruence between morphological and molecular approach in understanding species relationship in *Ipomoea* spp.: a rare event in taxonomy. *Asian J. Plant Sci. 10*, 263–268.

De Queiroz, K.A. (2005). A unified concept of species and its consequences for the future of taxonomy. *Proc. Calif. Acad. Sci. 56(1)*, 196–215.

De Queiroz, K.A. (2007). Species concepts and species delimitation. *Syst. Biol. 56(6)*, 879–886.

Demissew, S., and Mill, R.R. (2009). Revision of the genus *Seddera* (Convolvulaceae). *Kew Bull. 64*, 197–233.

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci. 57*, 1–13.

Denham, T. (2013). Ancient and historic dispersals of sweet potato in Oceania. *Proc. Natl. Acad. Sci. U.S.A. 110*, 1982–1983.

Deroin, T. (1999). Ontogeny and phylogeny in Convolvulaceae-*Ipomoeeae*: preliminary comparative remarks on ovary morphology. *Syst. Geogr. Pl. 68*, 225.

Dhillon, N.P.S., and Ishiki, K. (1999). Genomic variation and genetic relationships in *Ipomoea* spp. *Plant. Breed. 118*, 161–165.

Diaz, J., Schmiediche, P., and Austin, D.F. (1996). Polygon of crossability between eleven species of *Ipomoea*: section *Batatas* (Convolvulaceae). *Euphytica 88*, 189–200.

Dixon, R.B. (1932). The problem of the sweet potato in Polynesia. *Am. Anthropol. 34*, 40–66.

Dorsey, G.A. (1893). The necropolis of Ancon, Peru. *Sci. Am. 36*, 14851–14852.

Doyle, J.J., and Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus 12*, 13–15.

Dressler, R.L. (1953). The pre-Columbian cultivated plants of Mexico. *Bot. Mus. Leaflets, Harvard Univ. 16*, 115–172.

Drouin, G., Daoud, H., and Xia, J. (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phyl. Evol. 49*, 827–831.

Duncan, T.M., and Rausher, M.D. (2013a). Evolution of the selfing syndrome in *Ipomoea*. *Front. Plant Sci. 4*, 1–8.

Duncan, T.M., and Rausher, M.D. (2013b). Morphological and genetic differentiation and reproductive isolation among closely related taxa in the *Ipomoea* series *Batatas*. *Amer. J. Bot. 100*, 2183–2193.

Edgar, R.C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics 5*, 113.

Edgar, R.C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res. 32*, 1792–1797.

Engel, F. (1970a). Exploration of the Chilca Canyon, Peru. *Curr. Anthropol. 11*, 55–58.

Engel, F. (1970b). La grotte du Mégathérium à Chilca et les écologies du Haut-Holocène péruvien. In *Échanges et communications, Mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60e anniversaire* (Mouton, La Haye), pp. 413–436.

Engel, F. (1970c). Recolección y cultivo en los Andes precolombinos. *An. Cient. 8*, 122–136.

Engel, F. (1973). New facts about pre-Columbian life in the Andean lomas. *Curr. Anthropol. 14*, 271–280.

Ennos, R.A. (1981). Quantitative studies of the mating system in two sympatric species of *Ipomoea* (Convolvulaceae). *Genetica 57*, 93–98.

Eserman, L.A. (2012). *Taxonomy and crossing relationships in a small group of morning glories (*Ipomoea *section* Pharbitis*)*. Southeastern Louisiana University.

Eserman, L.A., Tiley, G.P., Jarret, R.L., Leebens-Mack, J.H., and Miller, R.E. (2014). Phylogenetics and diversification of morning glories (tribe *Ipomoeeae*, Convolvulaceae) based on whole plastome sequences. *Amer. J. Bot. 101*, 92–103.

Eserman, L.A., Jarret, R.L., and Leebens-Mack, J.H. (2018). Parallel evolution of storage roots in morning glories (Convolvulaceae). *BMC Plant Biol. 18*.

Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics 164*, 1567–1587.

Fehren-Schmitz, L., Jarman, C.L., Harkins, K.M., Kayser, M., Popp, B.N., and Skoglund, P. (2017). Genetic ancestry of Rapanui before and after European contact. *Curr. Biol. 27(20)*, 3209–3215.E6.

Felger, R.S., Austin, D.F., Van Devender, T., Sánchez-Escalante, J.J., and Costea, M. (2012). Convolvulaceae of Sonora, Mexico. I. *Convolvulus*, *Cressa*, *Dichondra*, *Evolvulus*, *Ipomoea*, *Jacquemontia*, *Merremia* and *Operculina*. *J. Bot. Res. Inst. Texas 6*, 459–527.

Feliner, G.N., and Rosselló, J.A. (2007). Better the devil you know? Guidelines for insightful utilization of *nrDNA ITS* in species-level evolutionary studies in plants. *Mol. Phyl. Evol. 44*, 911–919.

Felsenstein, J. 2005. *PHYLIP (Phylogeny Inference Package) version 3.6*. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Fernández-Pozo, N., Menda, M., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., Yan, A. & Mueller, L.A. (2015). The Sol Genomics Network (SGN) — from genotype to phenotype to breeding. *Nucl. Ac. Res.* 43 (D1): D1036-D1041.

Florence, J., and Lorence, D.H. (1997). Introduction to the flora and vegetation of the Marquesas Islands. *Allertonia 7*, 226–237.

Florence, J., Waldren, S., and Chepstow-Lusty, A.J. (1995). The flora of the Pitcairn Islands: a review. *Biol. J. Linn. Soc. 56*, 79–119.

Folk, R.A., Mandel, J.R., and Freudenstein, J.V. (2017). Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Syst. Biol. 66*, 320–337.

Folorunso, A.E. (2013). Taxonomic evaluation of fifteen species of *Ipomoea* L. (Convolvulaceae) from South-Western Nigeria using foliar micromorphological characters. *Not. Sci. Biol. 5*, 156.

Folorunso, A.E., Illoh, H.C., and Olorungbeja, J.A. (2013). Numerical taxonomy of some *Ipomoea* (Linn.) species in South-West Nigeria. *Ife J. Sci. 15*, 63–74.

Food and Agriculture Organization of the United Nations (2018). *FAOSTAT Statistics Database*.

Fosberg, F.R., and Sachet, M.-H. (1977). *Flora of Micronesia, 3: Convolvulaceae* (Washington: Smithsonian Institution Press).

Franklin, J., Keppel, G., and Whistler, W.A. (2008). The vegetation and flora of Lakeba, Nayau and Aiwa Islands, Central Lau Group, Fiji. *Micronesica 40*, 169–225.

Friedberg, C. (1959). Contribution à l'étude ethnobotanique des tombes précolombiennes de Lauri (Pérou) (suite et fin) - Part II. *J. Agr. Trad. Bot. App. 6*, 405–434.

Friederici, G. (1929). Zu den vorkolumbischen Verbindungen der Südsee-Völker mit Amerika. *Anthropos 24*, 441–487.

Friedlaender, J.S., Friedlaender, F.R., Reed, F.A., Kidd, K.K., Kidd, J.R., Chambers, G.K., Lea, R.A., Loo, J.-H., Koki, G., Hodgson, J.A., et al. (2008). The genetic structure of Pacific islanders. *PLoS Genet. 4*, e19.

Gentry, H.S. (1942). *Rio Mayo Plants. A study of the flora and vegetation of the valley of the rio Mayo, Sonora* (Washington, DC: Carnegie Institution of Washington).

González-Pérez, E., Esteban, E., Vía, M., García-Moro, C., Hernández, M., and Moral, P. (2006). Genetic change in the Polynesian population of Easter Island: evidence from *Alu* insertion polymorphisms. *Ann. Human Genet. 70*, 829–840.

Goodwin, Z.A., Harris, D.J., Filer, D., Wood, J.R.I., and Scotland, R.W. (2015). Widespread mistaken identity in tropical plant collections. *Curr. Biol. 25*, R1066–R1067.

Gosling, A.L., and Matisoo-Smith, E.A. (2018). The evolutionary history and human settlement of Australia and the Pacific. *Curr. Opin. Gen. Dev. 53*, 53–59.

Gray, A. (1878). Order XCIV. Convolvulaceae. In *Synoptical Flora of North America* (New York: American Book Company), pp. 207–223.

Gray, A., and Trumbull, J.H. (1883). Review of De Candolle's "Origin of cultivated plants" with annotations upon certain American species. *Amer J. Sci. 25*, 241–255.

Green, R.C. (2000). A range of disciplines support a dual origin for the bottle gourd in the Pacific. *J. Pol. Soc. 109*, 191–197.

Green, R.C. (2005). Sweet potato transfers in Polynesian prehistory. In *The sweet potato in Oceania: a reappraisal* (Sidney: University of Sidney), p. 43–62.

Grisebach, A. (1864). *Flora of the British West Indian islands* (London: L. Reeve & Co., Ltd).

Gunn, C.R. (1969). History and taxonomy of the purple moonflower, *Ipomoea turbinata* Lagasca y Segura. *Proc. Assoc. Off. Seed Anal. 59*, 116–123.

Guppy, H.B. (1887). Chapter XI. The journal of Gallego - prefatory remarks. In *The Solomon Islands and Their Natives* (London: Swan Sonnenschein, Lowrey), pp. 192–245.

Guppy, H.B. (1906). *Observations of a naturalist in the Pacific between 1896 and 1899* (London: MacMillan and Co., Limited).

Hackney, L.A. of, and Thomson, B. (1901). *The discovery of the Solomon Islands by Alvaro de Mendaña in 1568* (London: They Hakluyt Society).

Hagelberg, E., Quevedo, S., Turbon, D., and Clegg, J.B. (1994). DNA from ancient Easter Islanders. *Nature 369*, 25–26.

Hallier, H. (1893). Versuch einer natürlichen Gliederung der Convolvulaceen auf morphologischer und anatomischer Grundlage. *Bot. Jahrb. Syst. 16*, 453–591.

Hallier, H. (1894). Convolvulaceae Africanae. *Bot. Jahrb. Syst. 28*, 28–54.

Hanbury, D. (1871). On a species of *Ipomoea*, affording Tampico jalap. *J. Linn. Soc., Bot. 11*, 279–284.

Handy, E.S.C. (1940). *The Hawaiian planter. His plants, methods and areas of cultivation* (Honolulu, Hawaii: Bishop Museum).

Hargreaves, B.J. (1994). Tubers and bulbs as an adaptive strategy in the Kalahari. In *The Biodiversity of African Plants. Proceedings of the 14th AETFAT Congress*, 22-27 August 1994, L.J.G. van der Maesen, X.M. van der Burgt, and J.M. van Medenbach de Rooy, eds. (Wageningen, The Netherlands), pp. 88–91.

Harries, H.C., and Clement, C.R. (2014). Long-distance dispersal of the coconut palm by migration within the coral atoll ecosystem. *Ann. Bot. 113*, 565–570.

Hather, J.G. (1988). *The morphological and anatomical interpretation and identification of charred vegetative parenchymatous plant remains.* Two volumes. University of London.

Hather, J.G. (2000). *Archaeological parenchyma* (London: Archetype).

Hather, J., and Kirch, P.V. (1991). Prehistoric sweet potato (*Ipomoea batatas*) from Mangaia Island, Central Polynesia. *Antiquity 65*, 887–893.

Hebert, P.D.N., Cywinska, A., Ball, S.L., and deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B: Biol. Sci. 270*, 313–321.

Hedtke, S.M., and Hillis, D.M. (2011). The potential role of androgenesis in cytoplasmic-nuclear phylogenetic discordance. *Syst. Biol. 60*, 87–96.

Heiser, C.B. (1979). *The gourd book* (Norman, USA: University of Oklahoma Press).

Heiser Jr., C.B. (1965). Cultivated plants and cultural diffusion in nuclear America. *Am. Anthropol. 67,* 930–949.

Hemsley, W.B. (1894). The flora of the Tonga or Friendly Islands, with descriptions of and notes on some new or remarkable plants, partly from the Solomon Islands. *J. Linn. Soc., Bot. 30,* 158–217.

Heyerdahl, Y. (1952). *American Indians in the Pacific. The theory behind the Kon-Tiki expedition* (George Allen & Unwin Ltd.).

Hillebrand, W. (1888). *Flora of the Hawaiian Islands: a description of their phanerogams and vascular cryptogams* (London: Williams & Norgate).

Ho, P.-T. (1955). The introduction of American food plants into China, part. 1. *Am. Anthropol., New. Ser. 57,* 191–201.

Hohna, S., Heath, T.H., Boussau, B., Landis, M.J., Ronquist, F., and Huelsenbeck, J.P. (2014). Probabilistic graphical model representation in phylogenetics. *Syst. Biol. 63,* 753–771.

Hohna, S., Heath, T.A., Bastien, B., Landis, M.J., Frederik, R., and Huelsenbeck, J. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol. 65,* 726–736.

Holm, L.G. (1997). *World weeds: natural histories and distribution* (New York: Wiley).

Hooker, J.D. (1867). *Handbook of the New Zealand flora* (London: L. Reeve & Co., Ltd).

Horak, M.J., and Wax, L.M. (1991). Germination and seedling development of bigroot morning glory (*Ipomoea pandurata*). *Weed Sci. 39,* 390–396.

Hornell, J. (1946). How did the sweet potato reach Oceania? *J. Linn. Soc., Bot. 53,* 41–62.

Horrocks, M. (2004). Polynesian plant subsistence in prehistoric New Zealand: A summary of the microfossil evidence. *New Zealand J. Bot. 42,* 321–334.

Horrocks, M., and Lawlor, I. (2006). Plant microfossil analysis of soils from Polynesian stonefields in South Auckland, New Zealand. *J. Arch. Sci. 33,* 200–217.

Horrocks, M., Irwin, G., Jones, M., and Sutton, D. (2004a). Starch grains and xylem cells of sweet potato (*Ipomoea batatas*) and bracken (*Pteridium esculentum*) in archaeological deposits from northern New Zealand. *J. Arch. Sci. 31,* 251–258.

Horrocks, M., Shane, P., Barber, I., D'Costa, D., and Nichol, S. (2004b). Microbotanical remains reveal Polynesian agriculture and mixed cropping in early New Zealand. *Rev. Palaeobot. Palynol. 131*, 147–157.

Horsburgh, K., and McCoy, M. (2017). Dispersal, isolation, and interaction in the islands of Polynesia: a critical review of archaeological and genetic evidence. *Diversity 9*, 37.

House, H.D. (1908). The North American species of the genus *Ipomoea*. *Ann. New York Acad. Sci. 18*, 181–263.

Howard, R.A., and McDonald, J.A. (1995). *Ipomoea sphenophylla* Urban recollected and neotypified. *Harv. Pap. Bot. 1*, 69–72.

Huang, J.C., and Sun, M. (2000). Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series *Batatas* (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theor. Appl. Genet. 100*, 1050–1060.

Huang, J., Corke, H., and Sun, M. (2002). Highly polymorphic AFLP markers as a complementary tool to *ITS* sequences in assessing genetic diversity and phylogenetic relationships of sweetpotato (*Ipomoea batatas* (L.) Lam.) and its wild relatives. *Genet. Resources Crop Evol. 49*, 541–550.

Hudjashov, G., Endicott, P., Post, H., Nagle, N., Ho, S.Y.W., Lawson, D.J., Reidla, M., Karmin, M., Rootsi, S., Metspalu, E., et al. (2018). Investigating the origins of eastern Polynesians using genome-wide data from the Leeward Society Isles. *Sci. Rep. 8*.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics 17*, 754–755.

Humboldt, A. von (1825). *Essai politique sur le royaume de la Nouvelle-Espagne* (Paris: Chez Antoine-Augustin Renouard).

Hurles, M.E., Maund, E., Nicholson, J., Bosch, E., Renfrew, C., Sykes, B.C., and Jobling, M.A. (2003). Native American Y chromosomes in Polynesia: the genetic impact of the Polynesian slave trade. *Am. J. Hum. Genet. 72*, 1282–1287.

Huxley, J.S. (1957). The three types of evolutionary process. *Nature 180*, 454–455.

Huxley, J.S. (1958). Evolutionary process and taxonomy with special reference to grades. *Uppsala Univ. Arsskr. 6*, 21–38.

IBM Corp (2016). *IBM SPSS Statistics* (Armonk, NY: IBM Corp.).

Infante-Betancour, J. (2014). Una nueva especie de *Ipomoea* (Convolvulaceae) para Colombia. *Caldasia 36*, 247–252.

Ivens, W.G. (1918). *Dictionary and grammar of the language of Sa'a and Ulawa, Solomon Islands* (Washington: The Carnegie Institution of Washington).

Jacobs, T.V. (2000). Underutilized edible plants from South Africa: a perspective. In *Managing Plant Genetic Diversity. Proceedings of an International Conference, Kuala Lumpur, Malaysia, 12-16 June 2000*, pp. 371–378.

Jarret, R.L., and Austin, D.F. (1993). Genetic diversity and systematic relationships in sweetpotato (*Ipomoea batatas* (L.) Lam.) and related species as revealed by RAPD analysis. *Genet. Resources Crop Evol. 41*, 165–173.

Jarret, R.L., Gawel, N., and Whittemore, A. (1992). Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. *J. Amer. Soc. Hort. Sci. 117*, 633–637.

Jayeola, A.A., and Oladunjoye, O.R. (2012). Systematic studies in some *Ipomoea* Linn. species using pollen and flower morphology. *Ann. West. Univ. Timisoara, Ser. Biol. 15*, 177–187.

Jett, S.C. (2017). *Ancient ocean crossings: reconsidering the case for contacts with the pre-Columbian Americas* (Tuscaloosa: The University of Alabama Press).

Joel, D.M., and Listone, A. (1986). New adventive weeds in Israel. *Israel J. Bot. 35*, 215–223.

John, H.S. (1976). A plant collection from Niue Island by Jensen in 1876 Pacific plant studies 29. *Bot. Mag. Tokyo 89*, 235–240.

Johnson, R.W. (1986). Four new species of *Ipomoea* L. (Convolvulaceae) from Australia. *Austrobaileya 2*, 217–223.

Johnson, R.W. (2012). New species and subspecies of *Ipomoea* L. (Convolvulaceae) from northern Australia and a key to the Australian species. *Austrobaileya 8*, 699–723.

Jones, A. (1964). Chromosome numbers in the genus *Ipomoea*. *J. Hered. 55*, 216–219.

Jones, A. (1967). Should Nishiyama's K123 (*Ipomoea trifida*) be designated *I. batatas*? *Econ. Bot. 21*, 163–166.

Jones, A. (1968). Chromosome numbers in *Ipomoea* and related genera. *J. Hered. 59*, 99–102.

Jones, A. (1970). Asynapsis in Ipomoea gracilis. *J. Hered. 61*, 151–152.

Jones, T.L. (2011). *Polynesians in America: pre-Columbian contacts with the New World* (Lanham: AltaMira Press).

Jones, K.M., and de Brauw, A. (2015). Using agriculture to improve child health: promoting orange sweet potatoes reduces diarrhea. *World Dev. 74*, 15–24.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). MapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics 29*, 1682–1684.

Katayama, K., Kobayashi, A., Sakai, T., Kuranouchi, T., and Kai, Y. (2017). Recent progress in sweetpotato breeding and cultivars for diverse applications in Japan. *Breed. Sci. 67*, 3–14.

Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res. 30*, 3059–3066.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol. 30*, 772–780.

Kaur, J., Rawat, A., Kumar, S., and Narain, S. (2016). Taxonomy, phytochemistry, traditional uses and cultivation of *Ipomoea aquatica* Forsk. *IJIR 2*, 408–412.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics 28*, 1647–1649.

Kenyan, F.M.G. (1928). A morphological and cytological study of *Ipomoea trifida*. *Bull. Torrey Bot. Club 55*, 499–512.

Khoury, C.K., Heider, B., Castañeda-Álvarez, N.P., Achicanoy, H.A., Sosa, C.C., Miller, R.E., Scotland, R.W., Wood, J.R.I., Rossel, G., Eserman, L.A., et al. (2015). Distributions, ex situ conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., series Batatas]. *Frontiers Pl. Sci. 6*, 251.

King, J.R., and Bamford, R. (1937). The chromosome number in *Ipomoea* and related genera. *J. Hered. 28*, 279–282.

Kirch, P., Coil, J., Hartshorn, A., Jeraj, M., Vitousek, P., and Chadwick, O. (2005). Intensive dryland farming on the leeward slopes of Haleakala, Maui, Hawaiian Islands: archaeological, archaeobotanical, and geochemical perspectives. *World Archaeol. 37*, 240–258.

Kirch, P.V., Steadman, D.W., Butler, V.L., Hather, J., and Weisler, M.I. (1995). Prehistory and human ecology in Eastern Polynesia: excavations at Tangatatau rockshelter, Mangaia, Cook Islands. *Archaeol. Ocean. 30*, 47–65.

Kirch, P.V., Hartshorn, A.S., Chadwick, O.A., Vitousek, P.M., Sherrod, D.R., Coil, J., Holm, L., and Sharp, W.D. (2004). Environment, agriculture, and settlement patterns in a marginal Polynesian landscape. *Proc. Natl. Acad. Sci. U.S.A. 101*, 9936–9941.

Kistler, L., Montenegro, A., Smith, B.D., Gifford, J.A., Green, R.E., Newsom, L.A., and Shapiro, B. (2014). Transoceanic drift and the domestication of African bottle gourds in the Americas. *Proc. Natl. Acad. Sci. U.S.A. 111*, 2937–2941.

Knapp, M., Horsburgh, K.A., Prost, S., Stanton, J.-A., Buckley, H.R., Walter, R.K., and Matisoo-Smith, E.A. (2012). Complete mitochondrial DNA genome sequences from the first New Zealanders. *Proc. Natl. Acad. Sci. U.S.A. 109*, 18350–18354.

Knowles, L.L., Carstens, B.C. & Weins, J. (2007). Delimiting species without monophyletic gene trees. *Syst. Biol. 56(6)*, 887–895.

Kobayashi, M. (1984). The *Ipomoea trifida* complex closely related to sweet potato. In *Proceedings of the 6th Symposium of the International Society of Tropical Root Crop* (Lima: Centro Internacional de la Papa), pp. 561–568.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resourc. 15*, 1179–1191.

Kress, W.J. (2017). Plant DNA barcodes: applications today and in the future. *J. Syst. Evol. 55*, 291–307.

Kurabachew, H. (2015). The role of orange fleshed sweet potato (*Ipomea batatas*) for combating vitamin A deficiency in Ethiopia: a review. *Int. J. Food Sci. Nutr. Eng. 5*, 141–146.

Kyndt, T., Quispe, D., Zhai, H., Jarret, R.L., Ghislain, M., Liu, Q., Gheysen, G., and Kreuze, J.F. (2015). The genome of cultivated sweet potato contains "*Agrobacterium*" T-DNAs with expressed genes: an example of a naturally transgenic food crop. *Proc. Natl. Acad. Sci. U.S.A.* 201419685.

Ladefoged, T.N., Graves, M.W., and Coil, J.H. (2005). The introduction of sweet potato in Polynesia: early remains in Hawai'i. *J. Pol. Soc. 114*, 359–374.

Lamarck, J.-B. (1793). *Tableau encyclopédique et méthodique des trois règnes de la nature* (A Paris: Chez Panckoucke).

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol. 10*, R25.

Lanning, E.P. (1965). Early man in Peru. *Sci. Am. 213*, 68–76.

Lejoly, J., and Lisowski, S. (1986). *Paralepistemon*, nouveau genre de Convolvulaceae d'Afrique tropicale. *Bull. Jard. Bot. Belg. 56*, 196.

Lepofsky, D. (2003). The ethnobotany of cultivated plants of the Maohi of the Society Islands. *Econ. Bot. 57*, 73–92.

Lie, B.A., Dupuy, B.M., Spurkland, A., Fernández-Viña, M.A., Hagelberg, E., and Thorsby, E. (2007). Molecular genetic studies of natives on Easter Island: evidence of an early European and Amerindian contribution to the Polynesian gene pool. *Tiss. Antig. 69*, 10–18.

Linajes, A., Rico-Gray, V., and Carrión, G. (1994). Traditional production system of the root of jalapa, *Ipomoea purga* (Convolvulaceae) in Central Veracruz, Mexico. *Econ. Bot. 48*, 84–89.

Linnaeus, C. von (1753). *Species Plantarum* (Holmiae: Impensis Laurentii Salvii).

Lipp, F.J. (1971). Ethnobotany of the Chinantec Indians, Oaxaca, Mexico. *Econ. Bot. 25*, 234–244.

Loytynoja, A., and Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science 320*, 1632–1635.

MacNeish, R.S., Patterson, T.C., and Browman, D.L. (1975). *The central Peruvian prehistoric interaction sphere* (Andover, Massachusetts: Robert S. Peabody Foundation).

Madani, I., and Majbour, M. (2017). Important anatomical features in the taxonomy of selected plant species from the genus *Ipomoea* (Convolvulaceae) in Sudan. *Sudan J. Sci. 9*.

Magallon, S., Gomez-Acevedo, S., Sanchez-Reyes, L.L., and Hernandez-Hernandez, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol. 207*, 437–453.

Malaspinas, A.-S., Lao, O., Schroeder, H., Rasmussen, M., Raghavan, M., Moltke, I., Campos, P.F., Sagredo, F.S., Rasmussen, S., Gonçalves, V.F., et al. (2014). Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. *Curr. Biol. 24*, R1035–R1037.

Malo, D. (1898). *Hawaiian antiquities* (Honolulu, Hawaii: Hawaiian Gazette Co., Ltd.).

Manos, P.S., Miller, R.E., and Wilkin, P. (2001). Phylogenetic analysis of *Ipomoea*, *Argyreia*, *Stictocardia*, and *Turbina* suggests a generalized model of morphological evolution in morning glories. *Syst. Bot. 26*, 585–602.

Marderosian, A.H.D. (1965). Nomenclatural history of the morning glory, *Ipomoea violacea* (L.). *Taxon 14*, 234.

Markham, C.R. (1864). *Contributions towards a grammar and dictionary of Quichua* (London: Trubner & Co.).

Martin, F.W. (1965). Incompatibility in the sweet potato. A review. *Econ. Bot. 19*, 406–415.

Martin, F.W., and Jones, A. (1972). The species of *Ipomoea* closely related to the sweet potato. *Econ. Bot. 26*, 201–215.

Martin, F.W., Ruberté, R.M., and Jones, A. (1974). A wild *Ipomoea* species closely related to the sweet potato. *Econ. Bot. 28*, 287–292.

Matisoo-Smith, E., and Ramírez, J.M. (2010). Human skeletal evidence of Polynesian presence in South America? Metric analyses of six crania from Mocha Island, Chile. *J. Pacif. Archaeol. 1*, 76–88.

Maxted, N., Ford-Lloyd, B.V., Jury, S., Kell, S., and Scholten, M. (2006). Towards a definition of a crop wild relative. *Biodiv. Cons. 15*, 2673–2685.

McCormick, F.A. (1916). Notes on the anatomy of the young tuber of *Ipomoea batatas* Lam. *Bot. Gaz. 61*, 388–398.

McDonald, A. (1987). Revision of *Ipomoea* section *Exogonium* (Choisy) Griseb. (Convolvulaceae). *Brenesia 28*, 41–87.

McDonald, J.A. (1995). Revision of *Ipomoea* section *Leptocallis* (Convolvulaceae). *Harv. Pap. Bot. 1*, 97–122.

McDonald, J.A. (2001). Revision of *Ipomoea* Series *Tyrianthinae* (Convolvulaceae). *Lundellia 4*, 76–93.

McDonald, J.A., and Austin, D.F. (1990). Changes and additions in *Ipomoea* sect. *Batatas*. *Brittonia 42*, 116–120.

McDonald, J.A., and Mabry, T.J. (1992). Phylogenetic systematics of New World *Ipomoea* (Convolvulaceae) based on chloroplast DNA restriction site variation. *Plant Syst. Evol. 180*, 243–259.

Meira, M., Silva, E.P. da, David, J.M., and David, J.P. (2012). Review of the genus *Ipomoea*: traditional uses, chemistry and biological activities. *Rev. Bras. Farmacogn. 22*, 682–713.

Meisner, C.D.F. (1869). Convolvulaceae. In *Flora Brasiliensis*, C.F. Martius, ed. Pp. 199–362.

Merrill, E.D. (1920). Comments on Cook's theory as to the American origin and prehistoric Polynesian distribution of certain economic plants, especially *Hibiscus tiliaceus* Linnaeus. *Philip. J. Sci. 17*, 377–384.

Meyen, F.J.F. (1836). *Grundriss der Pflanzengeographie mit ausführlichen Untersuchungen über das Vaterland, den Anbau und den Nutzen der vorzüglichsten Culturpflanzen* (Berlin: Haude und Spenersche buchhandlung (S.J. Joseephy)).

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE) (New Orleans, LA)*, pp. 1–8.

Miller, R.E., Rausher, M.D., and Manos, P.S. (1999). Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on *ITS* and *Waxy* sequences. *Syst. Bot. 24*, 209–227.

Miller, R.E., Buckley, T.R., and Manos, P.S. (2002). An examination of the monophyly of morning glory taxa using Bayesian phylogenetic inference. *Syst. Biol. 51*, 740–753.

Miller, R.E., McDonald, J.A., and Manos, P.S. (2004). Systematics of *Ipomoea* subgenus Quamoclit (Convolvulaceae) based on *ITS* sequence data and a Bayesian phylogenetic analysis. *Amer. J. Bot. 91*, 1208–1218.

Miquel, F.A.W. (1856). *Flora van Nederlandsch Indië* (Amsterdam: C. G. van der Post).

Mirarab, S., and Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics 31*, i44–i52.

Miryeganeh, M., Takayama, K., Tateishi, Y., and Kajita, T. (2014). Long-distance dispersal by sea-drifted seeds has maintained the global distribution of *Ipomoea pes-caprae* subsp. *brasiliensis* (Convolvulaceae). *PLoS ONE 9*, e91836.

Missouri Botanical Garden (2018). *Tropicos.org*.

Mongalo, N.I., and Makhafola, T.J. (2018). Ethnobotanical knowledge of the lay people of Blouberg area (Pedi tribe), Limpopo Province, South Africa. *J. Ethnobiol. Ethnomed. 14*.

Montenegro, Á., Avis, C., and Weaver, A. (2008). Modelling the prehistoric arrival of the sweet potato in Polynesia. *J. Arch. Sci. 35*, 355–367.

Moonlight, P.W., Ardi, W.H., Padilla, L.A., Chung, K.-F., Fuller, D., Girmansyah, D., Hollands, R., Jara-Muñoz, A., Kiew, R., Leong, W.-C., et al. (2018). Dividing and conquering the fastest-growing genus: towards a natural sectional classification of the mega-diverse genus *Begonia* (Begoniaceae). *Taxon 67*, 267–323.

Moreno Gómez, J. (2010). De las Indias al Mediterráneo: la batata/patata, fruto de Málaga por antonomasia. *Isla de Arriarán 35*, 47–67.

Moreno-Mayar, J.V., Rasmussen, S., Seguin-Orlando, A., Rasmussen, M., Liang, M., Flåm, S.T., Lie, B.A., Gilfillan, G.D., Nielsen, R., Thorsby, E., et al. (2014). Genome-wide ancestry patterns in Rapanui suggest pre-European admixture with Native Americans. *Curr. Biol. 24*, 2518–2525.

Mueller-Dombois, D., and Fosberg, F.R. (1998). *Vegetation of the tropical Pacific islands* (New York: Springer).

Muñoz-Rodríguez, P., Carruthers, T., Wood, J.R.I., Williams, B.R.M., Weitemier, K., Kronmiller, B., Ellis, D., Anglin, N.L., Longway, L., Harris, S.A., et al. (2018). Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to Polynesia. *Curr. Biol. 28*, 1246–1256.e12.

MYcroarray (2015). *MYbaits. In-solution sequence capture for targeted High-Throughput Sequencing*.

Naciri, Y. & Linder, H.P. (2015). Species delimitation and relationships: the dance of the seven veils. *Taxon 64(1)*, 3–16.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Mol. Biol. Evol. 32*, 268–274.

Nishiyama, I. (1959). Collecting the sweet potato and its allied species in USA and Mexico. *Jap. J. Breed. 9*, 73–78.

Nishiyama, I. (1971). Evolution and domestication of the sweet potato. *Bot. Mag. Tokyo 84*, 377–387.

Nishiyama, I., and Teramura, T. (1962). Mexican wild forms of sweet potato. *Econ. Bot. 16*, 305–314.

Nishiyama, I., Fujise, K., Teramura, T., and Miyazaki, T. (1961a). Studies of sweet potato and its related species: I. Comparative investigations on the chromosome numbers and the main plant characters of *Ipomoea* species in section *Batatas*. *Jap. J. Breed. 11*, 37–43.

Nishiyama, I., Fujise, K., Teramura, T., and Miyazaki, T. (1961b). Studies of sweet potato and its related species. II. Physiological and ecological characters of K123. *Jap. J. Breed. 11*, 261–268.

Nishiyama, I., Miyazaki, T., and Sakamoto, S. (1975). Evolutionary autoploidy in the sweet potato (*Ipomoea batatas* (L.) Lam. and its progenitors. *Euphytica 24*, 197–208.

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V. & Dubchak, I. The genome portal of the Department of Energy Joint Genome Instute: 2014 updates. *Nucl. Ac. Res.* 42 (D1): D26-D31.

Nordenskiöld, E. (1931). Origin of the Indian civilizations in South America. *Comp. Ethnogr. Studies 9*.

O'Brien, P.J. (1972). The sweet potato: its origin and dispersal. *Am. Anthropol. 74*, 342–365.

O'Donell, C. (1959). Las especies americanas de *Ipomoea* L., sect. *Quamoclit* (Moench) Griseb. *Lilloa 39*, 19–65.

Ogunwenmo, K.O. (2003). Cotyledon morphology: an aid in identification of *Ipomoea* taxa (Convolvulaceae). *Feddes Repert. 114*, 198–203.

Ogunwenmo, K.O. (2008). Evolutionary and taxonomic studies of *Ipomoea* L. sect. *Involucratae* Bak. & Randle (Convolvulaceae) in Nigeria. *Feddes Repert. 110*, 499–514.

Oracion, M.Z., Niwa, K., and Shiotani, I. (1990). Cytological analysis of tetraploid hybrids between sweet potato and diploid *Ipomoea trifida* (H.B.K.) Don. *Theor. Appl. Genet. 80*, 617–624.

Orjeda, G., Freyre, R., and Iwanaga, M. (1991). Use of *Ipomoea trifida* germ plasm for sweet potato improvement. 3. Development of 4x interspecific hybrids between *Ipomoea batatas* (L.) Lam. (2n = 6x = 90) and *I. trifida* (H.B.K.) G. Don. (2n = 2x = 30) as storage-root initiators for wild species. *Theor. Appl. Genet. 83*, 159–163.

Ozias-Akins, P., and Jarret, R.L. (1994). Nuclear DNA content and ploidy levels in the genus *Ipomoea*. *J. Amer. Soc. Hort. Sci. 119*, 110–115.

Park, I., Yang, S., Kim, W.J., Noh, P., Lee, H.O., and Moon, B.C. (2018). The complete chloroplast genomes of six *Ipomoea* species and indel marker development for the discrimination of authentic *Pharbitidis* semen (seeds of *I. nil* or *I. purpurea*). *Front. Plant Sci. 9*.

Patterson, T.C. (1971). Central Peru: its population and economy. *Archaeology 24*, 307–315.

Patterson, T.C., and Lanning, E.P. (1964). Changing settlement patterns on the central Peruvian coast. *Ñawpa Pacha 2*, 113–123.

Patterson, T.C., and Lanning, E.P. (1966). Correction to Ñawpa Pacha 2. *Ñawpa Pacha 4*, 155.

Patterson, T.C., and Moseley, M.E. (1968). Late pre-ceramic and Early ceramic cultures of the central coast of Peru. *Ñawpa Pacha 6*, 115–133.

Pearsall, D.M. (1993). Contributions of phytolith analysis for reconstructing subsistence: examples from research in Ecuador. In *Current Research in Phytolith Analysis: Applications in Archaeology and Paleoecology* (University of Pennsylvania: Museum Applied Science Center for Archaeology & University of Pennsylvania), pp. 108–124.

Pennington, R.T. & Lavin, M. The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytol. 210(1)*, 25–37.

Perry, L. (2002). Starch granule size and the domestication of manioc (*Manihot esculenta*) and sweet potato (*Ipomoea batatas*). *Econ. Bot. 56*, 335–349.

Pezo-Lanfranco, L., Eggers, S., Petronilho, C., Toso, A., da Rocha Bandeira, D., Von Tersch, M., dos Santos, A.M.P., Ramos da Costa, B., Meyer, R., and Colonese, A.C. (2018). Middle Holocene plant cultivation on the Atlantic Forest coast of Brazil? *Royal Soc. Open Sci. 5*, 180432.

Pickersgill, B., and Heiser Jr., C.B. (1978). Origins and distribution of plants domesticated in the New World tropics. In *Advances in Andean Archaeology*, D.L. Browman, ed. (The Hague: Mouton Publishers), pp. 133–165.

Pigafetta, A. (1524). *Primo viaggio intorno al globo terracqueo ossia ragguaglio della navigazione alle indie orientali per la via d'Occidente* (Milan: Giuseppe Galeazzi).

Piperno, D.R., and Pearsall, D.M. (1998). From small-scale horticulture to the formative period. In *The origins of agriculture in the Lowland Neotropics* (Elsevier), pp. 243–320.

Powell, D.A. (1979). The Convolvulaceae of the Lesser Antilles. *J. Arn. Arb. 60*, 219–271.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol. 26*, 1641–1650.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood trees for large alignments. *PLoS ONE 5*, e9490.

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

Rajapakse, S., Nilmalgoda, S.D., Molnar, M., Ballard, R.E., Austin, D.F., and Bohac, J.R. (2004). Phylogenetic relationships of the sweetpotato in *Ipomoea* series *Batatas* (Convolvulaceae) based on nuclear *β-amylase* gene sequences. *Mol. Phyl. Evol. 30*, 623–632.

Ratan, A. (2009). *Assembly algorithms for Next Generation Sequence data.* Pennsylvania State University.

Raya Rao, S.S., and Leela, M. (2008). Leaf architecture in relation to taxonomy: *Ipomoea* L. *Feddes Repert. 101*, 611–616.

Reichert, E.T. (1913). *The differentiation and specificity of starches in relation to genera, species, etc.; stereochemistry applied to protoplasmic processes and products, and as strictly scientific basis for the classification of plants and animals* (Washington, DC: The Carnegie Institution of Washington).

Reinhardt, E. (1911). *Kulturgeschichte der Nutzpflanzen* (München: Berlag von Ernst Reinhardt).

Reiseberg, L.H., and Soltis, D.E. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trends Pl. 5*, 65–84.

Rensch, K.H. (1991). Polynesian plant names: linguistic analysis and ethnobotany, expectations and limitations. In *Islands, plants, and Polynesians: an introduction to Polynesian ethnobotany. Proceedings of a symposium*, P.A. Cox, S.A. Banack, and Brigham Young University--Hawaii Campus, eds. (Portland, Or: Dioscorides Press), pp. 97–111.

Richardson, J.B. (1972). The pre-Columbian distribution of the bottle gourd (*Lagenaria siceraria*): a re-evaluation. *Econ. Bot. 26*, 265–273.

Ridley, H.N. (1930). *The dispersal of plants throughout the world* (Lloyds Bank Buildings, Ashford, Kent: L. Reeve & Co., Ltd).

Rieseberg, L.H. & Brouillet, K. (1994). Are many plant species paraphyletic? *Taxon 43*, 21–32.

Roberty, G. (1952). Genera Convolvulacearum. *Candollea 14*, 11–60.

Rochebrune, A.-T. (1879). Recherches d'ethnographie botanique sur la flore des sépultures péruviennes d'Ancon. *Act. Soc. Lin. Bordeaux 33*, 343–358.

Ronquist, F., and Huelsenbeck, J.P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics 19*, 1572–1574.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol. 61*, 539–542.

Rosendahl, P., and Yen, D.E. (1971). Fossil sweet potato remains from Hawaii. *J. Pol. Soc. 80*, 378–385.

Roullier, C., Rossel, G., Tay, D., Mckey, D., and Lebot, V. (2011). Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. *Mol. Ecol. 20*, 3963–3977.

Roullier, C., Duputié, A., Wennekes, P., Benoit, L., Fernández Bringas, V.M., Rossel, G., Tay, D., McKey, D., and Lebot, V. (2013a). Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). *PLoS ONE 8*, e62707.

Roullier, C., Kambouo, R., Paofa, J., McKey, D., and Lebot, V. (2013b). On the origin of sweet potato (*Ipomoea batatas* (L.) Lam.) genetic diversity in New Guinea, a secondary centre of diversity. *Heredity 110*, 594–604.

Roullier, C., Benoit, L., McKey, D.B., and Lebot, V. (2013c). Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. *Proc. Natl. Acad. Sci. U.S.A. 110*, 2205–2210.

Ruby, J.G., Bellare, P., and DeRisi, J.L. (2013). PRICE: software for the targeted assembly of components of (meta) genomic sequence data. *G3: Genes, Genomes, Genetics 3*, 865–880.

Rumphius, G.E. (1747). *Herbarium amboinense: plurimas conplectens arbores, frutices, herbas, plantas terrestres & aquaticas, quae in Amboina et adjacentibus reperiuntur insulis* (Amstelaedami: Fransicum Changuion, Joannem Catuffe, Hermannum Uytwerf).

Sachet, M.-H. (1975). *Flora of the Marquesas* (Washington: Smithsonian Institution Press).

Safford, W.E. (1917). Food-plants and textiles of ancient America. In *Proceedings of the 19th International Congress of Americanists*, F.W. Hodge, ed. (Washington, DC), pp. 12–20.

Sagra, R. de la (1850). *Historia física, política y natural de la isla de Cuba* (Paris: A. Bertrand).

Sardos, J., Rouard, M., Hueber, Y., Cenci, A., Hyma, K.E., van den Houwe, I., Hribova, E., Courtois, B., and Roux, N. (2016). A genome-wide association study on the seedless phenotype in banana (*Musa* spp.) reveals the potential of a selected panel to detect candidate genes in a vegetatively propagated crop. *PLoS ONE 11*, e0154448.

Särkinen, T., Bohs, L., Olmstead, R.G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol. 13:214*.

Sauer, C.O. (1950). Cultivated plants of South and Central America. In *Handbook of South American Indians*, J.H. Steward, ed. (Washington, DC: United States Government Printing Office), pp. 487–544.

Scotland, R.W., and Wood, J.R.I. (2012). Accelerating the pace of taxonomy. *Trends Ecol. Evol. 27*, 415–416.

Seemann, B. (1865). *Flora Vitiensis: a description of the plants of the Viti or Fiji islands, with an account of their history, uses, and properties* (London: L. Reeve).

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E., and Small, R.L. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Amer. J. Bot. 92*, 142–166.

Shaw, J., Lickey, E.B., Schilling, E.E., and Small, R.L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Amer. J. Bot. 94*, 275–288.

Shaw, J., Shafer, H.L., Leonard, O.R., Kovach, M.J., Schorr, M., and Morris, A.B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *Amer. J. Bot. 101*, 1987–2004.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Syst. Biol. 51*, 492–508.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. System Biol. 7*, 539.

Silvestre Domingo, S. (2004). *Ipomoea triloba* L. (convolvulaceae) una nueva especie alóctona para la península. *Lagascalia 24*, 63–66.

Simão-Bianchini, R., and Rubens Pirani, J. (2005). Duas novas espécies de Convolvulaceae de Minas Gerais, Brasil. *Hoehnea 32*, 295–300.

Simon, E. (1914). Introduction of the sweet potato. *Trans. As. Soc. Jap. 42*, 711–724.

Smith, A.C. (1991). *Flora Vitiensis nova: a new Flora of Fiji (spermatophytes only)* (Lawai, Kauai, Hawaii: Pacific Tropical Botanical Garden,).

Smith, R.A., and Rausher, M.D. (2007). Close clustering of anthers and stigma in *Ipomoea hederacea* enhances prezygotic isolation from *Ipomoea purpurea*. *New Phytol. 173*, 641–647.

Smith Jr., C.E. (1968). The New World centers of origin of cultivated plants and the archaeological evidence. *Econ. Bot. 22*, 253–266.

South African National Biodiversity Institute (2018). *SANBI*.

Spate, O.H.K. (2004). *The Spanish Lake. The Pacific since Magellan* (Canberra: ANU E Press).

Srisuwan, S., Sihachakr, D., and Siljak-Yakovlev, S. (2006). The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. *Pl. Sci. (Elsevier) 171*, 424–433.

Srivastava, D., and Shukla, K. (2015). *Ipomoea cairica*: a medicinal weed with promising health benefits. *IJIRR 2*, 687–694.

Srivastava, G., Mehrotra, R.C., and Dilcher, D.L. (2018). Paleocene *Ipomoea* (Convolvulaceae) from India with implications for an East Gondwana origin of Convolvulaceae. *Proc. Natl. Acad. Sci. U.S.A. 115*, 6028–6033.

St. John, H., and Mason, L.E. (1953). Vernacular names of the plants of Bikini, Marshall Islands. Pacific Studies 12. *Pac. Sci. 7*, 165–168.

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

Staples, G.W. (2009). *Merremia pacifica* (Convolvulaceae) recharacterised, with notes on other Pacific species. *Kew Bull. 64*, 333–338.

Staples, G.W. (2010). A checklist of *Merremia* (Convolvulaceae) in Australasia and the Pacific. *Gard. Bull. Sing. 61*, 483–522.

Staples, G.W., and Traiperm, P. (2017). A nomenclatural review of *Argyreia* (Convolvulaceae). *Taxon 66*, 445–477.

Staples, G.W., Wiersema, J.H., Chambers, N.A., and Austin, D.F. (2005a). The restoration of *Ipomoea muricata* (L.) Jacq. (Convolvulaceae). *Taxon 54*, 1075–1079.

Staples, G.W., Na Songkhla, B., Kunhwasi, C., and Traiperm, P. (2005b). Annotated checklist of Thai Convolvulaceae. *Thay Forest Bull. (Bot.) 33*, 171–184.

Staples, G.W., Traiperm, P., Sugau, J.B., and Pornpongrungrueng, P. (2014). *Ipomoea cambodiensis* Gagnep. & Courchet (Convolvulaceae) recharacterised with notes on its distribution and ecology. *Adansonia 36*, 351–357.

Stefanovic, S., Krueger, L., and Olmstead, R.G. (2002). Monophyly of the Convolvulaceae and circumscription of their major lineages based on DNA sequences of multiple chloroplast loci. *Amer. J. Bot. 89*, 1510–1522.

Stefanovic, S., Austin, D.F., and Olmstead, R.G. (2003). Classification of Convolvulaceae: a phylogenetic approach. *Syst. Bot. 28*, 791–806.

Stokes, J.F.G. (1932). Spaniards and the sweet potato in Hawaii and Hawaiian-American contacts. *Am. Anthropol. 34*, 594–600.

Storey, A.A., and Matisoo-Smith, E.A. (2014). No evidence against Polynesian dispersal of chickens to pre-Columbian South America. *Proc. Natl. Acad. Sci. U.S.A. 111*, E3583–E3583.

Storey, A.A., Quiroz, D., Beavan, N., and Matisoo-Smith, E. (2013). Polynesian chickens in the New World: a detailed application of a commensal approach: Polynesian chickens in the New World. *Archaeol. Ocean. 48*, 101–119.

Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., and Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-Generation Sequencing for plant systematics. *Amer. J. Bot. 99*, 349–364.

Swamy, J., and Ramana, P.V. (2018). Notes on the taxonomy and distribution of the Bengal morning glory *Ipomoea rubens* Choisy (Convolvulaceae) in India. *J. Threat. Taxa 10*, 11992.

Swofford, D.L. (2002). *Phylogenetic Analysis Using Parsimony (*and other methods)* (Sunderland, Massachusetts: Sinauer Associates).

Sykes, W.R. (1981). The vegetation of Late, Tonga. *Allertonia 2*, 323–353.

Sykes, B.C., Leiboff, A., Low-Beer, J., Tetzner, S., and Richards, M. (1995). The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet. 57*, 1463–1475.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol. 56*, 564–577.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol. 30*, 2725–2729.

Tello, J. (1924). Arte antiguo peruano. *Inca, Rev. Est. Antropol. 2*.

The International Potato Center (2006). *The world sweetpotato atlas* (The International Potato Center).

Thiers, B. (2018). *Index Herbariorum: a global directory of public herbaria and associated staff*.

Thomson, V.A., Lebrasseur, O., Austin, J.J., Hunt, T.L., Burney, D.A., Denham, T., Rawlence, N.J., Wood, J.R., Gongora, J., Girdland Flink, L., et al. (2014). Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. *Proc. Natl. Acad. Sci. U.S.A. 111*, 4826–4831.

Thorsby, E. (2012). The Polynesian gene pool: an early contribution by Amerindians to Easter Island. *Phyl. Trans. R. Soc. B: Biol. Sci. 367*, 812–819.

Thulin, M. (2003). Notes on *Convolvulus*, *Astripomoea*, *Ipomoea* and *Merremia* (Convolvulaceae) from the Horn of Africa. *Nordic J. Bot. 23*, 629–640.

Thunberg, C.P. (1784). *Systema Vegetabilium secundum classes, ordines, genera, species, cum characteribus et differentiis* (Gottingae: Typis et Impensis Jo. Christ. Dieterich).

Ting, Y.C., and Kehr, A.E. (1953). Meiotic studies in the sweet potato (*Ipomoea batatas* Lam.). *J. Hered. 44*, 207–211.

Ting, Y.C., Kehr, A.E., and Miller, J.C. (1957). A cytological study of the sweet potato plant *Ipomoea batatas* (L.) Lam. and its related species. *Amer. Naturalist 91*, 197–203.

Towle, M.A. (1958). *The ethnobotany of pre-Columbian Peru as evidenced by archaeological remains.* Columbia University.

Turei, M., and Kapiti, P. (1912). The history of "Horouta" canoe and the introduction of the kumara into New Zealand. *J. Pol. Soc. 21*, 152–163.

Ugent, D., and Peterson, L.W. (1988). Archaeological remains of potato and sweet potato in Peru. *CIP Circular 16*, 1–10.

Ugent, D., Pozorski, S., and Pozorski, T. (1981). Prehistoric remains of the sweetpotato from the Casma valley of Peru [*Ipomoea batatas*]. *Phytologia 49*, 401–415.

Valva, V.L., and Sabato, S. (1983). Nomenclature and typification of *Ipomoea imperati* (Convolvulaceae). *Taxon 32*, 110.

Van Ooststroom, S.J. (1953). Convolvulaceae. In *Flora Malesiana* (Djakarta: Noordhoff-Kolff N.V.), pp. 388–512.

Vasconcelos, L.V., Simão-Bianchini, R., and França, F. (2016). Two new species of *Ipomoea* (Convolvulaceae) from the Chapada Diamantina of Bahia, Brazil. *Brittonia 68*, 142–147.

Verdcourt, B. (1957). Typification of the subdivisions of *Ipomoea* L. (Convolvulaceae) with particular regard to the East African species. *Taxon 6*, 150–152.

Von Jacquin, N.J. (1790). Observationes Botanicae. In *Collectanea* (Vindobonae: Officina Wappleriana), pp. 93–226.

Wallin, P., Stevenson, C., and Ladefoged, T.N. (2005). Sweet potato production on Rapa Nui. In *The sweet potato in Oceania: a reappraisal* (Sidney: University of Sidney), pp. 85–88.

Watling, J., Shock, M.P., Mongeló, G.Z., Almeida, F.O., Kater, T., De Oliveira, P.E., and Neves, E.G. (2018). Direct archaeological evidence for Southwestern Amazonia as an early plant domestication and food production centre. *PLoS ONE 13*, e0199868.

Watson, J.B. (1977). Piggs, fodder and the Jones effect in Postipomoean New Guinea. *Ethnology 16*, 57–70.

Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., and Liston, A. (2014). Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *App. Pl. Sci. 2*, 1400042.

Welman, W.G., and Meeuse, A.D.J. (1998). Revised key to *Ipomoea* (Convolvulaceae) in southern Africa. *Bothalia 28*, 19–24.

Wesolowski, V., Ferraz Mendonça de Souza, S.M., Reinhard, K.J., and Ceccantini, G. (2010). Evaluating microfossil content of dental calculus from Brazilian sambaquis. *J. Arch. Sci. 37*, 1326–1338.

Whistler, W.A. (1988). Ethnobotany of Tokelau: the plants, their Tokelau names and their uses. *Econ. Bot. 42*, 155–176.

Whistler, W.A. (1990). The other Polynesian gourd. *Pac. Sci. 44*, 115–122.

Whitaker, T.W., and Carter, G.F. (1954). Oceanic drift of gourds. Experimental observations. *Amer. J. Bot. 41*, 697.

Wilf, P., Carvalho, M.R., Gandolfa, M.A., and Cuneo, R.N. (2017). Eocene lantern fruits from Gondwanan Patagonia and the early origins of Solanaceae. *Science 355*, 71–75.

Wilkin, P. (1999). A morphological cladistic analysis of the *Ipomoeeae* (Convolvulaceae). *Kew Bull. 54*, 853–876.

Williams, B.R.M., Mitchell, T.C., Wood, J.R.I., Harris, D.J., Scotland, R.W., and Carine, M.A. (2014). Integrating DNA barcode data in a monographic study of *Convolvulus*. *Taxon 63*, 1287–1306.

Wittmack, L. (1887). XIII. Plants and Fruits. In *The necropolis of Ancon in Peru: a contribution to our knowledge of the culture and industries of the Empire of the Incas*

being the results of excavations made on the spot, W. Reiss, and A. Stübel, eds. (Berlin; New York: A. Asher & Co.). Plates 105-107.

Wollstein, A., Lao, O., Becker, C., Brauer, S., Trent, R.J., Nürnberg, P., Stoneking, M., and Kayser, M. (2010). Demographic history of Oceania inferred from genome-wide data. *Curr. Biol. 20*, 1983–1992.

Wood, J.R.I., and Scotland, R.W. (2017a). Notes on *Ipomoea* L. (Convolvulaceae) in Cuba and neighbouring islands with a checklist of species found in Cuba. *Kew Bull. 72*.

Wood, J.R.I., and Scotland, R.W. (2017b). Misapplied names, synonyms and new species of *Ipomoea* (Convolvulaceae) from South America. *Kew Bull. 72*.

Wood, J.R.I., and Scotland, R.W. (2017c). Notes on *Ipomoea* (Convolvulaceae) from the Amazonian periphery. *Kew Bull. 72*.

Wood, J., R.M. Williams, B., Mitchell, T.C., Carine, M., Harris, D., and Scotland, R. (2015a). A foundation monograph of *Convolvulus* L. (Convolvulaceae). *Phytokeys 51*, 1–282.

Wood, J.R.I., Carine, M.A., Harris, D., Wilkin, P., Williams, B., and Scotland, R.W. (2015b). *Ipomoea* (Convolvulaceae) in Bolivia. *Kew Bull. 70*, 71.

Wood, J.R.I., de Arrúa, R.D., de Rojas, G.D., and Scotland, R.W. (2016a). Two over-looked species of *Ipomoea* L. (Convolvulaceae) from Paraguay. *Kew Bull. 71*.

Wood, J.R.I., Urbanetz, C., and Scotland, R.W. (2016b). *Ipomoea pantanalensis*, a new species of *Ipomoea* L. (Convolvulaceae) from the Pantanal, Brazil. *Kew Bull. 71*.

Wood, J.R.I., Buril, M.T., and Scotland, R.W. (2017a). Remarkable disjunctions in *Ipomoea* species (Convolvulaceae) from NE Brazil and Central America and their taxonomic implications. *Kew Bull. 72*.

Wood, J.R.I., Muñoz-Rodríguez, P., Degen, R., and Scotland, R.W. (2017b). New species of *Ipomoea* (Convolvulaceae) from South America. *Phytokeys 88*, 1–38.

Wood, J.R.I., Vasconcelos, L.V., Simão-Bianchini, R., and Scotland, R.W. (2017c). New species of *Ipomoea* (Convolvulaceae) from Bahia. *Kew Bull. 72*.

Yan, L., Lai, X., Li, X., Wei, C., Tan, X., and Zhang, Y. (2015). Analyses of the complete genome and gene expression of chloroplast of sweet potato [*Ipomoea batata*]. *PLoS ONE 10*, e0124083.

Yang, J., Moeinzadeh, M., Kuhl, H., Helmuth, J., Xiao, P., Haas, S., Liu, G., Zheng, J., Sun, Z., Fan, W., et al. (2017). Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nat. Plants 3*, 696–703.

Yang, Y. & Smith, S.A. (2014). Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol. 31*, 3081–3092.

Yatsckievych, G., and Mason Jr., C.T. (1984). A taxonomic study of *Ipomoea tenuiloba* Torrey (Convolvulaceae) with notes on related species. *Madroño 31*, 102–108.

Yen, D.E. (1960). The sweet potato in the Pacific: the propagation of the plant in relation to its distribution. *J. Pol. Soc. 69*, 368–375.

Yen, D.E. (1971). Construction of the hypothesis for distribution of the sweet potato. In *Man across the sea. Problems of pre-Columbian contacts* (Austin: The University of Texas Printing Division), p. 328–42.

Yen, D.E. (1974). *The sweet potato and Oceania. An essay in ethnobotany* (Honolulu, Hawaii: Bishop Museum Press).

Zhang, D., Rossel, G., Kriegner, A., and Hijmans, R. (2004). AFLP assessment of diversity in sweetpotato from Latin America and the Pacific region: its implications on the dispersal of the crop. *Genet. Resources Crop Evol. 51*, 115–120.

Zizka, G. (1991). *Flowering plants of Easter Island* (Frankfurt: Palmengarten, Stadt Frankfurt am Main).

CREDITS OF THE IMAGES

All images included in this thesis were generated or taken by the author except:

- **Figure 2.7** in page 27. *Ipomoea batatas* specimen collected in the Society Islands by Joseph Banks and Daniel Solander, held at the Natural History Museum in London (BM001040624). Reproduced here under the terms of the Creative Commons CC0 1.0 Universal License (*https://creativecommons.org/publicdomain/zero/1.0/*).
- **Figure 3.15** in page 70. *Ipomoea setosa* specimen in Bolivia. © 2015 John R.I. Wood, reproduced here with permission.
- **Figure 3.20** in page 78. Storage roots of *Ipomoea lilloana*. © 2018 John R.I. Wood, reproduced here with permission.
- **Figures 4.2, 4.3 and 4.10** in pages 94, 95 and 111. Several species of *Ipomoea* closely related to the sweet potato. © 2015 John R.I. Wood and Robert W. Scotland, reproduced here with permission.
- **Figure 4.4** in page 97. © 2015 Khoury, Heider, Castañeda-Álvarez, Achicanoy, Sosa, Miller, Scotland, Wood, Rossel, Eserman, Jarret, Yencho, Bernau, Juarez, Sotelo, de Haan and Struik. Image published in *Frontiers in Plant Science* 6: 251 (*https://doi.org/10.3389/fpls.2015.00251*) reproduced here under the terms of the Creative Commons Attribution License (CC BY) (*http://creativecommons.org/licenses/by/4.0/*).
- **Figures 4.16** (page 117), **4.18** (page 119), **4.21** (page 122), **4.22** (page 124) and **5.4–5.9** (pages 161–165). © 2018 Elsevier Ltd. Images published in *Current Biology* 28(8): P1246-P1256 (*https://doi.org/10.1016/j.cub.2018.03.020*). Included here in agreement with Elsevier's policy that allows authors of papers to include the content of a paper in a thesis, provided it is not published commercially (*https://www.elsevier.com/about/our-business/policies/copyright#Author-rights*).
  Picture of seeds in Figure 5.4 taken by John Baker.
- **Figure 5.2** in page 151. Storage roots of different species of morning glories. © 2018 Eserman, Jarret and Leebens-Mack. Image published in *BMC Plant Biology* 18: 95 (*https://doi.org/10.1186/s12870-018-1307-4*) and reproduced here under the terms of the Creative Commons Attribution 4.0 International License (CC BY) (*http://creativecommons.org/licenses/by/4.0/*).

INDEX OF SCIENTIFIC NAMES

APPENDIX 1 | GENOMIC DIVERSITY IN SWEET POTATO

Table 1. Descriptive statistics for the analysis of diversity between sweet potato lineages using chloroplast data.

	LINEAGE			Statistic	Std. Error
GENETIC DISTANCE	1	Mean		.00025285	.000013930
		95% Confidence Interval for Mean	Lower Bound	.00022540	
			Upper Bound	.00028029	
		5% Trimmed Mean		.00023399	
		Median		.00019360	
		Std. Deviation		.000211715	
		Minimum		.000000	
		Maximum		.001057	
		Range		.001057	
		Interquartile Range		.000156	
		Skewness		1.499	.160
		Kurtosis		1.494	.319
	2	Mean		.00015818	.000003265
		95% Confidence Interval for Mean	Lower Bound	.00015177	
			Upper Bound	.00016458	
		5% Trimmed Mean		.00015042	
		Median		.00013800	
		Std. Deviation		.000107346	
		Minimum		.000000	
		Maximum		.000615	
		Range		.000615	
		Interquartile Range		.000125	
		Skewness		1.135	.074
		Kurtosis		1.219	.149

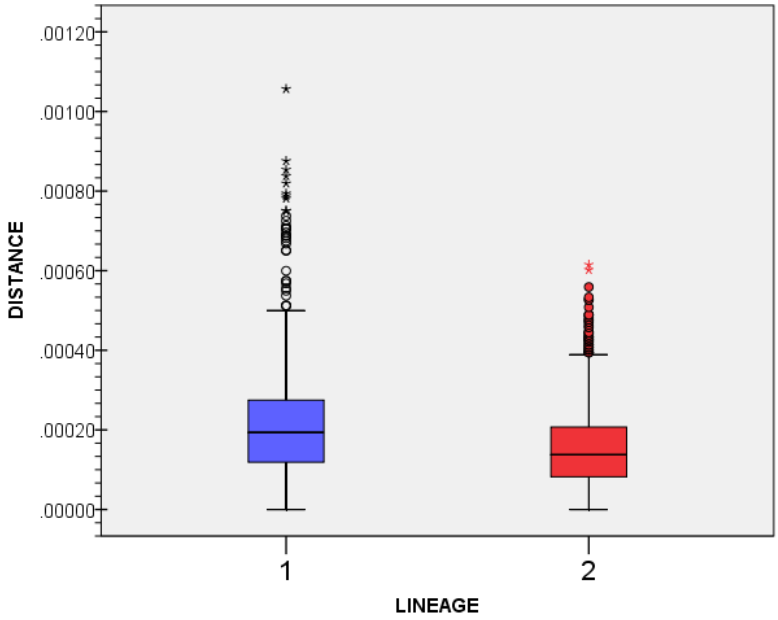


Figure 1. Average genetic diversity in the two sweet potato lineages. Pairwise distances (p-values) between pairs of accessions in each lineage, whole chloroplast genome.

Table 2. Test of normality of genetic distances inferred from whole chloroplast genome. The data does not follow a Normal distribution (p-value < 0.0005).

		Kolmogorov-Smirnov			Shapiro-Wilk		
	LINEAGE	Statistic	df	Sig.	Statistic	df	Sig.
GENETIC DISTANCE	1	.224	231	.000	.815	231	.000
	2	.118	1081	.000	.919	1081	.000

Table 3. Descriptive statistics for the analysis of diversity between sweet potato lineages using nuclear data.

	LINEAGE			Statistic	Std. Error
GENETIC DISTANCE	**1**	Mean		.00868406	.000085673
		95% Confidence Interval for Mean	Lower bound	.00851525	
			Upper bound	.00885286	
		5% Trimmed Mean		.00867910	
		Median		.00868700	
		Std. Deviation		.001302120	
		Minimum		.005222	
		Maximum		.012055	
		Range		.006833	
		Interquartile Range		.001890	
		Skewness		.032	.160
		Kurtosis		-.419	.319
	2	Mean		.00817708	.000038326
		95% Confidence Interval for Mean	Lower bound	.00810187	
			Upper bound	.00825229	
		5% Trimmed Mean		.00818254	
		Median		.00816900	
		Std. Deviation		.001205911	
		Minimum		.004887	
		Maximum		.012419	
		Range		.007532	
		Interquartile Range		.001628	
		Skewness		-.006	.078
		Kurtosis		-.057	.155

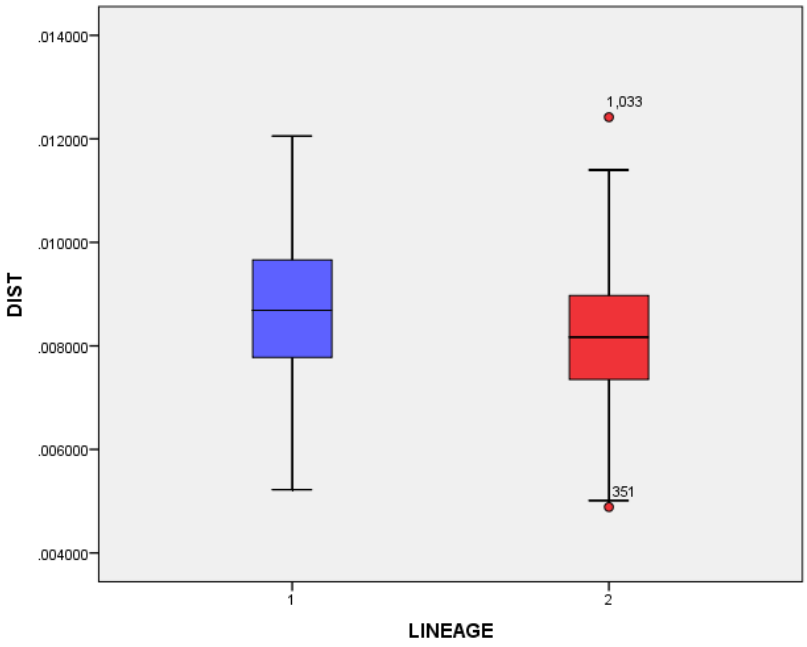


Figure 2. Average genetic diversity in the two sweet potato lineages. Pairwise distances (p-values) between pairs of accessions in each lineage, nuclear genes.

Table 4. Test of normality of genetic distances inferred from nuclear data. The data follows a Normal distribution (p-value > 0.05).

		Kolmogorov-Smirnov			**Shapiro-Wilk**		
	LINEAGE	Statistic	df	Sig.	Statistic	df	Sig.
GENETIC	**1**	.044	231	.200*	.995	231	.632
DISTANCE	**2**	.016	990	.200*	.998	990	.263

Group Statistics (CHLOROPLAST DATA)

	LINEAGE	N	Mean	Std. Deviation
GENETIC DISTANCE	1	231	.00025285	.000211715
	2	1081	.00015818	.000107346

NULL HYPOTHESIS	TEST	Sig.	Decision
"The distribution of average distances is the same across categories of lineage"	Independent samples Mann-Whitney U test	.000	Reject the null hypothesis

Group Statistics (NUCLEAR DATA)

	LIN.	N	Mean	Std. Deviation	Std. Error Mean
GENETIC DISTANCE	1	231	.00868406	.001302120	.000085673
	2	990	.00817708	.001205911	.000038326

Independent Samples Test (NUCLEAR DATA)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
GENETIC DISTANCE	Equal variances assumed	3.712	.054	5.666	1219	.000	.000506981	.000089484	.000331422	.000682539
	Equal variances not assumed			5.402	328.213	.000	.000506981	.000093855	.000322347	.000691614

APPENDIX 2 | INFRAGENERIC CLASSIFICATION OF THE GENUS *IPOMOEA*

Infrageneric classification of the genus *Ipomoea* as outlined by Austin (1979, 1980, 1996)

Subgen. *Ipomoea* L.

Sect. *Ipomoea* [Type species: *I. pes-tigridis* L.]

Ser. *Ipomoea*

Ser. *Involucratae* (Baker & Rendle) D.F.Austin [Type species: *I. involucrata* Beauv.]

Sect. Pharbitis (Choisy) Griseb. [Type species: *I. purpurea* (L.) Roth

Ser. *Pharbitis*

Ser. *Heterophyllae* (House) D.F.Austin [Type species: *I. heterophylla* Ortega]

Ser. *Tyrianthinae* (House) D.Austin [Type species: *I. tyrianthina* Lindley (= *I. orizabensis* Ledeb. ex Steud.)]

Subgen. *Quamoclit* (Moench) Clarke

Sect. *Calonyction* (Choisy) Griseb. [Type species: *I. alba* L.]

Sect. *Dasychaetia* Hall.f. [Type species: *I. linosepala* Hall.f.]

Sect. *Exogonium* (Choisy) Griseb. [Type species: *I. bracteata* Cav.]

* Sect. *Leptocallis* (G.Don) J.A.McDonald [Type species: *I. pedatisecta* M.Martens & Galeotti (= *I. ternifolia* Cav.)]

** Sect. *Microsepalae* (House) D.F.Austin [Type species: *I. microsepala* Benth.]

Sect. *Mina* (Cerv.) Griseb.

Sect. *Orthipomoea* Choisy in DC. [Type species: *I. polymorpha* Roem. & Sch.)

Sect. *Quamoclit* [Type species: *I. coccinea* L.]

Sect. *Tricolor* [Type species: *I. tricolor* Cav.]

Subgen. *Eriospermum* Hallier f.

Sect. *Eriospermum* [Type species: *I. digitata* L.]

Ser. *Anisomeres* (House) D.F.Austin [Type species: *I. anisomeres* Robins & Bartl.]

Ser. *Arborescentes* (Choisy) D.F.Austin [Type species: *I. arborescens* G.Don]

*** Ser. *Batatas* (Choisy) Griseb. [Type species: *I. batatas* (L.) Lam.]

Ser. *Bombycospermum* (C.Presl.) D.F.Austin [Type species *I. bombycina* (Choisy) Benth.]

Ser. *Eriospermum* (Hallier f.) D.F.Austin

Ser. *Jalapae* (House) D.F.Austin [Type species: *I. jalapa* (L.) Pursh.]

Ser. *Mirandinae* D.F.Austin [Type species: *I. mirandina* (Pittier) O'Donell]

Ser. *Setosae* (House) D.F.Austin [Type species: *I. setosa* Ker Gawl.

Ser. *Suffruticosae* (Choisy) D.F.Austin [Type species: *I. suffruticosa* Burch.]

Ser. *Dactylophyllae* (House) D.F.Austin [Type species: *I. horsfalliae* Hook.]

Sect. *Acmostemon* (Pilger) Verdcourt [Type species: *I. verbascoidea* Choisy]

Sect. *Poliothamnus* (Hallier f.) Verdc. [Type species: *I. hildebrandtii* Vatke]

Sect. *Xerophyta* (Baker & Rendle) D.F.Austin [Type species: *I. donaldsonii* Rendle]

Sect. *Erpipomoea* Choisy [Type species: *I. pes-caprae* (L.) R.Br.

* Recognised as Ser. *Pedatisectae* (House) D.F.Austin in 1979 and 1980.

** Recognised at the series level in 1979 and 1980.

*** Recognised at the section level in Subgen. *Quamoclit* in 1979 and 1980.

APPENDIX 3 | ADDITIONAL PHYLOGENIES

CONTENT

A-G. Approximate Maximum Likelihood of the main clades of *Ipomoea* inferred from *ITS* sequences.

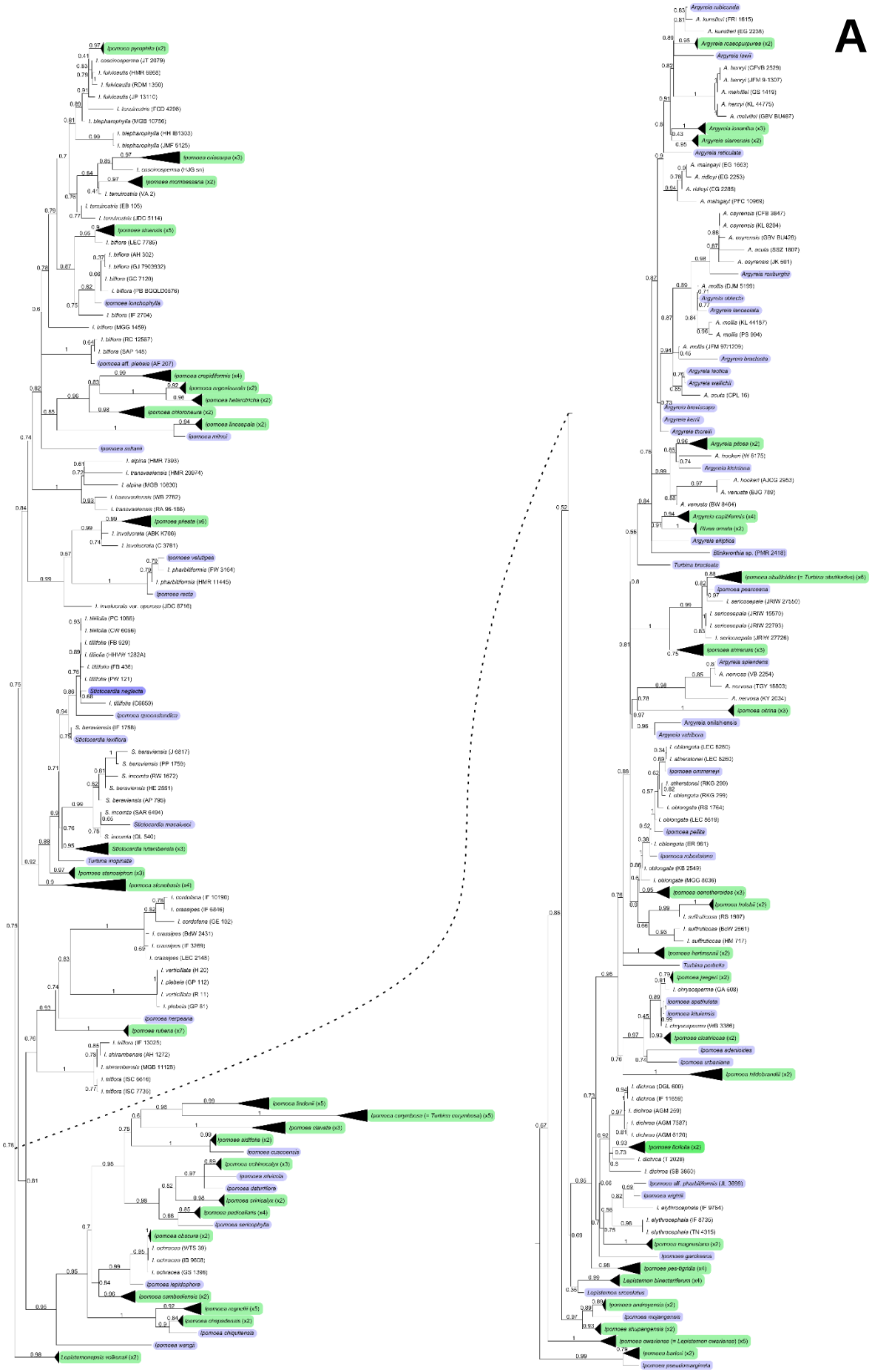
A, OW CLADE; B, CLADE A1; C, CLADE A2; D, CLADE B (left, B1; right, B2); E, CLADE C; F, CLADE D; G, CLADE E. Green circles in all phylogenies indicate 100% support (Shimodaira test, 1,000 replicates). Green boxes indicate monophyletic species and purple boxes, species represented by a single specimen, the monophyly of which cannot be assessed.

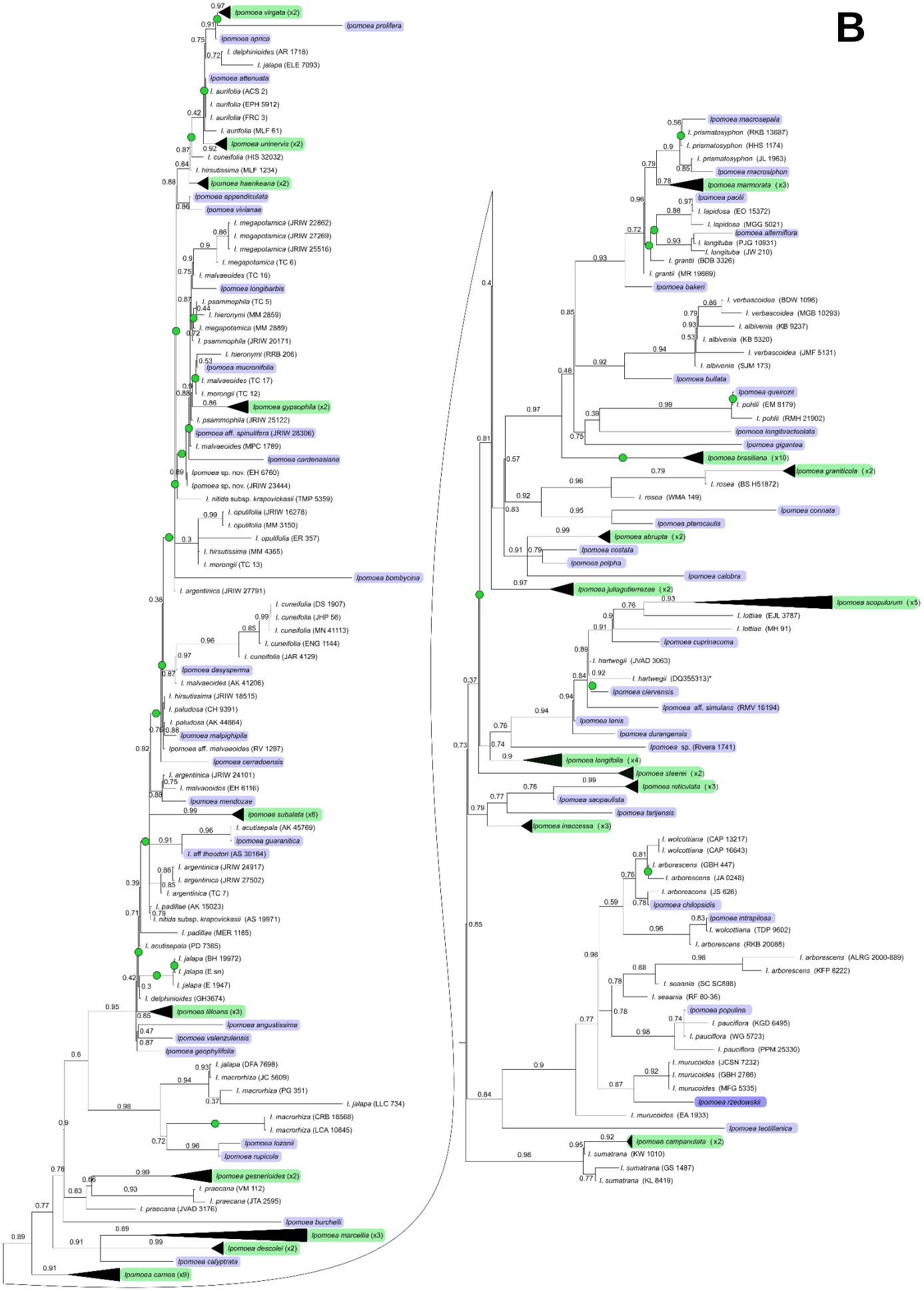
H. Nuclear phylogeny of CLADE A3 (Batatas) with geographical patterns

Approximate Maximum Likelihood inferred from concatenated alignments of nuclear coding regions. Green circles indicate 100% support (Shimodaira test, 1,000 replicates). Yellow shade, South America (incl. Colombia); red, Venezuela; green, North America; light blue, Central America and the Caribbean, purple, Old World.

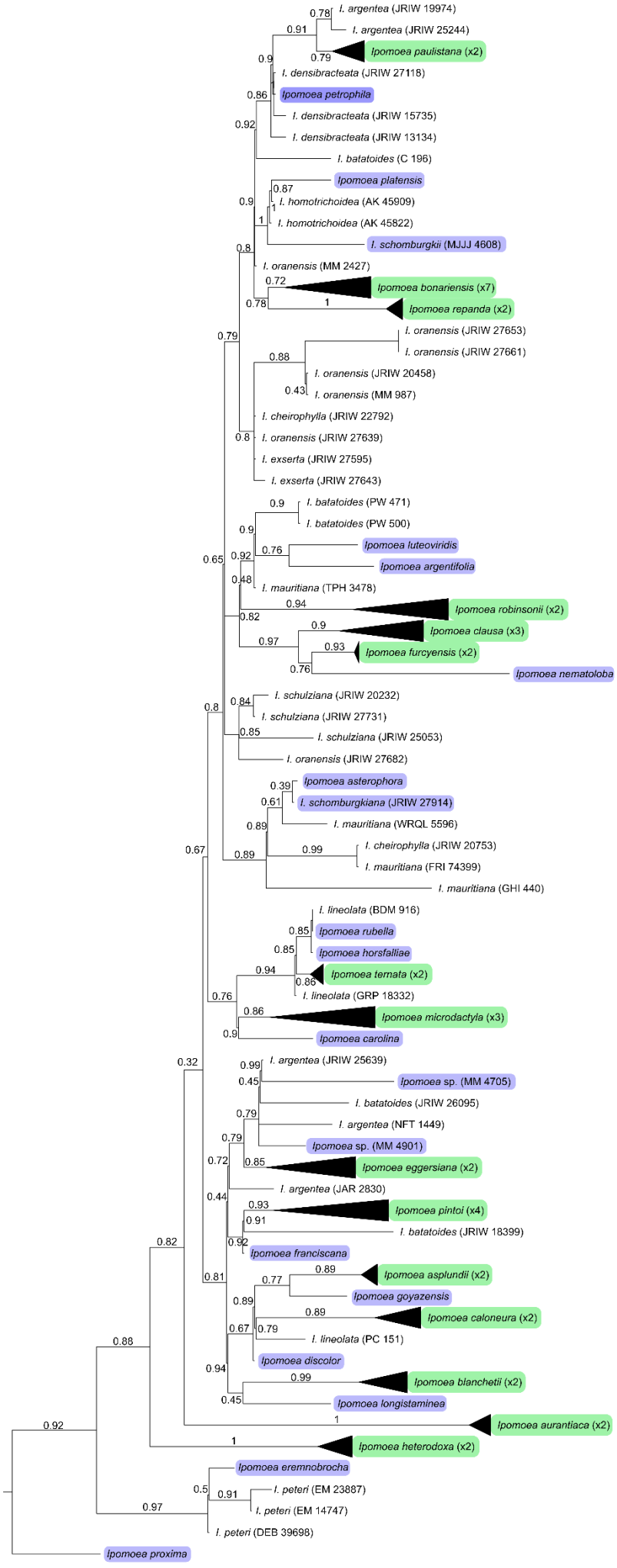
I. Time-calibrated chloroplast phylogeny of 94 specimens of *I. trifida* and *I. batatas*, including Banks and Solander specimen.

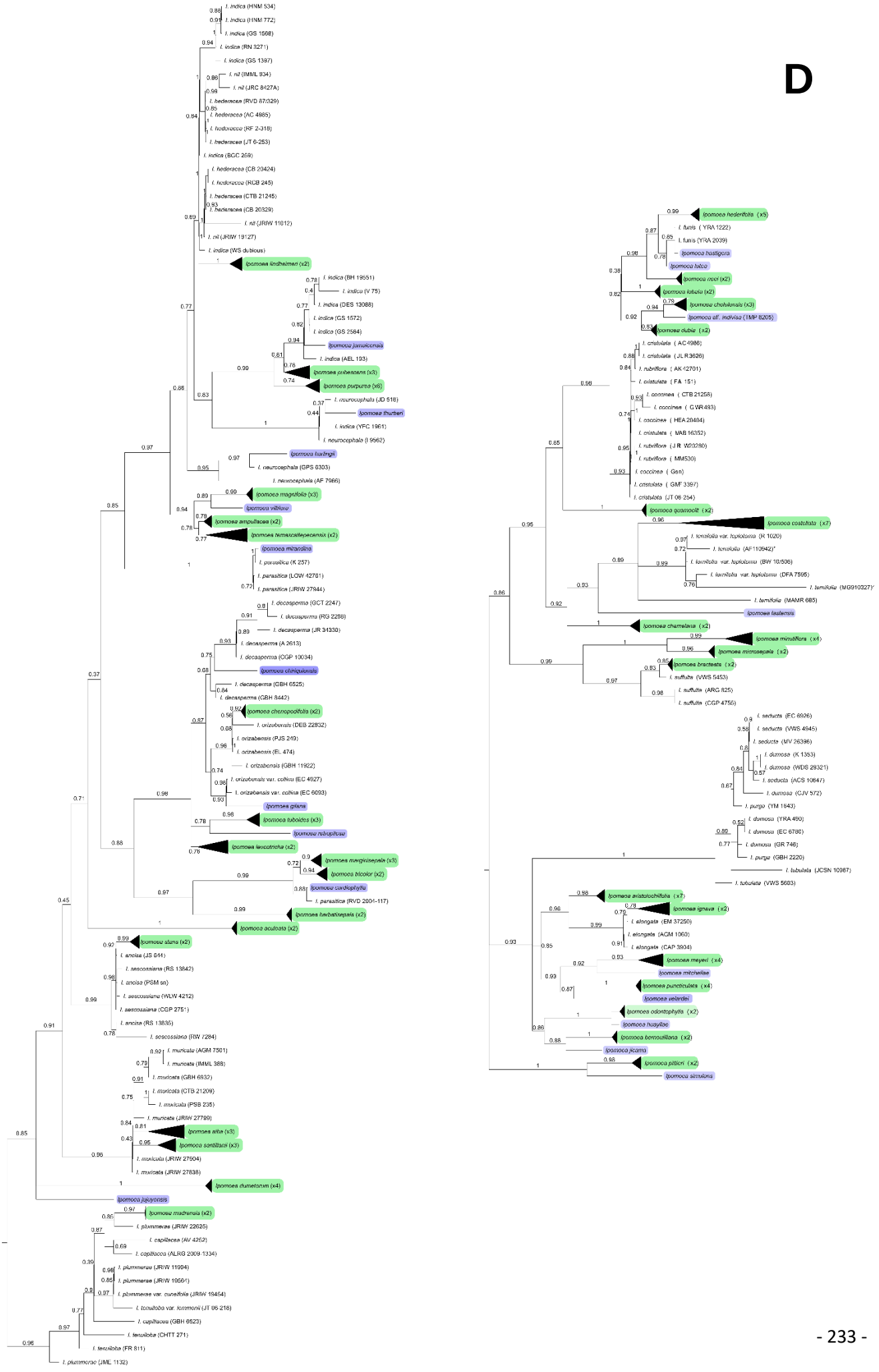
Tree generated from a multispecies coalescent analysis of 94 specimens of *Ipomoea trifida* and *I. batatas*. The figure shows coalescent times between individual chloroplast samples, with blue error bars representing the 95% highest posterior density. Green, Banks and Solander specimen; blue, sweet potato chloroplast lineage 1; red, sweet potato chloroplast lineage 2; yellow, *I. trifida*. Coalescent times between some *I. batatas* samples, including Banks and Solander's specimen, predate humans, whilst for other samples overlaps with the human era.

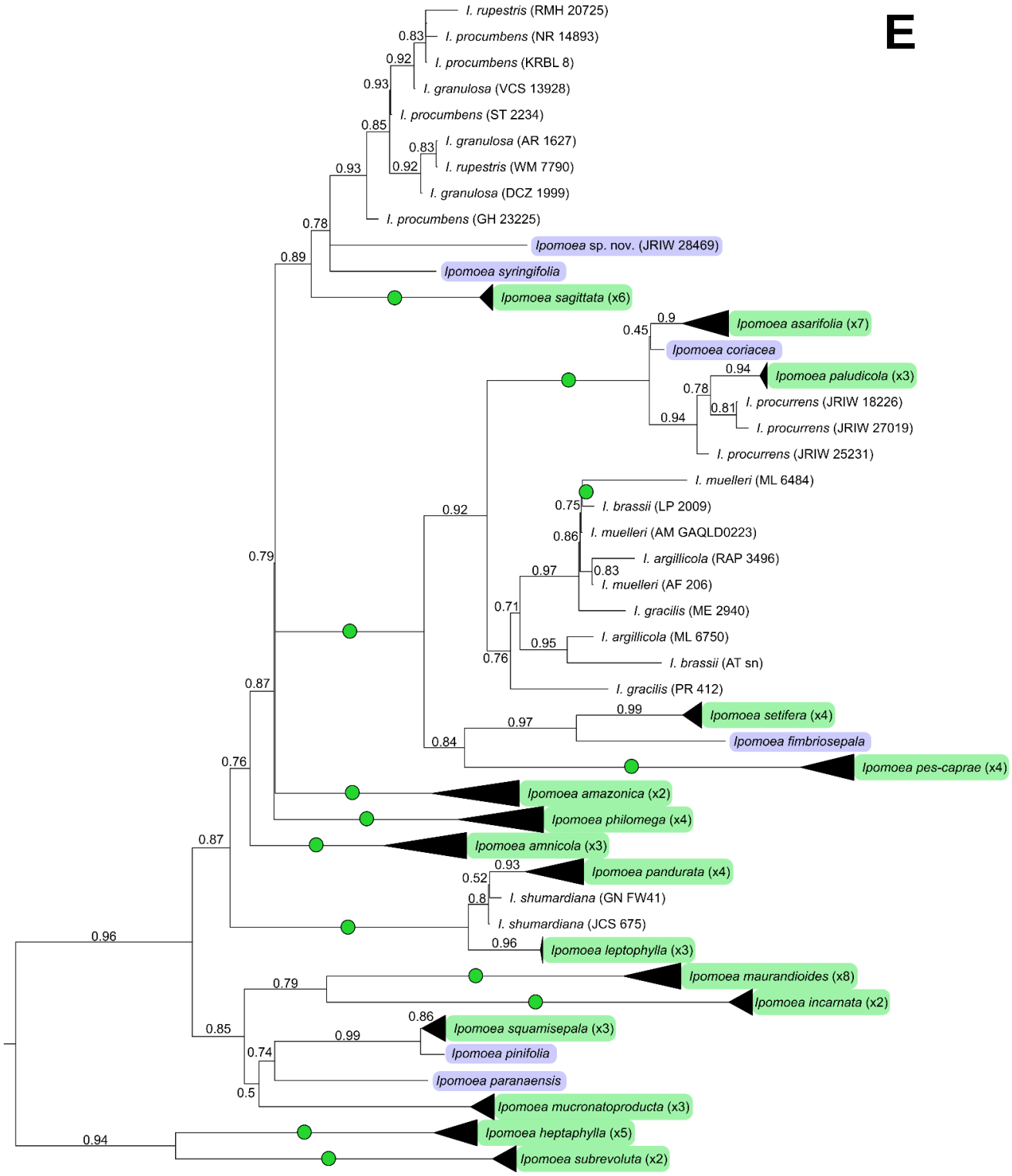

A


B
Ipomoea virgata (x2)
Ipomoea prolifera
Ipomoea aprica
I. delphinioides (AR 1718)
I. jalapa (ELE 7093)
Ipomoea attenuata
I. aurifolia (ACS 2)
I. aurifolia (EPH 5912)
I. aurifolia (FRC 3)
I. aurifolia (MLF 61)
Ipomoea uninervis (x2)
I. cuneifolia (HIS 32032)
I. hirsutissima (MLF 1234)
Ipomoea haenkeana (x2)
Ipomoea appendiculata
Ipomoea vivianae
I. megapotamica (JRIW 22862)
I. megapotamica (JRIW 27269)
I. megapotamica (JRIW 25516)
I. megapotamica (TC 6)
I. malvaeoides (TC 16)
Ipomoea longibarbis
I. psammophila (TC 5)
I. hieronymi (MM 2859)
I. megapotamica (MM 2889)
I. psammophila (JRIW 20171)
I. hieronymi (RRB 206)
Ipomoea mucronifolia
I. malvaeoides (TC 17)
I. morongii (TC 12)
Ipomoea gypsophila (x2)
I. psammophila (JRIW 25122)
Ipomoea aff. spinulifera (JRIW 28306)
I. malvaeoides (MPC 1789)
Ipomoea cardenasiana
Ipomoea sp. nov. (EH 6760)
Ipomoea sp. nov. (JRIW 23444)
I. nitida subsp. krapovickasii (TMP 5359)
I. opulifolia (JRIW 16278)
I. opulifolia (MM 3150)
I. opulifolia (ER 357)
I. hirsutissima (MM 4365)
I. morongii (TC 13)
Ipomoea bombycina
I. argentinica (JRIW 27791)
I. cuneifolia (DS 1907)
I. cuneifolia (JHP 56)
I. cuneifolia (MN 41113)
I. cuneifolia (ENG 1144)
I. cuneifolia (JAR 4129)
Ipomoea dasysperma
I. malvaeoides (AK 41206)
I. hirsutissima (JRIW 18515)
I. paludosa (CH 9391)
I. paludosa (AK 44864)
Ipomoea malpighiifolia
Ipomoea aff. malvaeoides (RV 1297)
Ipomoea cerradoensis
I. argentinica (JRIW 24101)
I. malvaeoides (EH 6116)
Ipomoea mendozae
Ipomoea subalata (x6)
I. acutisepala (AK 45769)
Ipomoea guaranitica
I. aff theodori (AS 30164)
I. argentinica (JRIW 24917)
I. argentinica (JRIW 27502)
I. argentinica (TC 7)
I. padillae (AK 15023)
I. nitida subsp. krapovickasii (AS 19971)
I. padillae (MER 1185)
I. acutisepala (PD 7385)
I. jalapa (BH 19972)
I. jalapa (E sn)
I. jalapa (E 1947)
I. delphinioides (GH3674)
Ipomoea lilloana (x3)
Ipomoea angustissima
Ipomoea valenzuelensis
Ipomoea geophyllifolia
I. jalapa (DFA 7698)
I. macrorhiza (JC 5609)
I. macrorhiza (PG 351)
I. jalapa (LLC 734)
I. macrorhiza (CRB 18568)
I. macrorhiza (LCA 10845)
Ipomoea lozanii
Ipomoea rupicola
Ipomoea gesnerioides (x2)
I. praecana (VM 112)
I. praecana (JTA 2595)
I. praecana (JVAD 3176)
Ipomoea burchelli
Ipomoea marcellia (x3)
Ipomoea descolei (x2)
Ipomoea calyptrata
Ipomoea carnea (x9)
Ipomoea macrosepala
I. prismatosyphon (RKB 13687)
I. prismatosyphon (HHS 1174)
I. prismatosyphon (JL 1963)
Ipomoea macrosiphon
Ipomoea marmorata (x3)
Ipomoea paolii
I. lapidosa (EO 15372)
I. lapidosa (MGG 5021)
Ipomoea alternifiora
I. longituba (PJG 10931)
I. longituba (JW 210)
I. grantii (BDB 3326)
I. grantii (MR 19869)
Ipomoea bakeri
I. verbascoidea (BDW 1096)
I. verbascoidea (MGB 10293)
I. albivenia (KB 9237)
I. albivenia (KB 5320)
I. verbascoidea (JMF 5131)
I. albivenia (SJM 173)
Ipomoea bullata
Ipomoea queirozii
I. pohlii (EM 8179)
I. pohlii (RMH 21902)
Ipomoea longibractoolata
Ipomoea gigantea
Ipomoea brasiliana (x10)
Ipomoea graniticola (x2)
I. rosea (BS H51872)
I. rosea (WMA 149)
Ipomoea connata
Ipomoea pterocaulis
Ipomoea abrupta (x2)
Ipomoea costata
Ipomoea polpha
Ipomoea calobra
Ipomoea juliagutierrezae (x2)
Ipomoea scopulorum (x5)
I. lottiae (EJL 3787)
I. lottiae (MH 91)
Ipomoea cuprinacoma
I. hartwegii (JVAD 3063)
I. hartwegii (DQ355313)*
Ipomoea ciervensis
Ipomoea aff. simulans (RMV 16194)
Ipomoea lonis
Ipomoea durangensis
Ipomoea sp. (Rivera 1741)
Ipomoea longifolia (x4)
Ipomoea steerei (x2)
Ipomoea reticulata (x3)
Ipomoea seopaulista
Ipomoea tanjensis
Ipomoea inaccessa (x3)
I. wolcottiana (CAP 13217)
I. wolcottiana (CAP 16643)
I. arborescens (GBH 447)
I. arborescens (JA 0248)
I. arborescens (JS 626)
Ipomoea chilopsidis
Ipomoea intrapilosa
I. wolcottiana (TDP 9602)
I. arborescens (RKB 20088)
I. arborescens (ALRG 2000-889)
I. arborescens (KFP 8222)
I. seaania (SC SC898)
I. seaania (RF 80-36)
Ipomoea populina
I. pauciflora (KGD 6495)
I. pauciflora (WG 5723)
I. pauciflora (PPM 25330)
I. murucoides (JCSN 7232)
I. murucoides (GBH 2786)
I. murucoides (MFG 5335)
Ipomoea rzedowskii
I. murucoides (EA 1933)
Ipomoea leolillanica
Ipomoea campanulata (x2)
I. sumatrana (KW 1010)
I. sumatrana (GS 1487)
I. sumatrana (KL 8419)

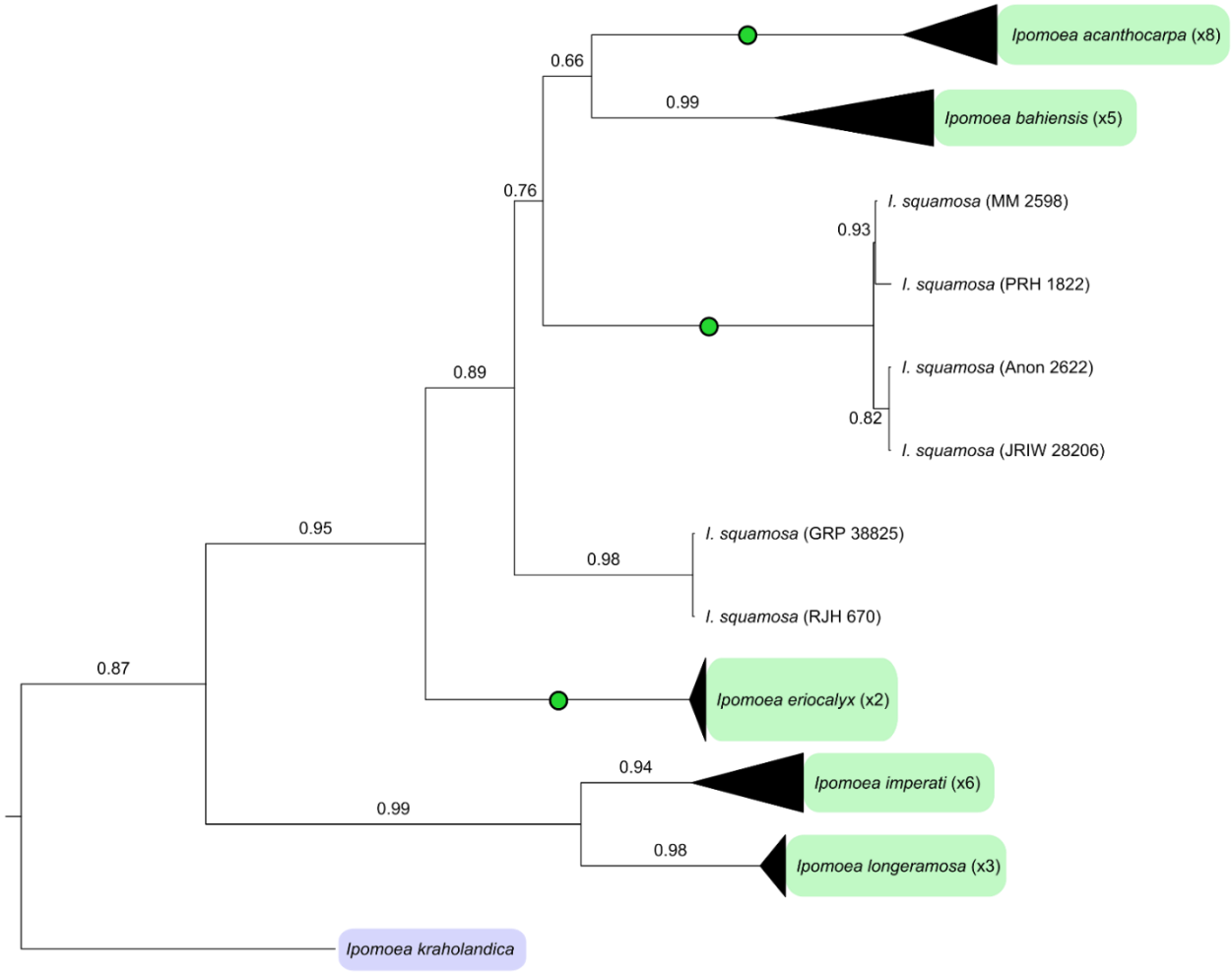
C


I. argentea (JRIW 19974)
I. argentea (JRIW 25244)
Ipomoea paulistana (x2)
I. densibracteata (JRIW 27118)
Ipomoea petrophila
I. densibracteata (JRIW 15735)
I. densibracteata (JRIW 13134)
I. batatoides (C 196)
Ipomoea platensis
I. homotrichoidea (AK 45909)
I. homotrichoidea (AK 45822)
I. schomburgkii (MJJJ 4608)
I. oranensis (MM 2427)
Ipomoea bonariensis (x7)
Ipomoea repanda (x2)
I. oranensis (JRIW 27653)
I. oranensis (JRIW 27661)
I. oranensis (JRIW 20458)
I. oranensis (MM 987)
I. cheirophylla (JRIW 22792)
I. oranensis (JRIW 27639)
I. exserta (JRIW 27595)
I. exserta (JRIW 27643)
I. batatoides (PW 471)
I. batatoides (PW 500)
Ipomoea luteoviridis
Ipomoea argentifolia
I. mauritiana (TPH 3478)
Ipomoea robinsonii (x2)
Ipomoea clausa (x3)
Ipomoea furcyensis (x2)
Ipomoea nematoloba
I. schulziana (JRIW 20232)
I. schulziana (JRIW 27731)
I. schulziana (JRIW 25053)
I. oranensis (JRIW 27682)
Ipomoea asterophora
I. schomburgkiana (JRIW 27914)
I. mauritiana (WRQL 5596)
I. cheirophylla (JRIW 20753)
I. mauritiana (FRI 74399)
I. mauritiana (GHI 440)
I. lineolata (BDM 916)
Ipomoea rubella
Ipomoea horsfalliae
Ipomoea ternata (x2)
I. lineolata (GRP 18332)
Ipomoea microdactyla (x3)
Ipomoea carolina
I. argentea (JRIW 25639)
Ipomoea sp. (MM 4705)
I. batatoides (JRIW 26095)
I. argentea (NFT 1449)
Ipomoea sp. (MM 4901)
Ipomoea eggersiana (x2)
I. argentea (JAR 2830)
Ipomoea pintoi (x4)
I. batatoides (JRIW 18399)
Ipomoea franciscana
Ipomoea asplundii (x2)
Ipomoea goyazensis
Ipomoea caloneura (x2)
I. lineolata (PC 151)
Ipomoea discolor
Ipomoea blanchetii (x2)
Ipomoea longistaminea
Ipomoea aurantiaca (x2)
Ipomoea heterodoxa (x2)
Ipomoea eremnobrocha
I. peteri (EM 23887)
I. peteri (EM 14747)
I. peteri (DEB 39698)
Ipomoea proxima

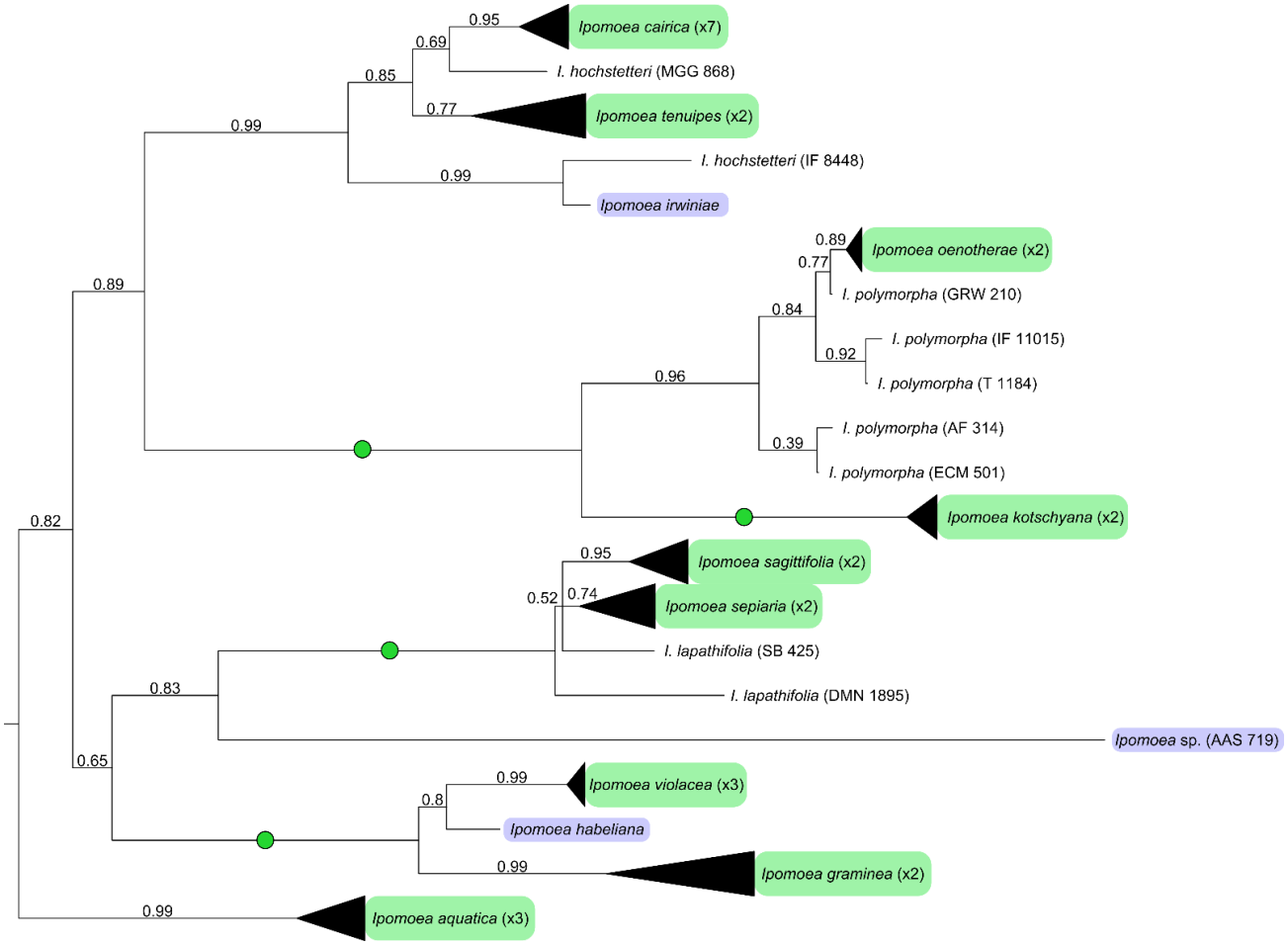

D
I. indica (HNM 534)
I. indica (IINM 772)
I. indica (GS 1568)
I. indica (RN 3271)
I. indica (GS 1397)
I. nil (IMML 934)
I. nil (JRC 8427A)
I. hederacea (RVD 87/329)
I. hederacea (AC 4985)
I. hederacea (RF 2-318)
I. hederacea (JT 6-253)
I. indica (BCC 259)
I. hederacea (CB 20424)
I. hederacea (RCB 245)
I. hederacea (CTB 21245)
I. hederacea (CB 20329)
I. nil (JRIW 11012)
I. nil (JRIW 19127)
I. indica (WS dubious)
Ipomoea lindheimeri (x2)
I. indica (BH 19551)
I. indica (V 75)
I. indica (DES 13088)
I. indica (GS 1572)
I. indica (GS 2584)
Ipomoea jamaicensis
I. indica (AEL 193)
Ipomoea pinnatisecta (x3)
Ipomoea purpurea (x6)
I. neurocephala (JD 518)
Ipomoea thurberi
I. indica (YPC 1961)
I. neurocephala (I 9562)
Ipomoea hartwegii
I. neurocephala (GPS 6303)
I. neurocephala (AF 7966)
Ipomoea magnifolia (x3)
Ipomoea villicaia
Ipomoea ampullacea (x2)
Ipomoea temascalepecensis (x2)
Ipomoea mirandina
I. parasitica (K 257)
I. parasitica (LOW 42781)
I. parasitica (JRIW 27944)
I. decasperma (GCT 2247)
I. decasperma (RG 2298)
I. decasperma (JR 34330)
I. decasperma (A 2613)
I. decasperma (CGP 10034)
Ipomoea cliniquensis
I. decasperma (GBH 6525)
I. decasperma (GBH 8442)
Ipomoea chenopodiifolia (x2)
I. orizabensis (DEB 22832)
I. orizabensis (PJS 249)
I. orizabensis (EL 474)
I. orizabensis (GBH 11922)
I. orizabensis var. collina (EC 4527)
I. orizabensis var. collina (EC 6093)
Ipomoea glissa
Ipomoea tuboides (x3)
Ipomoea rubropilosa
Ipomoea leucotricha (x2)
Ipomoea marginisepala (x3)
Ipomoea tricolor (x2)
Ipomoea cardiophylla
I. parasitica (RVD 2004-117)
Ipomoea barbatisepala (x2)
Ipomoea acutiloba (x2)
Ipomoea stuma (x2)
I. ancisa (JS 944)
I. seaconsiana (RS 13842)
I. ancisa (PSM sn)
I. seaconsiana (WLW 4212)
I. seaconsiana (CGP 2751)
I. ancisa (RS 13835)
I. seaconsiana (RW 7284)
I. muricata (AGM 7501)
I. muricata (IMML 388)
I. muricata (GBH 6932)
I. muricata (CTB 21209)
I. muricata (PSB 235)
I. muricata (JRIW 27799)
Ipomoea alba (x3)
Ipomoea santillanii (x3)
I. muricata (JRIW 27934)
I. muricata (JRIW 27838)
Ipomoea dumetorum (x4)
Ipomoea jalapensis
Ipomoea madrensis (x2)
I. plummerae (JRIW 22625)
I. capillacea (AV 4252)
I. capillacea (ALRG 2009-1334)
I. plummerae (JRIW 11994)
I. plummerae (JRIW 19564)
I. plummerae var. cuneifolia (JRIW 19454)
I. tenuiloba var. tommonii (JT 06-218)
I. capillacea (GBH 6523)
I. tenuiloba (CHTT 271)
I. tenuiloba (FR 811)
I. plummerae (JME 1132)
Ipomoea hederifolia (x5)
I. funis (YRA 1222)
I. funis (YRA 2039)
Ipomoea hastigera
Ipomoea latica
Ipomoea meyeri (x2)
Ipomoea lobata (x2)
Ipomoea cholulensis (x3)
Ipomoea aff. indivisa (IMP 8205)
Ipomoea dubia (x2)
I. cristulata (AC 4986)
I. cristulata (JLR 3626)
I. rubriflora (AK 42701)
I. cristulata (FA 151)
I. coccinea (CTB 21258)
I. coccinea (GWR 493)
I. coccinea (HEA 20484)
I. cristulata (MAB 16352)
I. rubriflora (JLR WB3280)
I. rubriflora (MM530)
I. coccinea (Gsn)
I. cristulata (GMF 3397)
I. cristulata (JT 06-254)
Ipomoea guamuchil (x2)
Ipomoea costellata (x7)
I. tenuiloba var. leptoloma (R 1020)
I. tenuiloba (AF110942)*
I. tenuiloba var. leptoloma (BW 10/506)
I. tenuiloba var. leptoloma (DFA 7595)
I. tenuiloba (MG910327)*
I. tenuiloba (MAMR 685)
Ipomoea lasiensis
Ipomoea chamelana (x2)
Ipomoea minutiflora (x4)
Ipomoea microsepala (x2)
Ipomoea bracteata (x2)
I. suffulta (VWS 5453)
I. suffulta (ARG 825)
I. suffulta (CGP 4755)
I. seducta (EC 6926)
I. seducta (VWS 4945)
I. seducta (MV 26396)
I. damosa (K 1353)
I. damosa (WDS 29321)
I. seducta (ACS 10647)
I. damosa (CJV 572)
I. purga (YM 1843)
I. damosa (YRA 490)
I. damosa (EC 6780)
I. damosa (GR 746)
I. purga (GBH 2220)
I. tubulata (JCSN 10987)
I. tubulata (VWS 5603)
Ipomoea aristolochiifolia (x7)
Ipomoea ignava (x2)
I. elongata (EM 37250)
I. elongata (AGM 1060)
I. elongata (CAP 3904)
Ipomoea meyeri (x4)
Ipomoea mitchellae
Ipomoea puncticulata (x4)
Ipomoea velardei
Ipomoea odontophylla (x2)
Ipomoea huaylae
Ipomoea bonvouloirana (x2)
Ipomoea jicama
Ipomoea pittieri (x2)
Ipomoea simulans

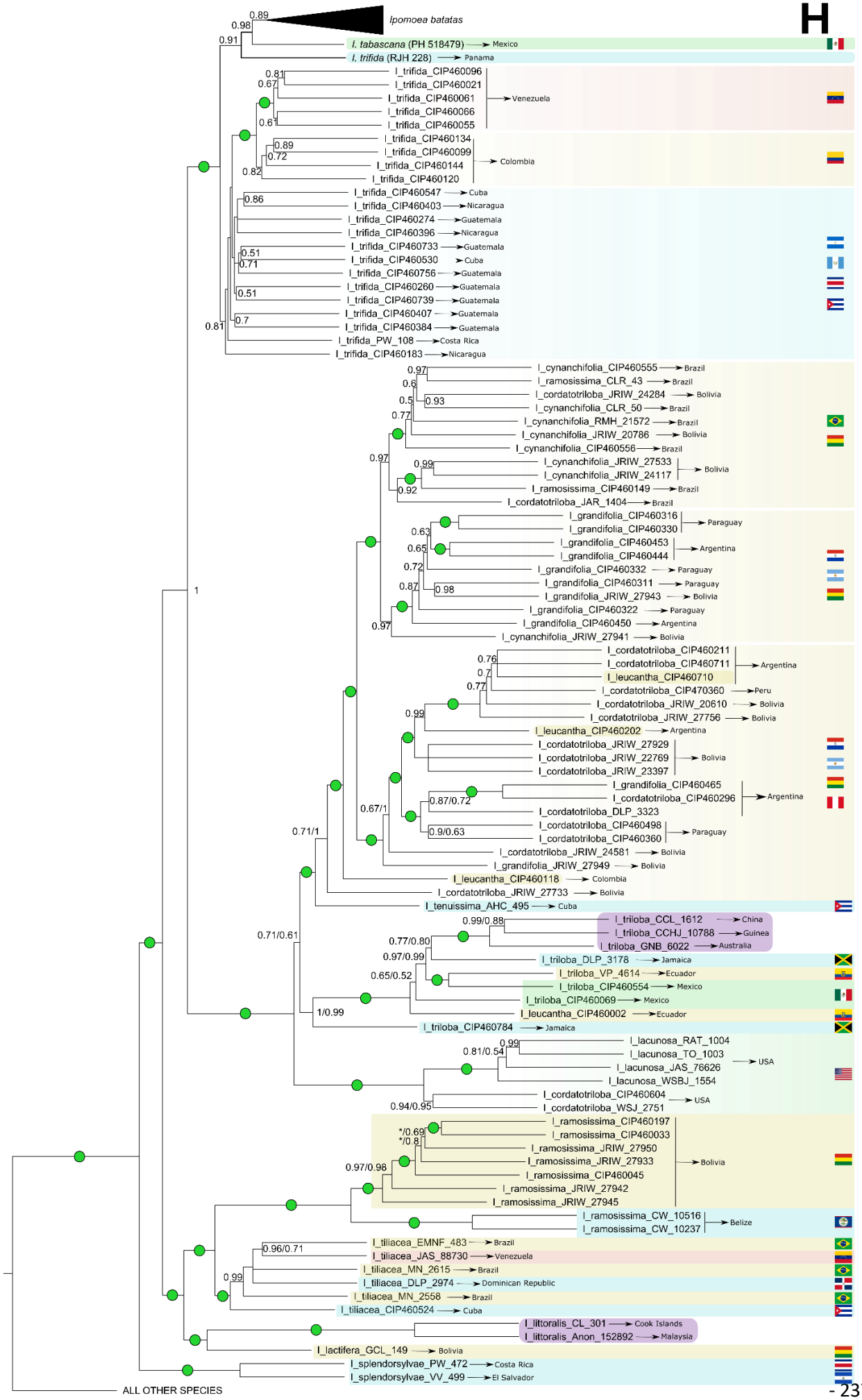

E
I. rupestris (RMH 20725)
I. procumbens (NR 14893)
I. procumbens (KRBL 8)
I. granulosa (VCS 13928)
I. procumbens (ST 2234)
I. granulosa (AR 1627)
I. rupestris (WM 7790)
I. granulosa (DCZ 1999)
I. procumbens (GH 23225)
Ipomoea sp. nov. (JRIW 28469)
Ipomoea syringifolia
Ipomoea sagittata (x6)
Ipomoea asarifolia (x7)
Ipomoea coriacea
Ipomoea paludicola (x3)
I. procurrens (JRIW 18226)
I. procurrens (JRIW 27019)
I. procurrens (JRIW 25231)
I. muelleri (ML 6484)
I. brassii (LP 2009)
I. muelleri (AM GAQLD0223)
I. argillicola (RAP 3496)
I. muelleri (AF 206)
I. gracilis (ME 2940)
I. argillicola (ML 6750)
I. brassii (AT sn)
I. gracilis (PR 412)
Ipomoea setifera (x4)
Ipomoea fimbriosepala
Ipomoea pes-caprae (x4)
Ipomoea amazonica (x2)
Ipomoea philomega (x4)
Ipomoea amnicola (x3)
Ipomoea pandurata (x4)
I. shumardiana (GN FW41)
I. shumardiana (JCS 675)
Ipomoea leptophylla (x3)
Ipomoea maurandioides (x8)
Ipomoea incarnata (x2)
Ipomoea squamisepala (x3)
Ipomoea pinifolia
Ipomoea paranaensis
Ipomoea mucronatoproducta (x3)
Ipomoea heptaphylla (x5)
Ipomoea subrevoluta (x2)

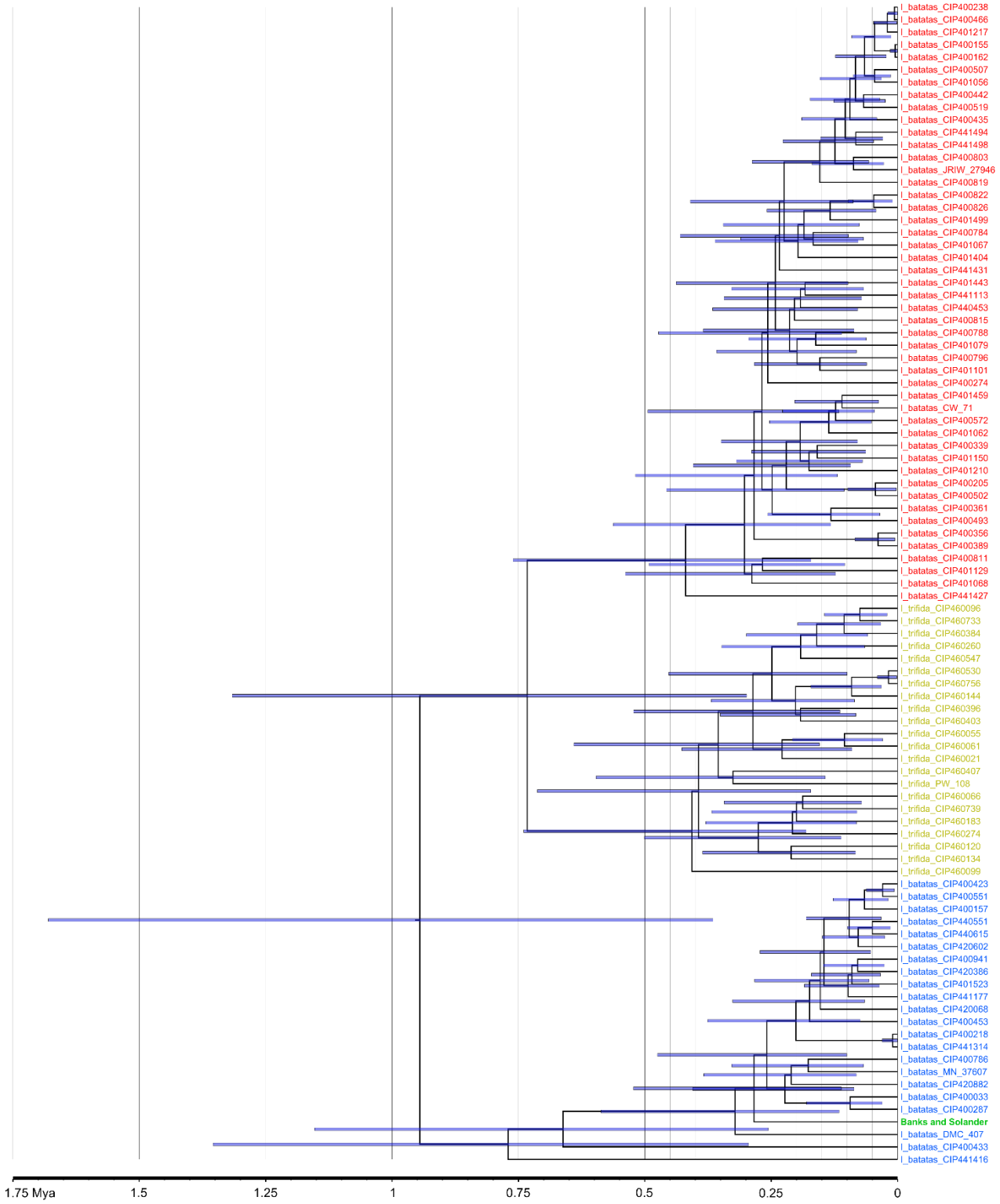
F


Ipomoea acanthocarpa (x8)
0.66
0.99
Ipomoea bahiensis (x5)
0.76
I. squamosa (MM 2598)
0.93
I. squamosa (PRH 1822)
0.89
I. squamosa (Anon 2622)
0.82
I. squamosa (JRIW 28206)
0.95
I. squamosa (GRP 38825)
0.98
I. squamosa (RJH 670)
0.87
Ipomoea eriocalyx (x2)
0.94
Ipomoea imperati (x6)
0.99
0.98
Ipomoea longeramosa (x3)
Ipomoea kraholandica

G


0.95
Ipomoea cairica (x7)
0.69
I. hochstetteri (MGG 868)
0.85
0.77
Ipomoea tenuipes (x2)
0.99
I. hochstetteri (IF 8448)
0.99
Ipomoea irwiniae
0.89
Ipomoea oenotherae (x2)
0.77
I. polymorpha (GRW 210)
0.84
I. polymorpha (IF 11015)
0.92
I. polymorpha (T 1184)
0.96
I. polymorpha (AF 314)
0.39
I. polymorpha (ECM 501)
0.89
Ipomoea kotschyana (x2)
0.82
0.95
Ipomoea sagittifolia (x2)
0.52
0.74
Ipomoea sepiaria (x2)
I. lapathifolia (SB 425)
I. lapathifolia (DMN 1895)
0.83
Ipomoea sp. (AAS 719)
0.65
0.99
Ipomoea violacea (x3)
0.8
Ipomoea habeliana
0.99
Ipomoea graminea (x2)
0.99
Ipomoea aquatica (x3)


H
Ipomoea batatas
I. tabascana (PH 518479) → Mexico
I. trifida (RJH 228) → Panama
I_trifida_CIP460096
I_trifida_CIP460021
I_trifida_CIP460061
I_trifida_CIP460066
I_trifida_CIP460055
→ Venezuela
I_trifida_CIP460134
I_trifida_CIP460099
I_trifida_CIP460144
I_trifida_CIP460120
→ Colombia
I_trifida_CIP460547 → Cuba
I_trifida_CIP460403 → Nicaragua
I_trifida_CIP460274 → Guatemala
I_trifida_CIP460396 → Nicaragua
I_trifida_CIP460733 → Guatemala
I_trifida_CIP460530 → Cuba
I_trifida_CIP460756 → Guatemala
I_trifida_CIP460260 → Guatemala
I_trifida_CIP460739 → Guatemala
I_trifida_CIP460407 → Guatemala
I_trifida_CIP460384 → Guatemala
I_trifida_PW_108 → Costa Rica
I_trifida_CIP460183 → Nicaragua
I_cynanchifolia_CIP460555 → Brazil
I_ramosissima_CLR_43 → Brazil
I_cordatotriloba_JRIW_24284 → Bolivia
I_cynanchifolia_CLR_50 → Brazil
I_cynanchifolia_RMH_21572 → Brazil
I_cynanchifolia_JRIW_20786 → Bolivia
I_cynanchifolia_CIP460556 → Brazil
I_cynanchifolia_JRIW_27533
I_cynanchifolia_JRIW_24117
→ Bolivia
I_ramosissima_CIP460149 → Brazil
I_cordatotriloba_JAR_1404 → Brazil
I_grandifolia_CIP460316
I_grandifolia_CIP460330
→ Paraguay
I_grandifolia_CIP460453
I_grandifolia_CIP460444
→ Argentina
I_grandifolia_CIP460332 → Paraguay
I_grandifolia_CIP460311 → Paraguay
I_grandifolia_JRIW_27943 → Bolivia
I_grandifolia_CIP460322 → Paraguay
I_grandifolia_CIP460450 → Argentina
I_cynanchifolia_JRIW_27941 → Bolivia
I_cordatotriloba_CIP460211
I_cordatotriloba_CIP460711
I_leucantha_CIP460710
→ Argentina
I_cordatotriloba_CIP470360 → Peru
I_cordatotriloba_JRIW_20610 → Bolivia
I_cordatotriloba_JRIW_27756 → Bolivia
I_leucantha_CIP460202 → Argentina
I_cordatotriloba_JRIW_27929
I_cordatotriloba_JRIW_22769
I_cordatotriloba_JRIW_23397
→ Bolivia
I_grandifolia_CIP460465
I_cordatotriloba_CIP460296
→ Argentina
I_cordatotriloba_DLP_3323
I_cordatotriloba_CIP460498
I_cordatotriloba_CIP460360
→ Paraguay
I_cordatotriloba_JRIW_24581 → Bolivia
I_grandifolia_JRIW_27949 → Bolivia
I_leucantha_CIP460118 → Colombia
I_cordatotriloba_JRIW_27733 → Bolivia
I_tenuissima_AHC_495 → Cuba
I_triloba_CCL_1612 → China
I_triloba_CCHJ_10788 → Guinea
I_triloba_GNB_6022 → Australia
I_triloba_DLP_3178 → Jamaica
I_triloba_VP_4614 → Ecuador
I_triloba_CIP460554 → Mexico
I_triloba_CIP460069 → Mexico
I_leucantha_CIP460002 → Ecuador
I_triloba_CIP460784 → Jamaica
I_lacunosa_RAT_1004
I_lacunosa_TO_1003
I_lacunosa_JAS_76626
→ USA
I_lacunosa_WSBJ_1554
I_cordatotriloba_CIP460604 → USA
I_cordatotriloba_WSJ_2751
I_ramosissima_CIP460197
I_ramosissima_CIP460033
I_ramosissima_JRIW_27950
I_ramosissima_JRIW_27933
I_ramosissima_CIP460045
I_ramosissima_JRIW_27942
I_ramosissima_JRIW_27945
→ Bolivia
I_ramosissima_CW_10516
I_ramosissima_CW_10237
→ Belize
I_tiliacea_EMNF_483 → Brazil
I_tiliacea_JAS_88730 → Venezuela
I_tiliacea_MN_2615 → Brazil
I_tiliacea_DLP_2974 → Dominican Republic
I_tiliacea_MN_2558 → Brazil
I_tiliacea_CIP460524 → Cuba
I_littoralis_CL_301 → Cook Islands
I_littoralis_Anon_152892 → Malaysia
I_lactifera_GCL_149 → Bolivia
I_splendorsylvae_PW_472 → Costa Rica
I_splendorsylvae_VV_499 → El Salvador
ALL OTHER SPECIES


I_batatas_CIP400238
I_batatas_CIP400466
I_batatas_CIP401217
I_batatas_CIP400155
I_batatas_CIP400162
I_batatas_CIP400507
I_batatas_CIP401056
I_batatas_CIP400442
I_batatas_CIP400519
I_batatas_CIP400435
I_batatas_CIP441494
I_batatas_CIP441498
I_batatas_CIP400803
I_batatas_JRIW_27946
I_batatas_CIP400819
I_batatas_CIP400822
I_batatas_CIP400826
I_batatas_CIP401499
I_batatas_CIP400784
I_batatas_CIP401067
I_batatas_CIP401404
I_batatas_CIP441431
I_batatas_CIP401443
I_batatas_CIP441113
I_batatas_CIP440453
I_batatas_CIP400815
I_batatas_CIP400788
I_batatas_CIP401079
I_batatas_CIP400796
I_batatas_CIP401101
I_batatas_CIP400274
I_batatas_CIP401459
I_batatas_CW_71
I_batatas_CIP400572
I_batatas_CIP401062
I_batatas_CIP400339
I_batatas_CIP401150
I_batatas_CIP401210
I_batatas_CIP400205
I_batatas_CIP400502
I_batatas_CIP400361
I_batatas_CIP400493
I_batatas_CIP400356
I_batatas_CIP400389
I_batatas_CIP400811
I_batatas_CIP401129
I_batatas_CIP401068
I_batatas_CIP441427
I_trifida_CIP460096
I_trifida_CIP460733
I_trifida_CIP460384
I_trifida_CIP460260
I_trifida_CIP460547
I_trifida_CIP460530
I_trifida_CIP460756
I_trifida_CIP460144
I_trifida_CIP460396
I_trifida_CIP460403
I_trifida_CIP460055
I_trifida_CIP460061
I_trifida_CIP460021
I_trifida_CIP460407
I_trifida_PW_108
I_trifida_CIP460066
I_trifida_CIP460739
I_trifida_CIP460183
I_trifida_CIP460274
I_trifida_CIP460120
I_trifida_CIP460134
I_trifida_CIP460099
I_batatas_CIP400423
I_batatas_CIP400551
I_batatas_CIP400157
I_batatas_CIP440551
I_batatas_CIP440615
I_batatas_CIP420602
I_batatas_CIP400941
I_batatas_CIP420386
I_batatas_CIP401523
I_batatas_CIP441177
I_batatas_CIP420068
I_batatas_CIP400453
I_batatas_CIP400218
I_batatas_CIP441314
I_batatas_CIP400786
I_batatas_MN_37607
I_batatas_CIP420882
I_batatas_CIP400033
I_batatas_CIP400287
Banks and Solander
I_batatas_DMC_407
I_batatas_CIP400433
I_batatas_CIP441416
1.75 Mya
1.5
1.25
1
0.75
0.5
0.25
0