Pablo Muñoz-Rodríguez, Tom Carruthers, John R.I. Wood, Bethany R.M. Williams, Kevin Weitemier, Brent Kronmiller, David Ellis, Noelle L. Anglin, Lucas Longway, Stephen A. Harris, Mark D. Rausher, Steven Kelly, Aaron Liston, Robert W. Scotland

Reconciling Conflicting Phylogenies in the Origin of Sweet Potato and Dispersal to Polynesia

Current Biology
2018
Volume 28
Issue 8
Pages 1246 – 1256.e12
https://doi.org/10.1016/j.cub.2018.03.020

Reconciling Conflicting Phylogenies in the Origin of Sweet Potato and Dispersal to Polynesia

Authors: Pablo Muñoz-Rodríguez[1], Tom Carruthers[1], John R.I. Wood[1], Bethany R.M. Williams[1], Kevin Weitemier[2], Brent Kronmiller[3], David Ellis[4], Noelle L. Anglin[4], Lucas Longway[2], Stephen A. Harris[1], Mark D. Rausher[5], Steve Kelly[1], Aaron Liston[2], Robert W. Scotland[1,6,*.]

Affiliations:

[1] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, United Kingdom.

[2] Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, United States of America.

[3] Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331, United States of America.

[4] International Potato Center, Av. La Molina 1895, La Molina, Lima, Peru.

[5] 3332 French Family Science Center. 124 Science Drive, Duke University, Durham, NC 27708, United States of America.

[6] Lead Contact

***Correspondance:** robert.scotland@plants.ox.ac.uk.

SUMMARY

Sweet potato is one of the world's most widely consumed crops, yet its evolutionary history is poorly understood. In this paper we present a comprehensive phylogenetic study of all species closely related to sweet potato and address several questions pertaining to sweet potato that remained unanswered. Our research combined genome skimming and target DNA capture to sequence the whole chloroplasts and 605 single copy nuclear regions from 199 specimens representing sweet potato and all its crop wild relatives (CWRs). We present

strongly supported nuclear and chloroplast phylogenies which demonstrate that sweet potato had an autopolyploid origin and *Ipomoea trifida* is its closest relative, confirming that no other extant species were involved in its origin. Phylogenetic analysis of nuclear and chloroplast genomes show conflicting topologies regarding the monophyly of sweet potato. The process of chloroplast capture explains these conflicting patterns, showing that *Ipomoea trifida* had a dual role in the origin of sweet potato, first as its progenitor and second as the species with which sweet potato introgressed so one of its lineages could capture an *I. trifida* chloroplast. In addition, we provide evidence that sweet potato was present in Polynesia in pre-human times. This, together with several other examples of long-distance dispersal in *Ipomoea*, negates the need to invoke ancient human-mediated transport as an explanation for its presence in Polynesia. These results have important implications for understanding the origin and evolution of a major global food crop and question the existence of pre-Columbian contacts between Polynesia and the American continent.

Keywords: chloroplast capture, crop wild relatives, target enrichment, *Ipomoea*, long-distance dispersal, phylogenomics, Polynesia, sweet potato.

INTRODUCTION

Two fundamental questions related to the origin and dispersal of sweet potato (*Ipomoea batatas* (L.) Lam.) remain unanswered. First, did the sweet potato evolve once or multiple times and what species were involved in its origin? Second, how did the sweet potato, a crop of American origin, come to be widespread in Polynesia before the arrival of the Europeans? Answering the first question requires knowledge of evolutionary relationships between the sweet potato and the species that are most closely related to it, often termed Crop Wild Relatives (CWRs, Figure 1). Understanding this relationship is the key to unraveling the origin of this crop and has implications for food security because these CWRs constitute potential sources of genetic variation for future crop improvement. In the case of sweet

potato, knowledge of these relationships is especially poor, even though it is a widely consumed crop [1,2] and an important resource for combating vitamin A deficiencies, estimated to affect over 190 million children worldwide [3]. Answering the second question (how did the sweet potato come to be present in Polynesia before the Europeans first arrived?) requires the consideration of two additional questions. First, what is the possibility of sweet potato dispersing from its native range in America to Polynesia by natural means (i.e. wind, water, birds)? Secondly, when did sweet potato colonize Polynesia?

In answer to the first question (did sweet potato evolve once or multiple times?), recent evidence provides support for both hypotheses. The analysis of AFLP fragments [4] and the identification of an *Agrobacterium*-mediated transposon in the nuclear genome of sweet potato cultivars —but absent in the sampled wild relatives [5], can be interpreted as evidence for a single origin. In contrast, the identification of two sweet potato gene pools from the analysis of chloroplast markers has been interpreted as evidence for multiple origins [6,7]. Distinguishing between these contrasting hypotheses is a prerequisite for answering other basic questions related to sweet potato such as the identity of its progenitor. Almost all wild relatives have been proposed as the progenitor of sweet potato, especially *Ipomoea trifida* Kunth and *I. triloba* L. [7–10], but conclusive evidence for any one species has hitherto been lacking. In addition, because it is unknown whether sweet potato evolved once or multiple times, it is impossible to be certain whether hexaploid sweet potato evolved by autopolyploidy from a single ancestor or by hybridization (allopolyploidy) between different species. Distinguishing between these very different routes to polyploidy is crucial for the correct interpretation and understanding of the sweet potato genome [11]. Lastly, the date of divergence between sweet potato and its wild relatives has never been comprehensively explored and, consequently, the chronology of sweet potato evolution relative to human and pre-human history is essentially unknown.

The question of how sweet potato, a crop of American origin, came to be widespread in Polynesia by the time Europeans first arrived, has been a source of controversy since the 19th century [12,13]. Several previous studies have sought to explain its presence in Polynesia as the result of three main transoceanic introductions. This so-called "tripartite hypothesis" [14–17] explains its presence in terms of two relatively well-documented introductions by Spanish and Portuguese travelers [17], and a presumed third earlier introduction in pre-Columbian times [14–16,18,19]. However, whether this earlier introduction existed and when it occurred remains highly speculative and controversial [17,19]. Most authors have explained the earlier introduction of sweet potato to Polynesia by postulating pre-Columbian human contacts between the two regions [16,18], rather than by considering transport by natural means such as wind, water or birds [20–22]. Additional evidence for the human transportation hypothesis was seen in the somewhat similar linguistic terms used for the crop in the two regions [15,23]. The possible human transportation of sweet potato to Polynesia has attracted broad attention in recent times —especially now that sweet potato appears to be the only remaining biological evidence for these alleged Pre-Columbian contacts; other supporting evidence from chicken and human DNA is now considered questionable [24,25].

In this paper, we present a comprehensive phylogenetic study of sweet potato and all its CWRs based on the whole chloroplast genome and 605 single-copy nuclear DNA regions. We provide answers to the questions about the origin of sweet potato outlined above and re-examine its arrival in Polynesia.

RESULTS AND DISCUSSION

Sweet potato's closest relative and autopolyploid origin

We produced separate nuclear and chloroplast phylogenies that strongly support the monophyly of series *Batatas*, irrespective of the type of analysis (see STAR Methods)

(Figures 2A, S1 and S2). These phylogenies also resolve well-supported relationships between sweet potato and all closely related species, providing the phylogenetic framework necessary for investigating the origin of the crop. The nuclear data shows *Ipomoea splendor-sylvae* House as sister to the rest of the *Batatas* group, whilst the chloroplast phylogeny has *I. splendor-sylvae* and *I. ramosissima* (Poir.) Choisy as sister taxa and together as sister to the other species in the *Batas* group. The section is then divided into a group of perennial species (*I. tiliacea* (Willd.) Choisy, *I. littoralis* Blume and *I. lactifera* J.R.I.Wood & Scotland, and also *I. ramosissima* in the nuclear phylogeny) and a second group containing two clades: one formed by six putative annual species (*I. triloba* L., *I. cordatotriloba* Dennst., *I. lacunosa* L., *I. grandifolia* (Dammer) O´Donell, *I. cynanchifolia* Meisn. and *I. tenuissima* Choisy), three of which are not monophyletic; and another formed by *I. batatas* and *I. trifida*. In addition, our results show that *I. leucantha* Jacq., previously identified as a hybrid [26], is polyphyletic (Figure S3), and confirm that *I. tabascana* J.A.McDonald & D.F.Austin is most likely a recent hybrid between *I. batatas* and *I. trifida* [10,27] (Figures S3 and S4C).

According to our analysis of nuclear data, sweet potato is monophyletic and *Ipomoea trifida* is its closest relative (Figure 2A). This result corroborates two previous studies that imply a single origin for the crop [4,5] and reject recent claims that advocate multiple origins based on the discovery of two sweet potato gene pools [6,7]. It is, therefore, reasonable to assume that sweet potato had a single origin and most probably evolved from *Ipomoea trifida*, a circum-Caribbean species.

Sweet potato is the only species in the *Batatas* group that is hexaploid (2n = 6X = 90), all other species being either diploid or tetraploid [28–30]. As a hexaploid entity, we would expect sweet potato to contain six alleles at each of the genetic loci analyzed in our study. We therefore estimated allelic variation within each specimen (see Haplotype identification in STAR Methods). The analysis of these alleles shows that, for the vast majority of gene trees,

all six putative alleles of hexaploid *I. batatas* are more closely related to each other than to alleles from any other species including *I. trifida* (Figure 2B). This strongly suggests an autopolyploid origin of sweet potato and provides no support for a hybrid (allopolyploid) origin involving any other species, including *I. triloba*, which has been proposed as progenitor of the crop by several authors [26,31].

Conflicting chloroplast phylogeny and chloroplast capture

In contrast to the nuclear data, the analysis of whole chloroplast genomes revealed the existence of two distinct sweet potato gene pools (here termed chloroplast lineage 1 and 2), as had previously been inferred from limited data [7] (Figure 3A-B). Our data show that chloroplast lineage 2 (CL2) is more closely related to *Ipomoea trifida*, whereas chloroplast lineage 1 (CL1) is sister group to these two (Figure S2).

All statistical tests and additional analyses conducted on the chloroplast data to challenge this result confirm the existence of two distinct sweet potato gene pools (Figures S5A–C). In addition, we visually explored the chloroplast alignment and discovered that there are no indels shared exclusively by the two sweet potato chloroplast gene pools, but both have unique indels and both also share indels with *Ipomoea trifida*, as would be expected if sweet potato contains two chloroplast haplotypes but inherited from *I. trifida* at different times.

If sweet potato had multiple origins, as suggested by these two independent chloroplast gene pools [6,7], or if it had gradually diversified from an ancestral polymorphism in *Ipomoea 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. We subsequently explored our nuclear data and one additional non-coding region (*ribosomal DNA Internal Transcribed Spacer*), which was assembled specifically because evidence of these two gene pools in the nuclear genome was allegedly found in this region [7]; no additional phylogenetic nor population structure analyses retrieved the two gene pools from

the nuclear data (Figure S4A-C). Also, we found no evidence that either incomplete lineage sorting or recombination affected the nuclear topology [32] (see description of phylogenetic analyses in STAR Methods). In summary, the conflicting topologies obtained for nuclear and chloroplast data have strong support and are consistent for all phylogenetic inference methods.

Given these findings, the evidence strongly suggests that the two distinct *Ipomoea batatas* chloroplast gene pools are the result of chloroplast capture from *I. trifida* following species divergence of *I. batatas* and *I. trifida*. Chloroplast capture is the introgression of a chloroplast genome from one plant species into another, sometimes with no evidence of nuclear gene flow [33], and is commonly proposed to explain inconsistencies between phylogenetic trees based on nuclear and chloroplast sequences [32,33].

In the context of these results, we consider several possible mechanisms of chloroplast capture can be supported by the data (Figures 4B-C). First, the result of the hybridization between a female *I. trifida* (diploid) and a male *I. batatas* (hexaploid) would be an entity carrying a *trifida-like* chloroplast. This entity, possibly allotetraploid, would later give rise to a new hexaploid form by further hybridization with *I. trifida*, i.e. generating a triploid entity that subsequently doubled to yield a hexaploid; less likely, the new hexaploid could also arise by additional autopolyploidization from the tetraploid intermediate and subsequent genome reduction. The newly formed hexaploid, coexisting with the original hexaploid *I. batatas*, would cross repeatedly with the original hexaploid lineage, progressively losing the *trifida* component of its nuclear genome while maintaining a *trifida-like* chloroplast (Figure 4B).

Since the result of this secondary contact and hybridization is a hexaploid entity with the same nuclear signature as the original sweet potato, but a captured chloroplast from *Ipomoea trifida*, one other possibility is that the phylogenetic pattern retrieved could be the result of an asymmetrical hybridization event, for which multiple examples have been described in plants

[34]. In this situation (Figure 4C), the entire nuclear genome would have been provided by an unreduced (hexaploid) sweet potato male gamete, whereas the chloroplast would have been inherited from an *I. trifida* maternal progenitor. The nuclear genome of the newly formed hexaploid entity would then be identical to that of the original *I. batatas*, thus showing a monophyletic sweet potato in the nuclear phylogeny, whereas the chloroplast phylogeny would reflect the chloroplast capture from *I. trifida*. If this mechanism is correct, it would explain the capture of the chloroplast by sweet potato without the need of a second polyploidization event.

Regardless of the exact mechanism of chloroplast capture, our results show that *I. batatas* evolved solely from *I. trifida* by autopolyploidization, and subsequently expanded its distribution range further beyond *I. trifida*'s natural distribution. Both species became reciprocally monophyletic over time and then hybridized, presumably over the sympatric area of their distribution, resulting in these populations of sweet potato with a different chloroplast. Meanwhile, other *I. batatas* populations retained the original chloroplast. Therefore, although *I. batatas* evolved from *I. trifida* by autopolyploidy, the chloroplast capture provides evidence that there was subsequent hybridization between the two species and so sweet potato contains two elements, one that is an autopolyploid (CL1) and another that is technically auto-allopolyploid (CL2).

Finally, additional sequencing revealed that two significant sweet potato varieties used in crop breeding research, Beauregard (orange-fleshed, low dry matter) and Tanzania (white-fleshed, high dry matter) (Dorcus Gemenet, pers. comm.), belong to CL1 (Figure S5D). This result possibly reflects the fact that sweet potato CL1 contains much phenotypic and genetic diversity, which would explain their use in contemporary crop breeding.

Divergence times in origin of sweet potato

In order to infer divergence times for sweet potato, and due to a lack of previous comprehensive divergence time estimates in Convolvulaceae, we first inferred a time-calibrated phylogeny for Convolvulaceae and Solanaceae. We then used a matrix containing samples throughout *Ipomoea* based on 21 nuclear regions for which there was high coverage (99%) to infer divergence times within the genus, including that of the crown node of series *Batatas*. Based on the ages inferred for this specific node, we inferred two more time-calibrated phylogenies of *Batatas*: one using plastome data and another using a matrix of the same 21 nuclear genes used to infer divergence times throughout *Ipomoea* (100% coverage).

According to our nuclear data, the clade including sweet potato and *Ipomoea trifida* diverged from its sister clade at least 1.5 million years ago and sweet potato diverged from *I. trifida* at least 800,000 years ago (red bar in Figure 5A). The hybridization between *I. trifida* and *I. batatas* that led to chloroplast capture then occurred within 56,000 years of the two species diverging (Figure 5B).

Minor temporal differences were noted in divergence time estimates inferred from the nuclear and the chloroplast datasets within *Ipomoea* series *Batatas*. One example is found in the inferred age of the clade containing *Ipomoea trifida* and *I. batatas*. This was approximately 200,000 years older when the analysis of the nuclear dataset was compared to the chloroplast data. This is likely to reflect the different evolutionary histories of these two genomes – a fact clearly demonstrated by the different topologies obtained from phylogenetic analyses of the nuclear and chloroplast datasets. These different topologies are likely to have a significant effect on divergence time estimates. As well as resolving different topologies, independent genomes are also likely to exhibit different patterns of molecular evolutionary rate variation between lineages. Given the inherent difficulties and inadequacies of current methods for accurately inferring patterns of rate heterogeneity, it is unsurprising that different age estimates are obtained from these two genomes [35,36].

Only one other study has explicitly estimated a divergence time for the split between *Ipomoea batatas* and *I. trifida* [11]. This study utilized an average mutation rate, which in turn had been calculated for *Arabidopsis thaliana* over a period of 30 generations [11,37]. The manner by which this rate was inferred may seem of little relevance for estimating divergence times in *Ipomoea* over timescales of hundreds of thousands of years. However, it is noteworthy that the timescale of events suggested by that paper is broadly congruent with the timescale we inferred for the divergence between *Ipomoea batatas* and *I. trifida* and subsequent chloroplast capture (from 380,000 years ago to 800,000 years ago). Nonetheless, the degree to which the study in [11] is congruent with our results is difficult to determine as it only analyzed nuclear data, meaning that it presented a less complete picture of the origin of sweet potato than our analysis of both nuclear and chloroplast data.

Ancestral population sizes

The depth of taxon sampling in our study allowed us to carry out a multispecies coalescent analysis of chloroplast data for all sampled specimens of *Ipomoea batatas* and *I. trifida*. This analysis was implemented in a Bayesian framework which simultaneously estimated coalescent times between different plastome lineages as well as ancestral population sizes. The aim of this analysis was to estimate effective population sizes for species and ancestral lineages within this clade [38], and to infer whether a population bottleneck was associated with the origin of sweet potato, or within the population in which chloroplast capture occurred.

This analysis unequivocally demonstrated that a population bottleneck affecting the entire clade of *Ipomoea batatas* and *I. trifida* occurred over 640,000 years ago. Subsequently, a bottleneck also affected the clade of *I. trifida* and *I. batatas* CL2, and this bottleneck ended over 370,000 years ago (Table S1). It can thus be inferred that the origin of sweet potato, and notably the chloroplast capture event, are likely to have occurred in ancestral populations

which were significantly smaller than extant populations. The population in which chloroplast capture occurred was at least 1/5 of the size of extant populations, potentially explaining the rapid spread of the captured chloroplast throughout the population.

Sweet potato presence in Polynesia

The inferred phylogeny of sweet potato and its CWRs presented in this paper (Figures 2, S1 and S2) confirmed that all species in this clade, with one exception, are restricted to the Americas. The exception, *Ipomoea littoralis* Blume, is distributed from Polynesia to Madagascar but is absent from the American continent [39,40]. *Ipomoea littoralis* diverged from its sister species *I. lactifera* J.R.I.Wood & Scotland more than 1.1 million years ago (blue bar in Figure 5A), strongly suggesting that the distribution of *I. littoralis* is best explained by natural dispersal of an ancestor of *I. littoralis* across the Pacific, followed by subsequent evolution into a different species. *Ipomoea littoralis* seeds are morphologically very similar to sweet potato seeds (Figure S6A) and, although as far as we know their buoyancy has not been tested, it has been shown that the seeds of several other *Ipomoea* species that live in similar environments can survive after floating long distances [41,42]. It would be very difficult to explain the distribution of *I. littoralis* and other widely distributed seashore species (e.g. *Ipomoea pes-caprae* (L.) R.Br., *I. violacea* L., *I. sagittata* Poir.) except in terms of long-distance dispersal by sea currents.

One other example of a highly disjunct distribution pattern within *Ipomoea* is that of *I. tuboides* O.Deg. & Ooststr. This species is endemic to the Hawaiian Islands but belongs to a clade dominated by Mexican species (Figure 6A). The time-calibrated phylogeny of this group shows that *I. tuboides* diverged from its sister species at least 1.1 million years ago (orange bar in Figure 6B), and the most likely explanation for its presence in Hawai'i, more than 5,000 kilometers from the Mexican coast, is 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 [43] are readily dispersed over very long distances. Long-distance dispersal can thus be considered the most plausible explanation of how sweet potato came to be distributed in pre-Columbian Polynesia.

In addition to other sources of data, specimens collected in Polynesia during the first European trips to the region are of extraordinary interest, as their study can help explain the early presence of sweet potato in Polynesia [16,44]. 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 S6B). This specimen is possibly the oldest sweet potato collection from Polynesia. We successfully sequenced Banks and Solander's specimen using genome skimming and the good quality of the sequences retrieved allowed us to assemble its whole chloroplast genome, as well as to identify fragments of multiple nuclear regions targeted in this study (see STAR Methods). Our analyses confirm that this specimen belongs to sweet potato CL1, i.e. the sweet potato chloroplast lineage that did not capture the chloroplast from *I. trifida* (Figure 7A) [16]. Also, the longer branch in the nuclear tree (Figure S7) indicates that this specimen is distinct from other specimens in that lineage. We subsequently used whole chloroplast data to estimate the divergence time of Banks and Solander's specimen from its closest relative; we constructed a conventional time-calibrated phylogeny and performed a coalescent analysis with all sequenced specimens of *I. batatas* and *I. trifida* (see STAR Methods). Both indicated that the lineage to which this specimen belongs diverged from its closest relative at least 111,500 years ago (at least 139,000 years ago in the coalescent analysis; Figure 7B). This result, together with the distinct admixture pattern (Figure 7C), is congruent with the long-term isolation of this distinct variety in comparison to varieties from Central and South America. In summary, our data strongly suggest that the presence of sweet potato in Polynesia predates

human colonization of the region by thousands of years, and consequently is most probably due to long-distance dispersal, which we have shown is a relatively common occurrence within the genus *Ipomoea*.

Conclusions

Our sequence data and species-level sampling represent the most comprehensive dataset yet published to address the origin and evolution of sweet potato. Our results convincingly demonstrate that nuclear and chloroplast genomes provide conflicting phylogenies for the relationship between *Ipomoea batatas* and *I. trifida*. We consider that the narrative most consistent with our results is that *I. batatas* evolved by autopolyploidy from *I. trifida*, within the current range of *I. trifida* in Central and Northern South America. Subsequent to the divergence of *I. batatas* from *I. trifida*, the two species hybridized and the footprint of that event is reflected by the presence of two strongly supported chloroplast lineages within *I. batatas* due to chloroplast capture.

Our time-calibrated phylogenies offer rough estimates of the chronology of sweet potato evolution. We acknowledge that estimating the ages of lineages from phylogenies [35,45,46] is fraught with potential errors, but we thought it important to assemble sufficient evidence to demonstrate successfully that sweet potato and its tuber evolved in the pre-human era.

Our results are also intriguing with regard to the presence of sweet potato in Polynesia, providing strong support for its presence there as a result of naturally occurring long-distance dispersal. Over the last twenty years, long-distance dispersal has emerged as a common explanation for disjunct patterns of plant distribution [47], thus the presence of an American plant in Polynesia is not as surprising as once thought. Several examples of similar, undoubtedly natural, long-distance dispersal in close relatives of sweet potato make it even less surprising. Additional support is provided by the earliest specimen of sweet potato

collected from Polynesia. This has a unique genetic signature suggesting it diverged from its other samples on the American continent more than 100,000 years ago. The evidence against human-mediated transport of sweet potato to Polynesia is, therefore, extremely strong.

ACKNOWLEDGMENTS

We acknowledge the financial support of The Leverhulme Trust for our *Ipomoea* Foundation Monograph project and the University of Oxford through The John Fell Fund for travel and sequencing costs. PMR was funded by a BBSRC scholarship granted through the Interdisciplinary Bioscience DTP Programme and received additional funding from a Santander Travel Award for his stay at Oregon State University. RWS and PMR acknowledge funding from the BBSRC GCRF-IAA fund (BB/GCRF-IAA/16 and BB/GCRF-IAA/17/16). TC was funded by a NERC scholarship granted through the Environmental Research DTP Programme. We thank the curators from BM, BOLV, CIP, E, HSB, HUEFS, K, LPB, MA, OXF, RB, SAN, US and USZ herbaria for granting access to their collections, and also Mark Carine at BM for his help in sampling Banks and Solander's specimen and his comments on the manuscript. We also thank David Swofford for his help with SVDQuartets and Richard Cronn for use of his lab in preparing DNA from the Banks & Solander specimen for sequencing and subsequent analysis. Finally, we thank Barbara Kennedy from Bishop Museum, Hawaii, for her help with Hawaiian material, and two anonymous reviewers for helpful comments.

AUTHOR CONTRIBUTIONS

Conceptualization, Supervision and Project Administration, R.W.S.; Funding Acquisition: R.W.S, P.M.R., T.C.; Methodology, R.W.S., A.L., S.K., P.M.R. and T.C.; Resources, J.R.I.W., B.R.M.W., P.M.R., D.E., N.L.A., L.G. and M.D.R.; Formal analysis and Investigation, P.M.R. and T.C.; Writing – Original draft, P.M.R., R.W.S., T.C. and J.R.I.W.; Writing – Review & Editing, all authors; Visualization, P.M.R., T.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Lebot, V. (2010). Tropical roots and tuber crops. In Soil, Plant Growth and Crop Production, W. H. Verheye, ed. (University of Gent, Belgium: EOLSS Publishers Co Ltd).

2. Food and Agriculture Organization of the United Nations (2017). FAOSTAT Statistics Database. Available at: http://faostat3.fao.org/home/ [Accessed December 4, 2017].

3. 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.

4. Zhang, D., Cervantes, J., Huamán, Z., Carey, E., and Ghislain, M. (2000). Assessing genetic diversity of sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars from tropical America using AFLP. Genet. Resources Crop Evol. *47*, 659–665.

5. 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.

6. 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.

7. Roullier, C., Duputié, A., Wennekes, P., Benoit, L., Fernández Bringas, V.M., Rossel, G., Tay, D., McKey, D., and Lebot, V. (2013). Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). PLoS ONE *8*, e62707.

8. 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.

9. Austin, D.F. (1988). 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 (CIP)), pp. 27–60.

10. 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.

11. 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.

12. Miquel, F.A.W. (1856). Flora van Nederlandsch Indië (Amsterdam: C. G. van der Post).

13. Candolle, A. de (1883). Origine des plantes cultivées 1st ed. (Paris: Germer Baillière).

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

15. Yen, D.E. (1974). The sweet potato and Oceania. An essay in Ethnobotany (Honolulu, Hawaii: Bishop Museum Press).

16. Roullier, C., Benoit, L., McKey, D.B., and Lebot, V. (2013). 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.

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

18. 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).

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

20. Bulmer, R. (1965). Birds as possible agents in the propagation of the sweet-potato. Emu *65*, 165–182.

21. Rossel, G., Kriegner, A., and Zhang, D.P. (1999). From Latin America to Oceania: the historic dispersal of sweetpotato re-examined using AFLP. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.504.5148.

22. Montenegro, Á., Avis, C., and Weaver, A. (2008). Modelling the prehistoric arrival of the sweet potato in Polynesia. J. Arch. Sci. *35*, 355–367.

23. O'Brien, P.J. (1972). The sweet potato: its origin and dispersal. Am. Anthropol. *74*, 342–365.

24. 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.

25. 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*, p3209-3215.e6.

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

27. McDonald, J.A., and Austin, D.F. (1990). Changes and additions in *Ipomoea* sect. *Batatas*. Brittonia *42*, 116–120.

28. 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.

29. Nishiyama, I., Fujise, K., Teramura, T., and Miyazaki, T. (1961). 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.

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

31. Yu, L.-X., Liu, M.-Y., CAO, Q.-H., Yu, Y.-C., Xie, Y.-P., Luo, Y.-H., Han, Y.-H., and Li, Z.-Y. (2014). Analysis of nrDNA ITS sequences in *Ipomoea batatas* and its relative wild species. Plant Sci. J. *32*, 40–49.

32. 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.

33. Reiseberg, L.H., and Soltis, D.E. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Pl. *5*, 65–84.

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

35. Britton, T. (2005). Estimating divergence times in phylogenetic trees without a molecular clock. Syst. Biol. *54*, 500–507.

36. Zhu, T., Mario, D.R., and Ziheng, Y. (2015). Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. Syst. Biol. *64*, 267–280.

37. Ossowski, S., Schneeberger, K., Lucas-lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D., and Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science *327*, 92–94.

38. Rannala, B., and Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics *164*, 1645–1656.

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

40. Khoury, C.. (2015). The conservation and use of crop genetic resources for food security. Available at: http://edepot.wur.nl/352830.

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

42. 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.

43. Ridley, H.N. (1930). The dispersal of plants throughout the world (Lloyds Bank Buildings, Ashford, Kent: L. Reeve & Co., Ltd).

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

45. Dos Reis, M., and Yang, Z. (2013). The unbearable uncertainty of Bayesian divergence time estimation: uncertainty in divergence time estimation. J. Syst. Evol. *51*, 30–43.

46. Wilf, P., and Escapa, I.H. (2015). Green Web or megabiased clock? Plant fossils from Gondwanan Patagonia speak on evolutionary radiations. New Phytol. *207*, 283–290.

47. Lavin, M., Schrire, B.P., Lewis, G., Pennington, R.T., Delgado-Salinas, A., Thulin, M., Hughes, C.E., Matos, A.B., and Wojciechowski, M.F. (2004). Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. Philos. Trans. Royal Soc. B *359*, 1509–1522.

48. Wood, J.R.I., Carine, M.A., Harris, D., Wilkin, P., Williams, B., and Scotland, R.W. (2015). *Ipomoea* (Convolvulaceae) in Bolivia. Kew Bull. *70*, 71.

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

50. MYcroarray (2015). MYbaits. In-solution sequence capture for targeted High-Throughput Sequencing.

51. 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. Appl. Pl. Sci. *2*, 1400042.

52. 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. Am. J. Bot. *99*, 349–364.

53. 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.

54. 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. Am. J. Bot. *94*, 275–288.

55. Ratan, A. (2009). Assembly algorithms for Next Generation Sequence data. Available at: https://etda.libraries.psu.edu/files/final_submissions/587.

56. 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.

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

58. 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.

59. 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.

60. 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.

61. 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.

62. 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.

63. 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.

64. Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

65. 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.

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

67. 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.

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

69. 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.

70. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood trees for large alignments. PLoS ONE *5*, e9490.

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

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

73. 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. Theor. Biol. *374*, 35–47.

74. Swofford, D.L. (2002). Phylogenetic Analysis Using Parsimony (*and other methods) (Sunderland, Massachusetts: Sinauer Associates).

75. 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.

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

77. Simmons, M.P., and Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. *49*, 369–381.

78. Müller, K. (2005). SeqState - primer design and sequence statistics for phylogenetic DNA data sets. Appl. Bioinf. *4*, 65–69.

79. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. *51*, 492–508.

80. 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.

81. Clement, M., Snell, Q., Walke, P., Posada, D., and Crandall, K. (2002). TCS: estimating gene genealogies. In Proc 16th Int Parallel Distrib Process Symp, p. 184.

82. 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. American Journal of Botany *101*, 1987–2004.

83. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution *30*, 2725–2729.

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

85. 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.

86. 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.

87. 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.

88. 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.

89. 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*.

90. 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.

91. Rambaut, A., Suchard, M.A., Xie, D., and Drummond, A.J. (2014). Tracer v1.6 Available at: http://beast.bio.ed.ac.uk/Tracer.

FIGURE TITLES AND LEGENDS

Figure 1. Sweet potato and five closely related species.

(A) *Ipomoea batatas* (sweet potato); (B) *I. trifida*; (C) *I. triloba*; (D) *I. ramosissima*; (E) *I. cordatotriloba* (South America); (F) *I. leucantha*.

Figure 2. Sweet potato evolved by autopolyploidy from *Ipomoea trifida*

(A) Nuclear phylogeny of *Ipomoea* sect. *Batatas* inferred from 307 nuclear regions that do not show recombination, using Astral-II. Values at the nodes are bootstrap support values (100 replicates from gene trees), and black dots indicate 100% support. Blue, perennial species; green, annual species; orange, sweet potato and *Ipomoea trifida*. All species except *Ipomoea cordatotriloba*, *I. grandifolia* and *I. cynanchifolia* are monophyletic with 100% support.

(B) Summary of gene tree topologies inferred considering multiple loci from all sweet potato specimens. Each bar represents a sweet potato specimen. Each of the colors represents the percentage of genes for which: only one haplotype could be distinguished (dark green); six distinct haplotypes were assembled and form a clade in the gene tree (light green); six haplotypes were assembled and at least one allele groups with alleles from other sweet potato specimens (yellow); six haplotypes were assembled and one or more alleles group with alleles from other species in *Ipomoea* series *Batatas* (red).

See also Figures S1–S4.

Figure 3. Conflicting topologies from nuclear and chloroplast genomes.

(A) Summarized nuclear and chloroplast phylogenies and population structure analyses (K = number of assumed ancestral populations) of *Ipomoea batatas* and *I. trifida*. The analysis of nuclear regions shows sweet potato is monophyletic, whereas the analysis of chloroplast genomes supports a non-monophyletic sweet potato. The population structure analysis of nuclear data shows no discernible internal structure in *Ipomoea batatas*, whereas there are two distinct groups in the chloroplast data.

(B) Integer neighbor-joining network and distribution map of all sweet potato and *I. trifida* specimens in this study. Yellow dots, *I. trifida*; blue and red dots, sweet potato chloroplast lineages. Hash marks in the network represent mutations. The green area on the map represents the global distribution of *I. trifida*.

See also Figures S1–S5 and Data S3.

Fig. 4. Hybridization of sweet potato and *Ipomoea trifida* following speciation.

(A) *Ipomoea batatas* most probably originated from *I. trifida* more than 800,000 years ago in a region between Central America and northern South America (blue ellipse in A_1), which is the current distribution of *I. trifida*. The two species subsequently diverged and *I. batatas* expanded its distribution range beyond its progenitor's. Further hybridization in the sympatric area within 56,000 years of speciation resulted in *I. batatas* capturing the chloroplast from *Ipomoea trifida* (red area in A_2), whereas the more distant chloroplast lineage of *Ipomoea batatas* maintained the original chloroplast. Following chloroplast capture, both lineages expanded their distribution area, either by natural means or by recent human transportation, and hence the pattern observed today (A_3).

(B and C) Two mechanisms of chloroplast capture by *Ipomoea batatas* from *I. trifida*. Orange and yellow represent *I. batatas* and *I. trifida* nuclear genomes respectively; purple represents

original *I. batatas* chloroplast; light green represents the captured *I. trifida* chloroplast; dark green represents current *I. trifida* chloroplast.

(B) The hybridization between *I. batatas* and *I. trifida* would produce a new allotetraploid entity, and subsequently a new hexaploid form by further hybridization with *I. trifida* and a new polyploidization. Subsequently, the nuclear component of *I. trifida* in the newly formed hexaploid entity (yellow color in the pie chart) would be lost after several generations of introgression with ancestral *I. batatas*, resulting in the conflicting topologies retrieved from nuclear and chloroplast genomes.

(C) Alternatively, 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*.

Figure 5. Sweet potato diverged from *Ipomoea trifida* in pre-human times.

Time-calibrated phylogenies of sweet potato and its CWRs inferred using (A) 21 nuclear regions and (B) whole chloroplasts respectively. The 95% HDP for the temporal duration of the branch ancestral to *Ipomoea trifida* and *I. batatas* CL2 in (B) is 4-56,000 years. Node bars represent 95% HDP 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.

See also Table S1.

Figure 6. Long distance dispersal in *Ipomoea* evidenced by a highly disjunct species nested within a Mexican clade.

(A) Distribution map of *Ipomoea tuboides* (photograph by Forest & Kim Starr) and its close relatives, and simplified nuclear phylogeny. *Ipomoea tuboides* is endemic to the Hawaiian Islands, but all its closest relatives are in Mexico and Central America, suggesting long-distance dispersal. Orange area indicates the global distribution 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 tree.

(B) Time-calibrated phylogeny of the species in the Tuboides group inferred using 21 nuclear regions. Node bars represent 95% HDP 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 7. Signature from Banks and Solander specimen is congruent with long-term isolation in Polynesia

(A) Position of Banks and Solander specimen (in green) in a chloroplast phylogenetic network in sweet potato chloroplast lineage 1.

(B) Coalescent dated phylogeny of *I. batatas* chloroplast lineage 1 shows 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% HDP intervals for node ages.

(C) Population structure analysis based on 5,735 nuclear variable positions.

See also Figures S6–S7 and Data S2.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources, reagents and scripts should be directed to and will be fulfilled by the Lead Contact, Professor Robert W. Scotland (robert.scotland@plants.ox.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Our dataset consists of 199 individuals representing all sixteen species in *Ipomoea* series *Batatas* and twenty-four other species across *Ipomoea* (passport data in Data S1). We included 72 *I. batatas* specimens from germplasm accessions and from different locations in America and the Old World. Most *I. batatas* and *I. trifida* samples were obtained from fresh material growing at the International Potato Center in Lima, Peru (CIP). DNA of the rest of specimens, including Banks and Solander's collection from Polynesia, was obtained from herbarium specimens collected between 1769 and 2014. All species from the Batatas group except three (*Ipomoea lactifera, I. tabascana* and *I. tenuissima*) are represented by multiple specimens from different geographical locations. *Ipomoea tenuissima* is a poorly known Caribbean species scarcely represented in herbaria, whereas both *I. tabascana* and *I. lactifera* are known from one and a few populations respectively [27,48].

METHOD DETAILS

Character sampling and target capture probes

We developed probes targeting 605 putative single copy nuclear regions of *Ipomoea* (see Data S1) through comparison of genomic data from *I. lacunosa* and coding sequence (CDS) of *Solanum tuberosum*. 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 bp. Along these loci, 100 bp RNA probes were developed by

MycroArray (Ann Arbor, MI), excluding probes with GC content < 25%. We also obtained the whole chloroplast genome of all specimens.

DNA extraction and library preparation

We extracted DNA from fresh material using CTAB method [49], and from herbarium specimens using the Plant Tissue Mini protocol for QIAGEN DNEasy Plant Mini Kit (Qiagen). We created genomic libraries using the NEBNext Ultra DNA Library Prep Kit for Illumina v.3.0. (New England BioLabs).

Hybridization and DNA sequencing

We implemented target enrichment using MYBaits [50] to capture nuclear regions of interest, following the protocol described in [51] 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 Internal Transcribed Spacer (*rDNA ITS*) region with genome skimming [52]. Sequencing was conducted using the Illumina HiSeq 3000 at the Center for Genome Research and Biocomputing, Oregon State University (Corvallis, United States). Sequences were trimmed for Illumina adapters and for quality, Q15 on the left and Q10 on the right of the reads. 100 bp paired reads were obtained.

Banks & Solander's specimen was sequenced using the MiSeq and 25bp paired reads, instead of target enrichment. We evaluated the degree of DNA damage in this specimen using mapDamage 2.0 [53] and found no signs of damage different from levels found in other herbarium specimens (see Data S2).

Beauregard and Tanzania sweet potato varieties were sequenced for the highly variable chloroplast *rpl32-trnL* region [54], using Sanger sequencing at Source BioScience.

Assembly of nuclear regions

We conducted a three-stage assembly process: first we generated draft gene assemblies with YASRA [55] that served as target regions in a second assembly run using PRICE [56]. We finally implemented SSPACE [57] to extend the gene assemblies. Final assembled contigs were aligned back to the reference sequences using BLASTN [58] to target assembled contig assignments.

Haplotype identification in nuclear data

We collected information on ploidy levels of the species from the literature and from CIP. We aligned the nuclear raw reads back to the assembled contigs using Bowtie [59]. From this alignment, we created a variant call file that described the SNPs found within the alignment. We then ran Hapcompass [60] 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 [61] considering independent alleles for all genes and samples, and found no significant intra-specimen variation (Figure 2B). We therefore conducted all subsequent phylogenetic analyses using consensus sequences.

Assembly of chloroplast genomes and rDNA ITS

We assembled the chloroplast genomes and the *ITS* region using SPAdes genome assembly algorithm [62], using as reference the chloroplast genome of *Ipomoea batatas* cultivar Xushu18 [63] and the full *ITS* fragment (including 5.8S region) of an *I. batatas* herbarium specimen (*C. Whitefoord 71*) previously sequenced using Sanger. Chloroplasts show the general structure in angiosperms, with one long single copy, one short single copy and two inverted repeats. Chloroplast size ranges from 160,382 to 174,715 base pairs, except for *Ipomoea lactifera* which presents several large deletions (150,628 base pairs).

Assembly of Banks and Solander specimen

The reads obtained using MySeq allowed us to target several fragments across the nuclear regions (1,016 reads mapped). We assembled into contigs only those read pairs where both reads matched the reference sequence at approximately the expected distance or 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. We further explored DNA degradation in this specimen by calculating base percentages in these variable positions and found no differences compared with more recent material (see Data S1).

Phylogenetic analysis of nuclear regions

We aligned every nuclear region individually using L-INS-I strategy in MAFFT v7.271 [64,65] (gap penalty = 1.53), and used default parameters in Gblocks [66,67] to remove poorly aligned positions from the alignment. We estimated evolutionary models for each region using jModelTest 2 [68] and obtained independent gene trees using default parameters in FastTree 2.1.9 [69,70]. In a dataset this large, neither intralocus recombination, incomplete lineage sorting (ILS) nor reticulation can be discounted [32]. Therefore, we ran multiple analyses to evaluate the effect of these processes. First, to reduce the possible effect of recombination, we ran the PHI statistical test [71] to identify those regions in our dataset likely to contain recombination (see Data S1). We ran all subsequent analyses using two data sets in parallel: one including all 605 regions, and another including only the 307 regions that did not show evidence of recombination according to the PHI test. In addition, to explore the effect of ILS we ran phylogenetic analyses using both coalescent-based and concatenated methods. First, we used gene trees as input to infer the species tree using Astral II [61]. Second, using the concatenated alignments we conducted Approximate Maximum Likelihood

as implemented in FastTree 2.1.9 [69,70], and SVDQuartets [72,73], a coalescent-based method available in PAUP 4.0 [74] (800,000,000 random quartets). We ran FastTree analysis using the CIPRES Science Gateway [75], and SVDQuartets using the supercomputer at University of Oxford Advanced Research Computing.

Phylogenetic analysis of chloroplast genomes

We aligned the chloroplast genomes using FFT-NS-2 strategy in MAFFT [64,65] (gap penalty = 1.53). The alignment was visually checked and minimal corrections were made in the poly-A and poly-T regions, only to minimize random alignment of these regions. We then used Gblocks [66,67] to remove poorly aligned positions and jModelTest 2.1.7 [68] to estimate the best substitution model for this alignment (GTR+I+G). We conducted Maximum Likelihood analysis using RAxML 8.0 [76] as implemented in CIPRES [75] (1,000 bootstrap replicates), and parsimony analysis using PAUP 4.0 [74] (1,000,000 trees based on 1,294 parsimony-informative characters, best tree = 2,631 steps). We also performed a parsimony analysis of 282 parsimony informative indels in PAUP (100,000 trees, best tree = 975 steps), coding them as presence/absence [77] using SeqState 1.4.1 [78].

To evaluate the robustness of the topology showing two sweet potato gene pools, we additionally produced an alternative topology enforcing sweet potato monophyly using RAxML [76]. We evaluated both topologies using the approximately unbiased test [79] as implemented in IQ-Tree 1.5.0a [80] (see Data S3).

We generated three phylogenetic networks: one including all *Ipomoea batatas* and *I. trifida* specimens, another one including all species in the group, and the third one including all *I. batatas* specimens plus Banks and Solander (675, 1,051 and 522 segregating sites respectively). We used the Integer Neighbor-Joining method implemented in PopART (ε=1) [81]. To further confirm our results, we ran independent phylogenetic analyses of the most

variable regions of the chloroplast [54,82] (Figure S5B). We also estimated pairwise distances (p-distance) between all sweet potato accessions using Mega 6.0 [83].

Finally, we generated one additional phylogenetic network using the *rpl32-trnL* chloroplast region to identify what chloroplast lineage the two varieties used in breeding programmes belong.

Analysis of population structure

We randomly extracted 3,000 variable positions from the alignments of nuclear regions and used them as input for Structure [84,85] 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). We 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=0.3081; K = 1–5; 3 runs), and 5,735 variable positions from the nuclear alignments including Banks and Solander specimen (λ=0.3483; K = 1–5; 3 runs).

Divergence time estimation and population size

We implemented divergence time estimation in RevBayes [86,87], a graphical modeling 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 (passport data in Data S1). 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 = 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 [88]. 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 sweet potato occurred in 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 fossil relatives to *Ipomoea*) [89,90]. 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.
A matrix containing samples from throughout *Ipomoea* based on 21 nuclear genes for which there was high coverage (99%) was then used to infer divergence times within the genus such as the crown nodes for *Ipomoea* series *Batatas* and the Tuboides clade. A GTR+G+I model was implemented, and branch-specific substitution rates were inferred using an uncorrelated lognormal relaxed clock with a standard deviation 0.2972. A single set of parameters for nucleotide substitution and branch-specific substitution rates were estimated 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 dataset nor the nuclear datasets 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.

We also conducted a multispecies coalescent analysis on all sequenced plastomes for *Ipomoea batatas* and *I. trifida*. We conducted this analysis to estimate effective population sizes for species and ancestral lineages within this clade [38] (*I. batatas* lineage 1, *I. batatas* lineage 2, *I. trifida*) and to infer when potential population bottlenecks associated with the origin of this crop are likely to have occurred. Of particular interest was whether a bottleneck is associated with the population in which chloroplast capture may have occurred (in the chloroplast phylogeny inferred in this study, this corresponds to the ancestral lineage of *I. trifida* and *I. batatas* lineage 2). In this analysis, we used fixed species and gene tree topologies in accordance with those inferred in phylogenetic analyses in this study. Specifically, *I. batatas* lineage 2 is designated as the sister taxon of *I. trifida*. A GTR+G+I model of sequence evolution was implemented, and overall rates of sequence evolution were assumed 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 (three taxa) was assumed to evolve under a constant rate of speciation and extinction. The age for the root node of the species tree was determined by the sampled ages for the equivalent node in the time-calibrated phylogeny for *Ipomoea* series *Batatas* inferred from the plastome dataset.

We performed all the analyses described above simultaneously in a single graphical model which was constructed in RevBayes. This allows uncertainty in parameter estimation to be integrated effectively across different analyses. Separate tree models (constant rate birth-death branching processes) were implemented in each analysis to account for the large variation in the intensity of taxon sampling. The model was run two times independently for 500,000 generations, sampling every 100 generations. Sufficient mixing and convergence between runs was assessed in Tracer v1.6 [91].

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.

QUANTIFICATION AND STATISTICAL ANALYSIS

The approximately unbiased test [79] to evaluate the robustness of the chloroplast topology was run in IQ-Tree 1.5.0a [80] using the RELL method with 100,000 resamplings.

DATA AND SOFTWARE AVAILABILITY

All data will be made available prior to publication via publicly accessible repositories.

DATA FILES

Data S1. Passport data of the specimens included in this study, Related to STAR Methods.

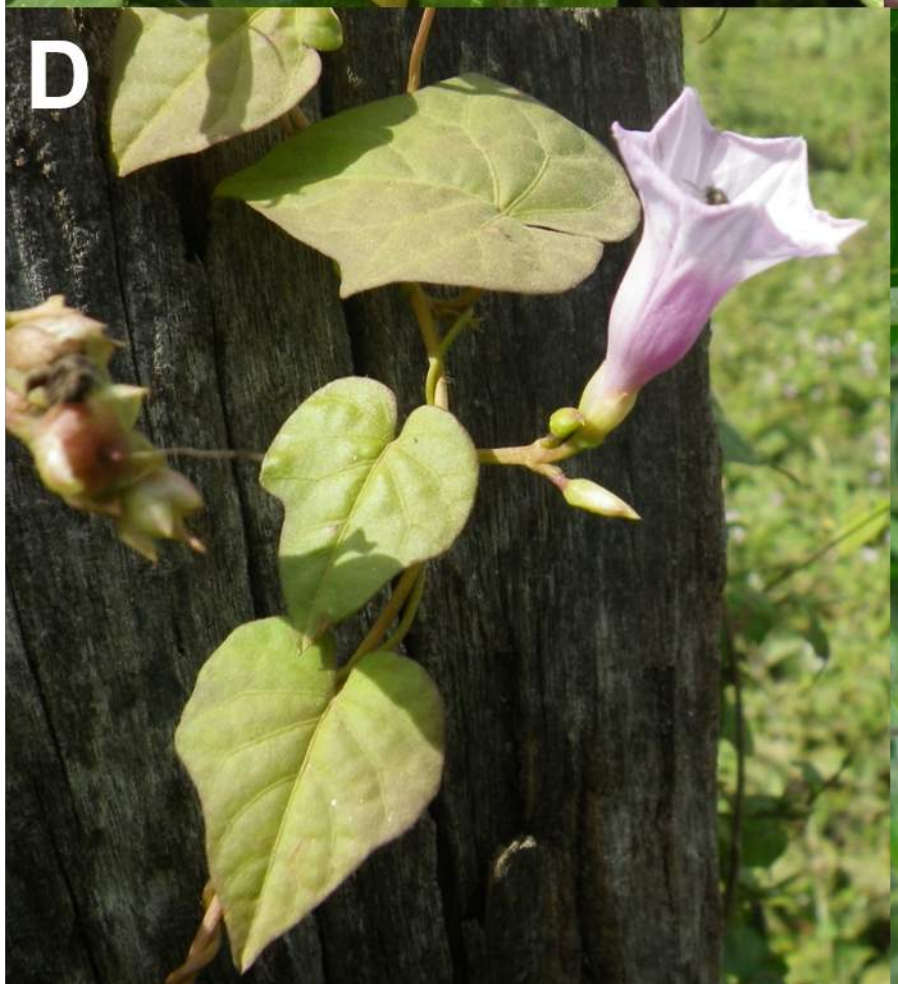
Data S2. MapDamage analysis of Banks and Solander's specimen, Related to Figure 7.

Data S3. Results of the Approximately Unbiased test, Related to Figure 3.


A


B


C


D


E


F

A

Ipomoea batatas
Ipomoea trifida
Ipomoea cordatotriloba (S.Am. specimens)
Ipomoea grandifolia
Ipomoea cynanchifolia
0.88
Ipomoea tenuissima
Ipomoea triloba
Ipomoea lacunosa
Ipomoea cordatotriloba (N.Am. specimens)
Ipomoea ramosissima
Ipomoea tiliacea
Ipomoea littoralis
Ipomoea lactifera
Ipomoea splendor-sylvae
ALL OTHER SPECIES

Sweet potato & closest relative

Annual species

Perennial species

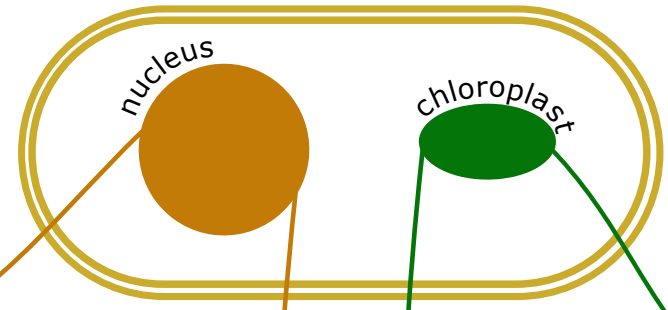
B

I_batatas_CIP400502
I_batatas_CIP441494
I_batatas_CIP400819
I_batatas_CIP401499
I_batatas_CIP401150
I_batatas_CIP400493
I_batatas_CIP400826
I_batatas_CIP400803
I_batatas_CIP441427
I_batatas_CIP400155
I_batatas_CIP400815
I_batatas_CIP401101
I_batatas_CIP400811
I_batatas_CW_71
I_batatas_CIP441498
I_batatas_CIP401062
I_batatas_CIP400788
I_batatas_CIP400238
I_batatas_CIP401443
I_batatas_CIP400784
I_batatas_CIP400822
I_batatas_CIP400466
I_batatas_CIP441113
I_batatas_CIP401067
I_batatas_CIP400435
I_batatas_CIP400356
I_batatas_CIP441431
I_batatas_CIP401210
I_batatas_CIP401217
I_batatas_CIP400274
I_batatas_CIP401056
I_batatas_CIP400519
I_batatas_CIP400389
I_batatas_CIP400162
I_batatas_CIP401129
I_batatas_JRIW_27946
I_batatas_CIP400205
I_batatas_CIP401404
I_batatas_CIP401068
I_batatas_CIP400442
I_batatas_CIP400572
I_batatas_CIP401079
I_batatas_CIP400507
I_batatas_CIP400796
I_batatas_CIP400339
I_batatas_CIP401459
I_batatas_CIP440453
I_batatas_CIP400361
I_batatas_CIP400433
I_batatas_CIP440551
I_batatas_MN_37607
I_batatas_CIP420882
I_batatas_DMC_407
I_batatas_CIP400218
I_batatas_CIP420602
I_batatas_CIP400423
I_batatas_CIP400551
I_batatas_CIP420068
I_batatas_CIP400786
I_batatas_CIP400157
I_batatas_CIP420386
I_batatas_CIP441314
I_batatas_CIP441416
I_batatas_CIP440615
I_batatas_CIP400033
I_batatas_CIP441177
I_batatas_CIP401523
I_batatas_CIP400941
I_batatas_CIP400453
I_batatas_CIP400287

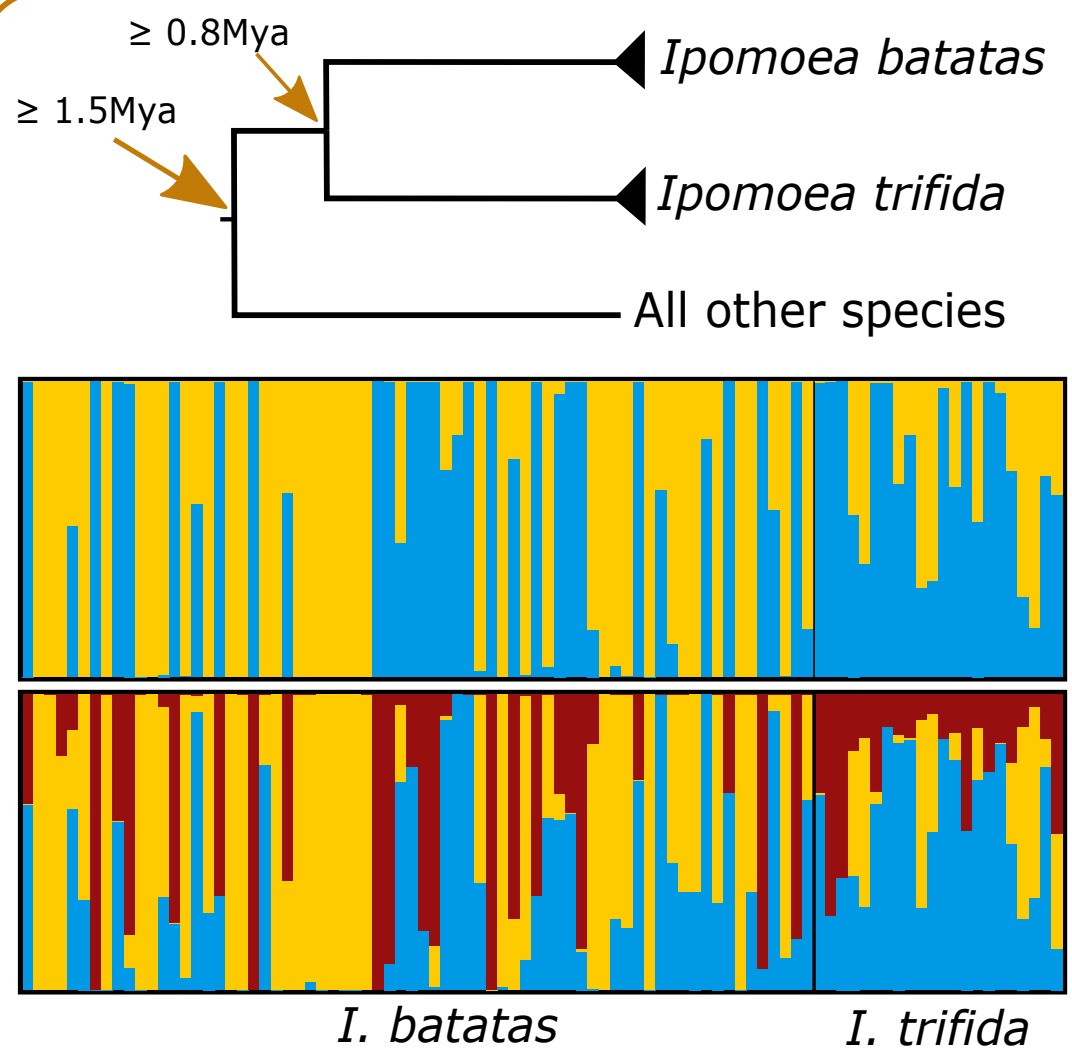
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Only one haplotype assembled
All six copies form a clade
At least one allele groups with alleles from other sweet potatoes
At least one allele groups with alleles from other species

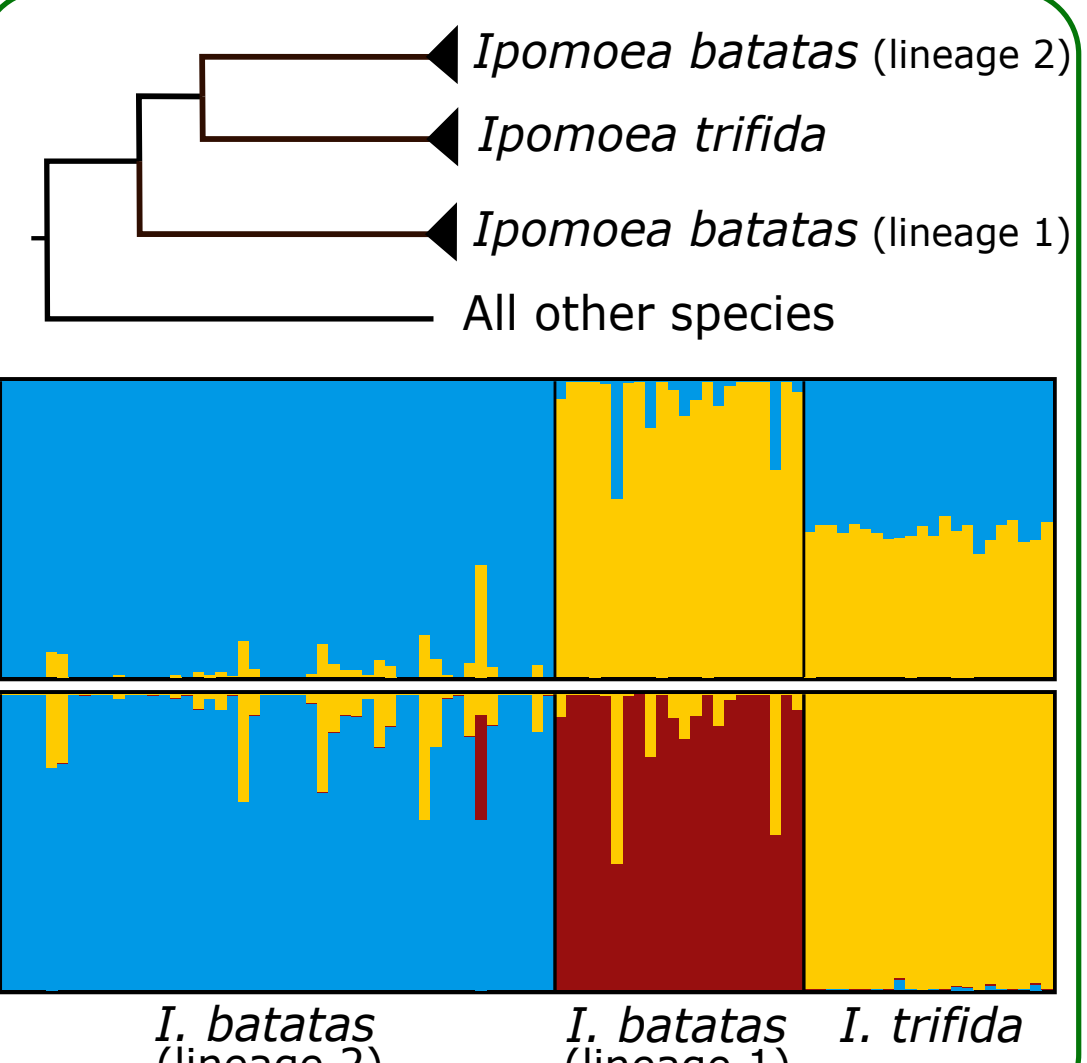
A



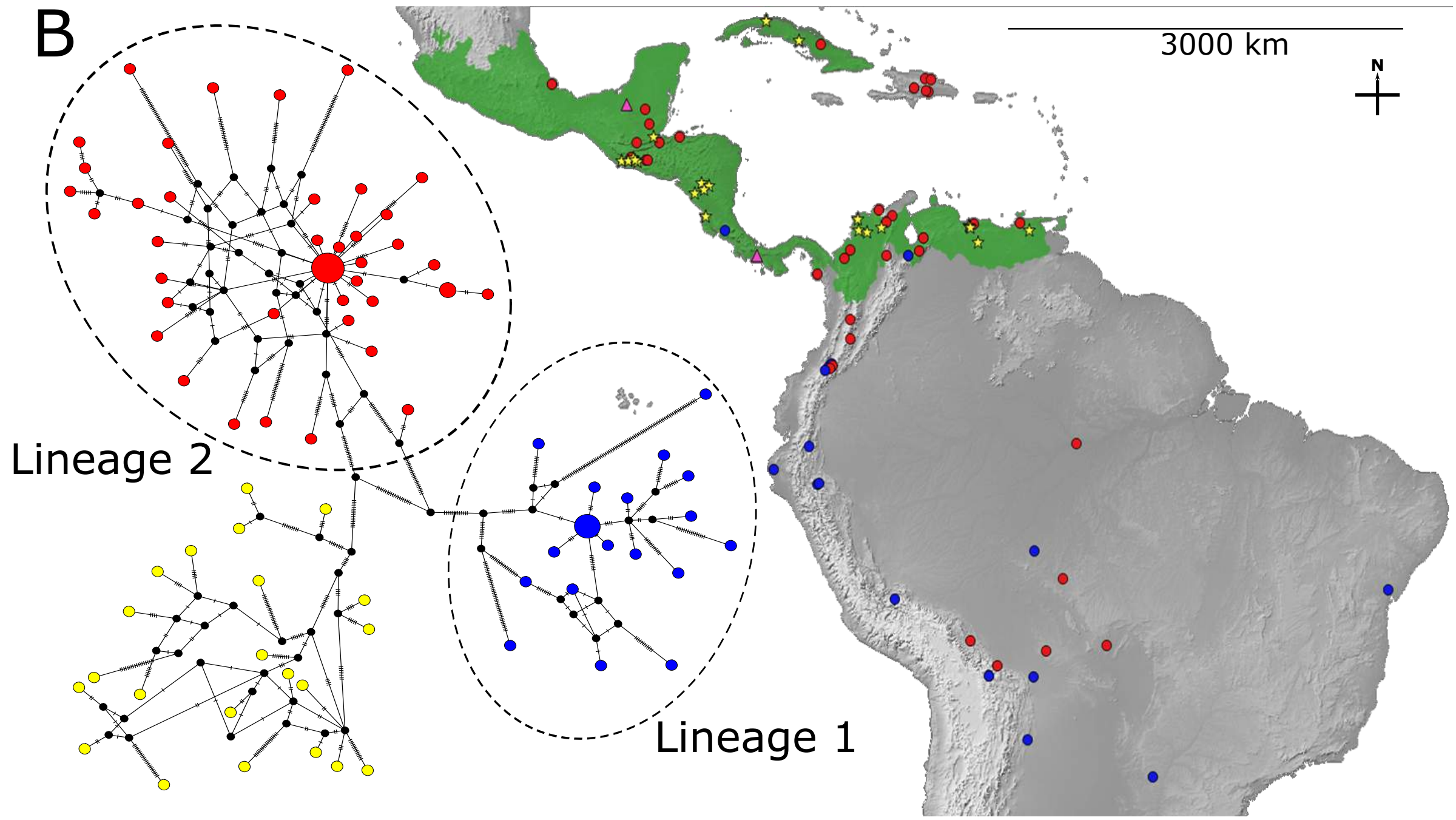
605 single copy regions

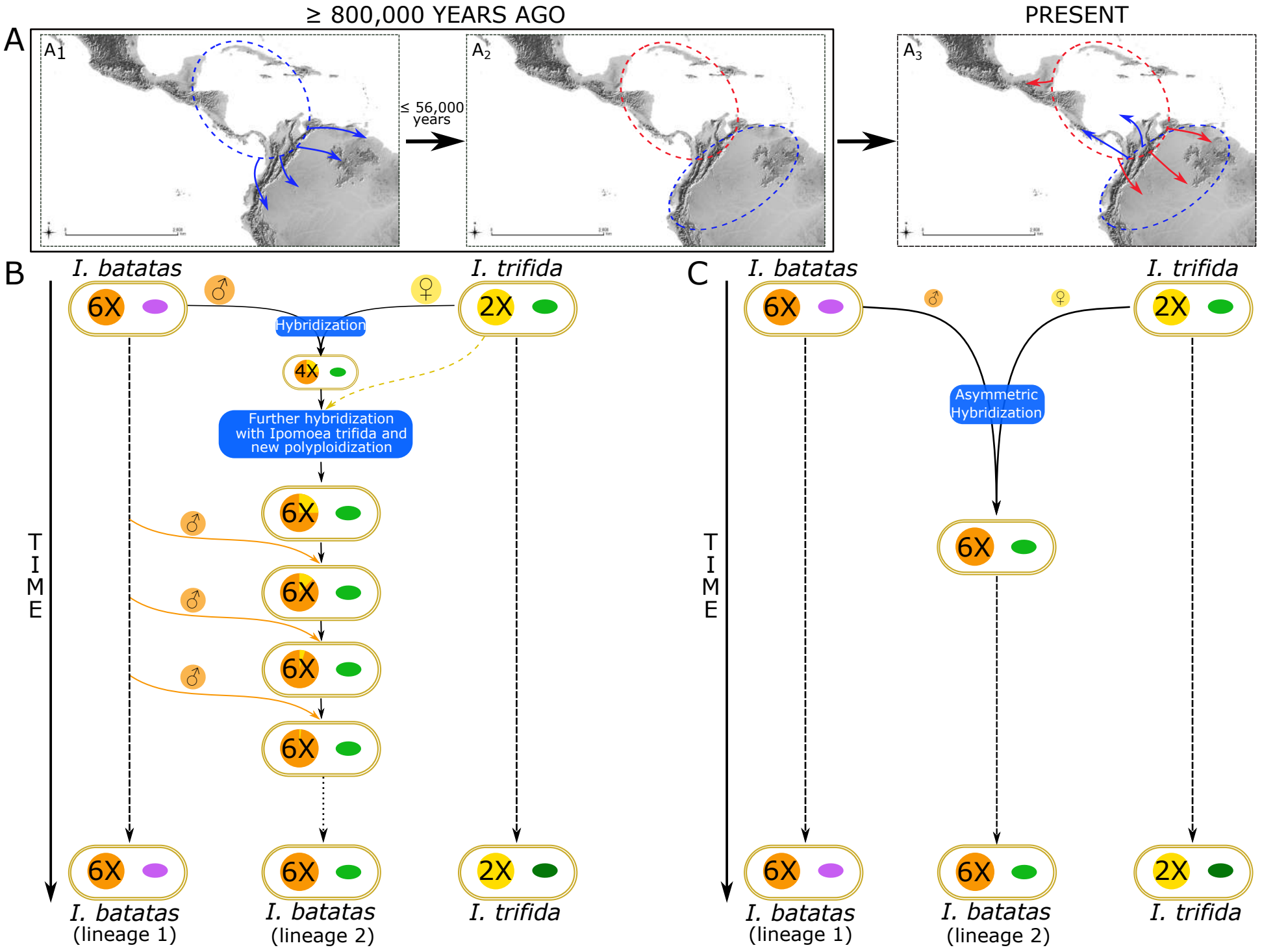


whole genome



B




≥ 800,000 YEARS AGO
PRESENT
A
A1
A2
A3
≤ 56,000 years
B
I. batatas
I. trifida
6X
2X
4X
Hybridization
Further hybridization with Ipomoea trifida and new polyploidization
TIME
I. batatas (lineage 1)
I. batatas (lineage 2)
I. trifida
C
Asymmetric Hybridization

A

Ipomoea batatas
Ipomoea trifida
Ipomoea cynanchifolia
Ipomoea grandifolia
Ipomoea cordatotriloba (South)
Ipomoea tenuissima
Ipomoea triloba
Ipomoea cordatotriloba (North)
Ipomoea lacunosa
Ipomoea lactifera
Ipomoea littoralis
Ipomoea ramosissima
Ipomoea tiliacea
Ipomoea splendor-sylvae

3.5 Mya 3 2.5 2 1.5 1 0.5 present

1.1Mya 0.8Mya

B

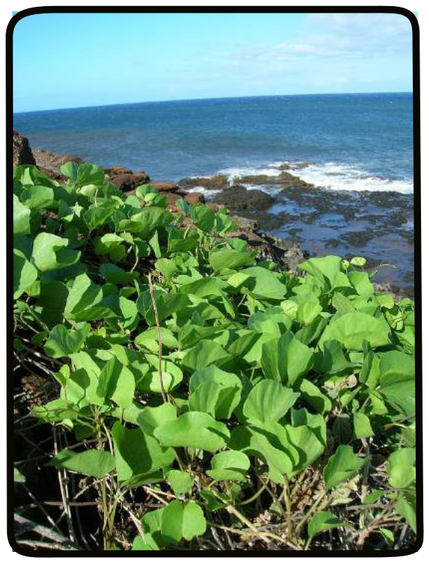
< 56,000 years

Ipomoea batatas (lineage 2)
Ipomoea trifida
Ipomoea batatas (lineage 1)
Ipomoea cordatotriloba (North)
Ipomoea lacunosa
Ipomoea tenuissima
Ipomoea triloba
Ipomoea cynanchifolia
Ipomoea grandifolia
Ipomoea cordatotriloba (South)
Ipomoea littoralis
Ipomoea tiliacea
Ipomoea ramosissima
Ipomoea splendor-sylvae

3.5 Mya 3 2.5 2 1.5 1 0.5 present

0.38Mya

A



Pacific
Ocean

5,200 km

1

≥ 1.1Mya

Ipomoea tuboides (1)
Ipomoea leucotricha (2)
Ipomoea decasperma (3)
Ipomoea orizabensis (4)
All other species

4
3
3
2

N
0
2,000
km

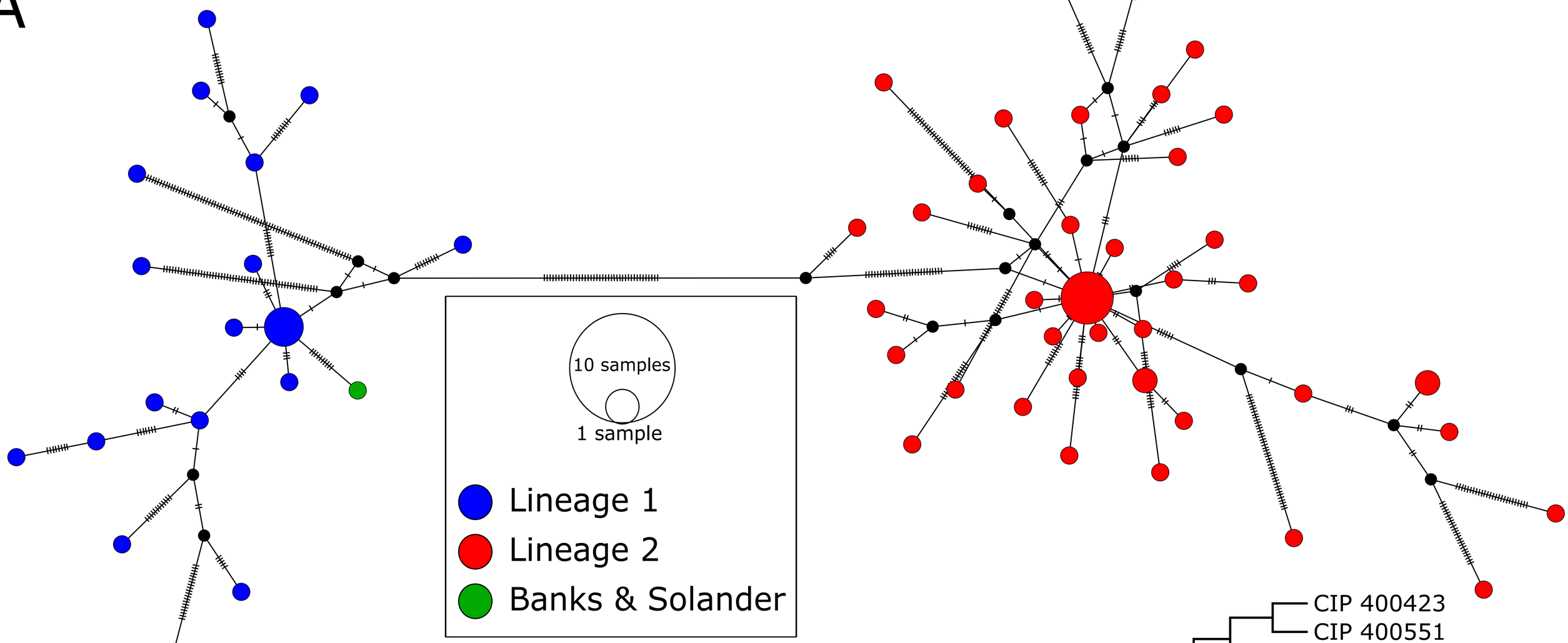
B

Ipomoea leucotricha
Ipomoea tuboides
Ipomoea decasperma
Ipomoea orizabensis
Ipomoea cardiophylla
Ipomoea marginisepala

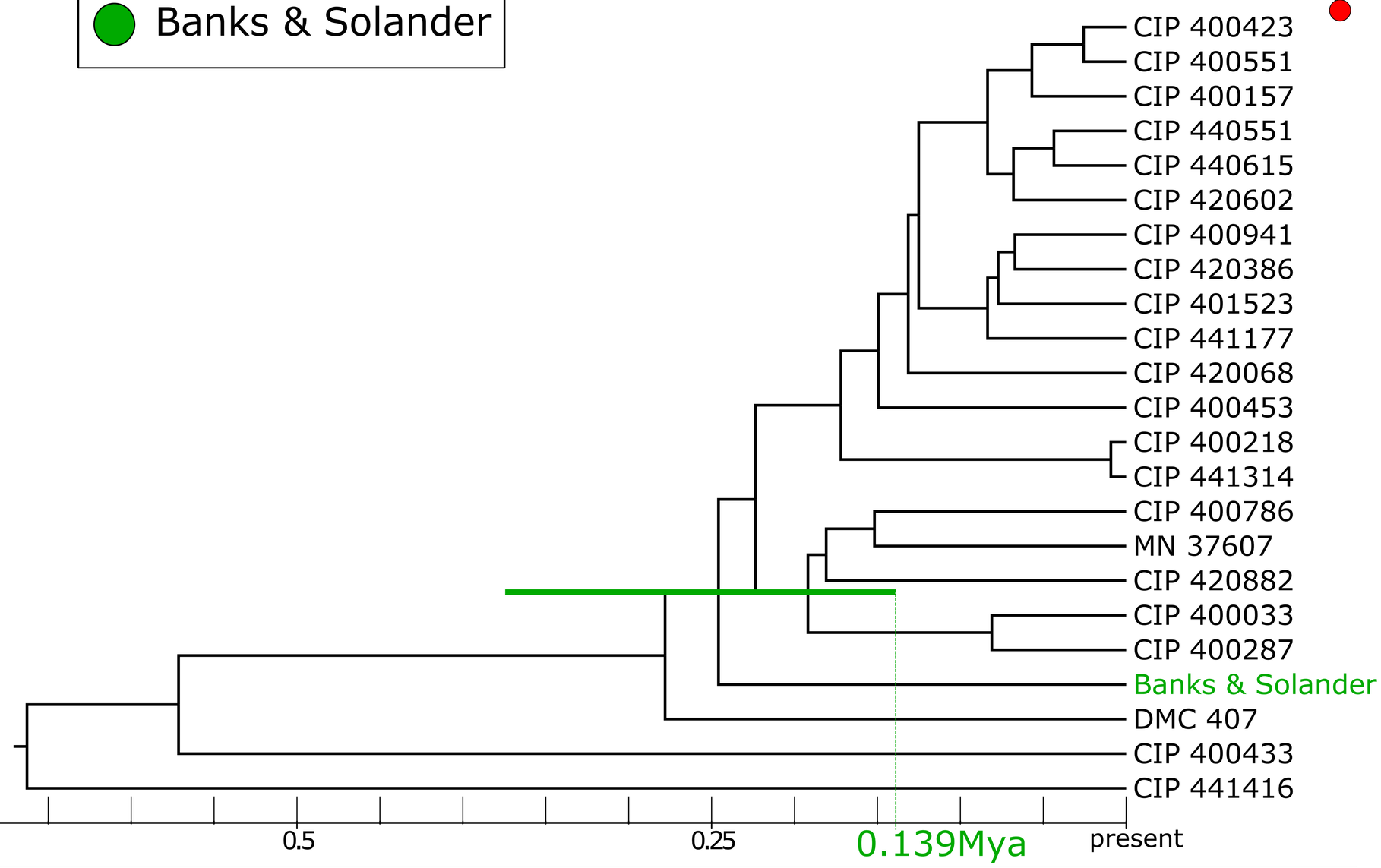
5 Mya
4
3
2
1
present

1.1 Mya

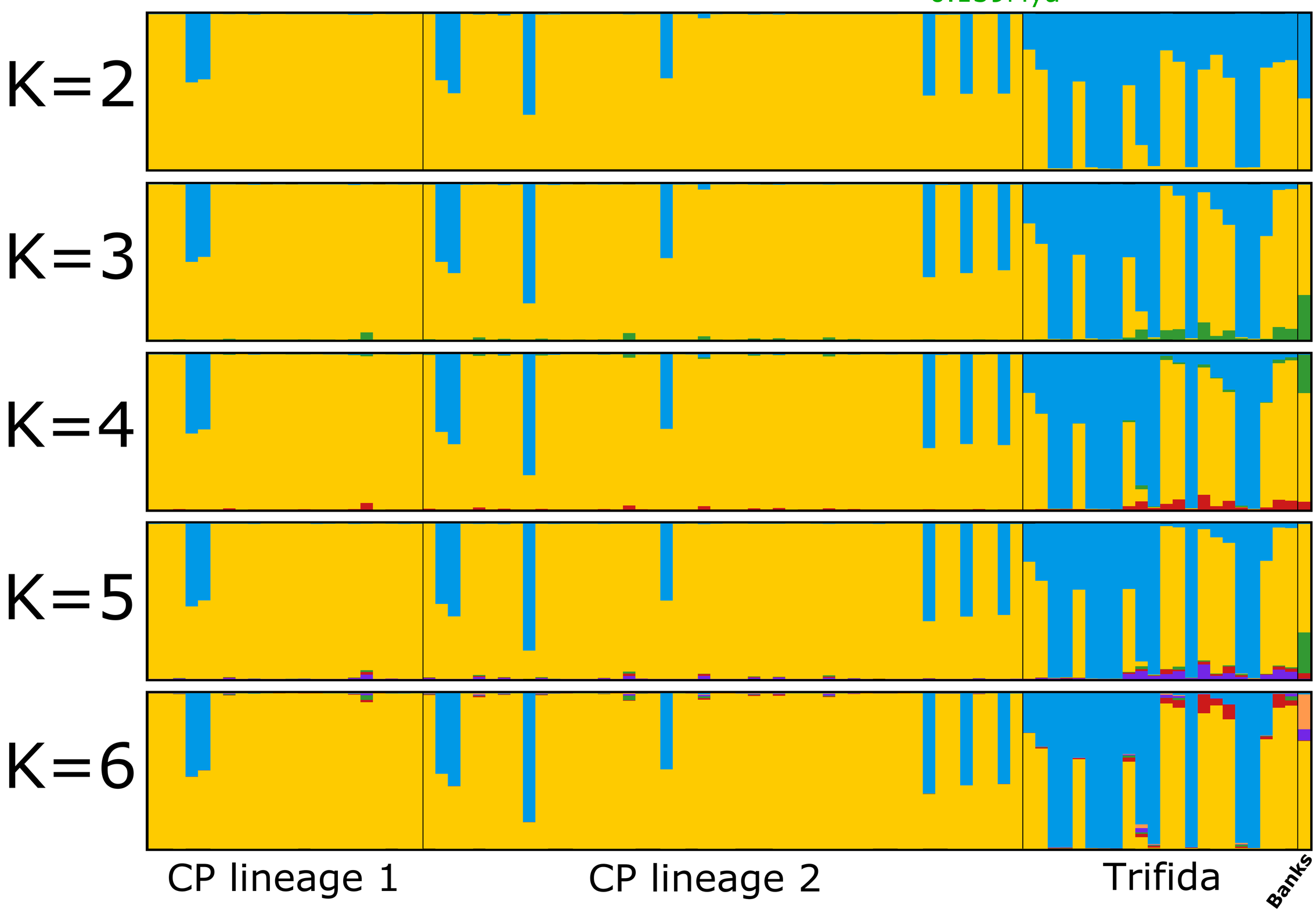
A


10 samples
1 sample
Lineage 1
Lineage 2
Banks & Solander

B


CIP 400423
CIP 400551
CIP 400157
CIP 440551
CIP 440615
CIP 420602
CIP 400941
CIP 420386
CIP 401523
CIP 441177
CIP 420068
CIP 400453
CIP 400218
CIP 441314
CIP 400786
MN 37607
CIP 420882
CIP 400033
CIP 400287
Banks & Solander
DMC 407
CIP 400433
CIP 441416
0.75
0.5
0.25
0.139Mya
present

C


K=2
K=3
K=4
K=5
K=6
CP lineage 1
CP lineage 2
Trifida
Banks

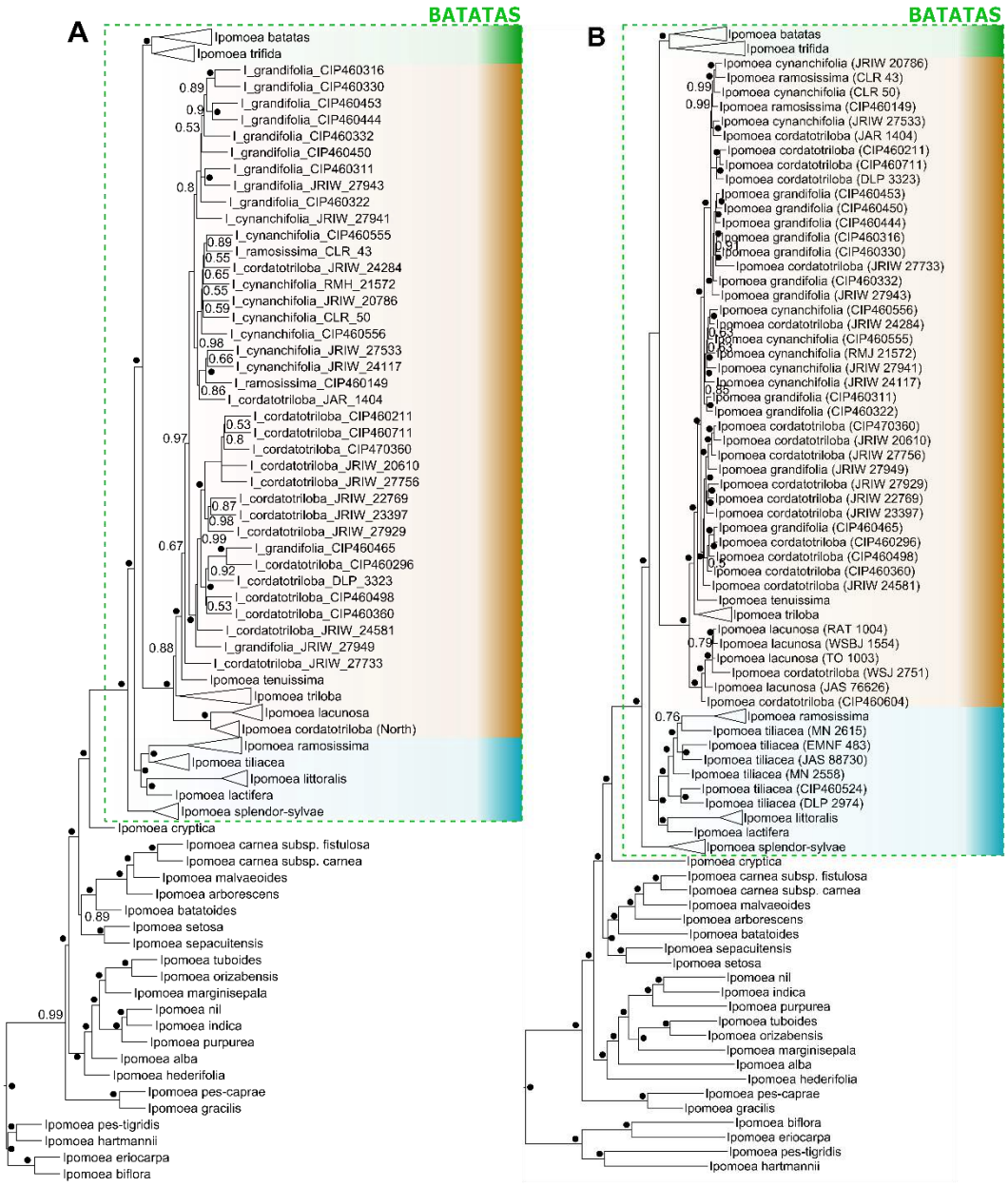


Figure S1. Nuclear phylogenies of sweet potato and its CWRs, Related to Figures 2 and 3.

Phylogeny of *Ipomoea* series *Batatas* (excluding hybrid species *I. leucantha* and *I. tabascana*) inferred from 307 nuclear regions that do not show recombination. Blue, perennial species; orange, annual species; green, sweet potato and *Ipomoea trifida*. Black dots indicate 100% support. Triangles represent monophyletic species with 100% support. (A) Inferred using Astral-II. Values at the nodes indicate bootstrap support for a partition (100 replicates from gene trees). (B) Inferred using Approximate Maximum Likelihood with all regions concatenated. Values at the nodes indicate local support values with the Shimodaira-Hasegawa test (1,000 resamples).

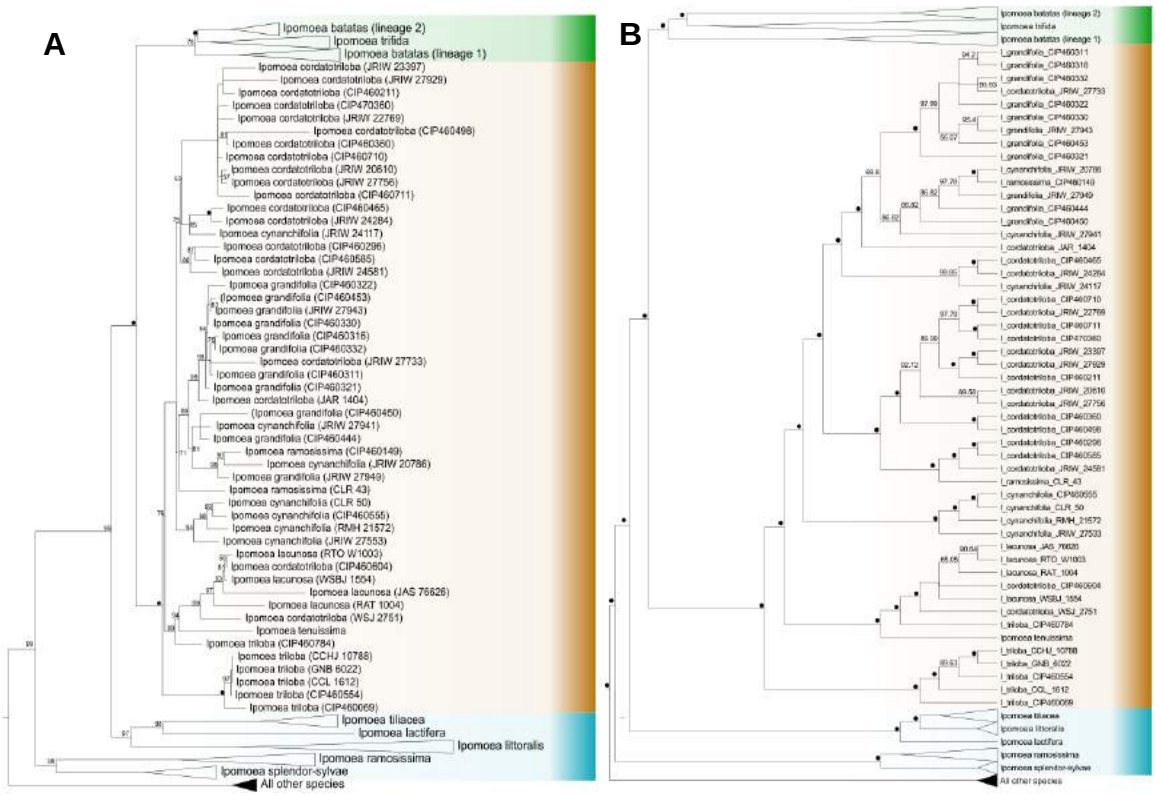


Figure S2. Chloroplast phylogenies of sweet potato and its CWRs, Related to Figures 2 and 3.

Phylogeny of *Ipomoea* series *Batatas* (excluding hybrid species *I. leucantha* and *I. tabascana*) inferred from whole chloroplast sequences. Blue, perennial species; orange, annual species; green, sweet potato and *I. trifida*. Triangles represent monophyletic species with 100% support.
(A) Maximum Likelihood (RAxML, 1,000 bootstrap replicates). Values at the nodes indicate bootstrap support. Black dots indicate 100% support.
(B) Maximum parsimony analysis using indels only, Majority Rule consensus. Values at the nodes indicate bootstrap support.

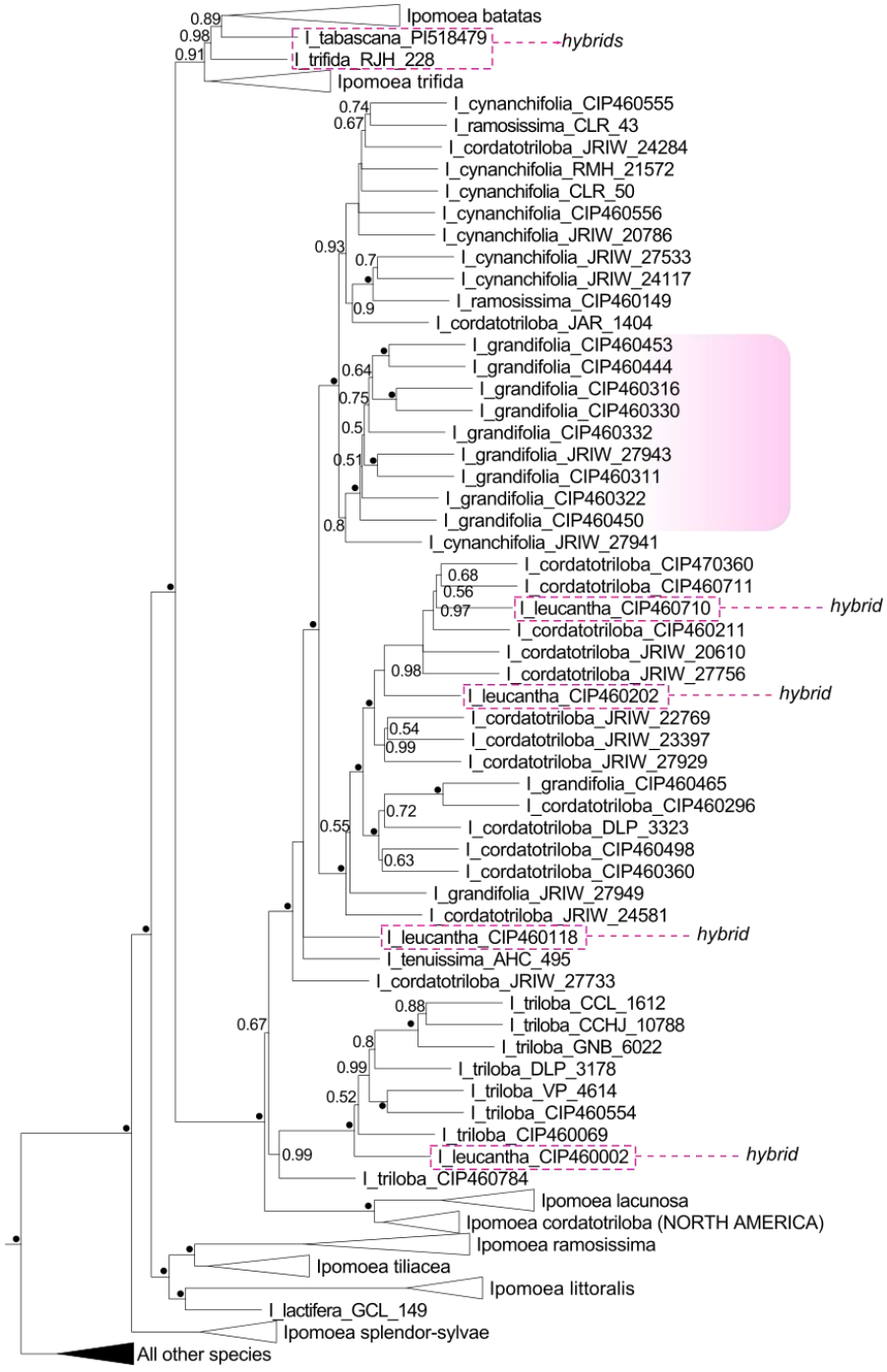


Figure S3. Nuclear phylogeny of sweet potato and its CWRs including hybrids, Related to Figure 2.

Approximate Maximum Likelihood phylogeny of *Ipomoea* series *Batatas* inferred from 307 nuclear regions that do not show recombination, showing the position of the hybrid species. Triangles represent monophyletic species with 100% support. Values at the nodes indicate bootstrap support (100 replicates from gene trees) for a partition. Black dots indicate 100% support. Purple dashes indicate putative hybrid specimens: *I. tabascana*, tetraploid *I. trifida* and *I. leucantha*, and purple shading area indicates most *I. grandifolia* specimens.

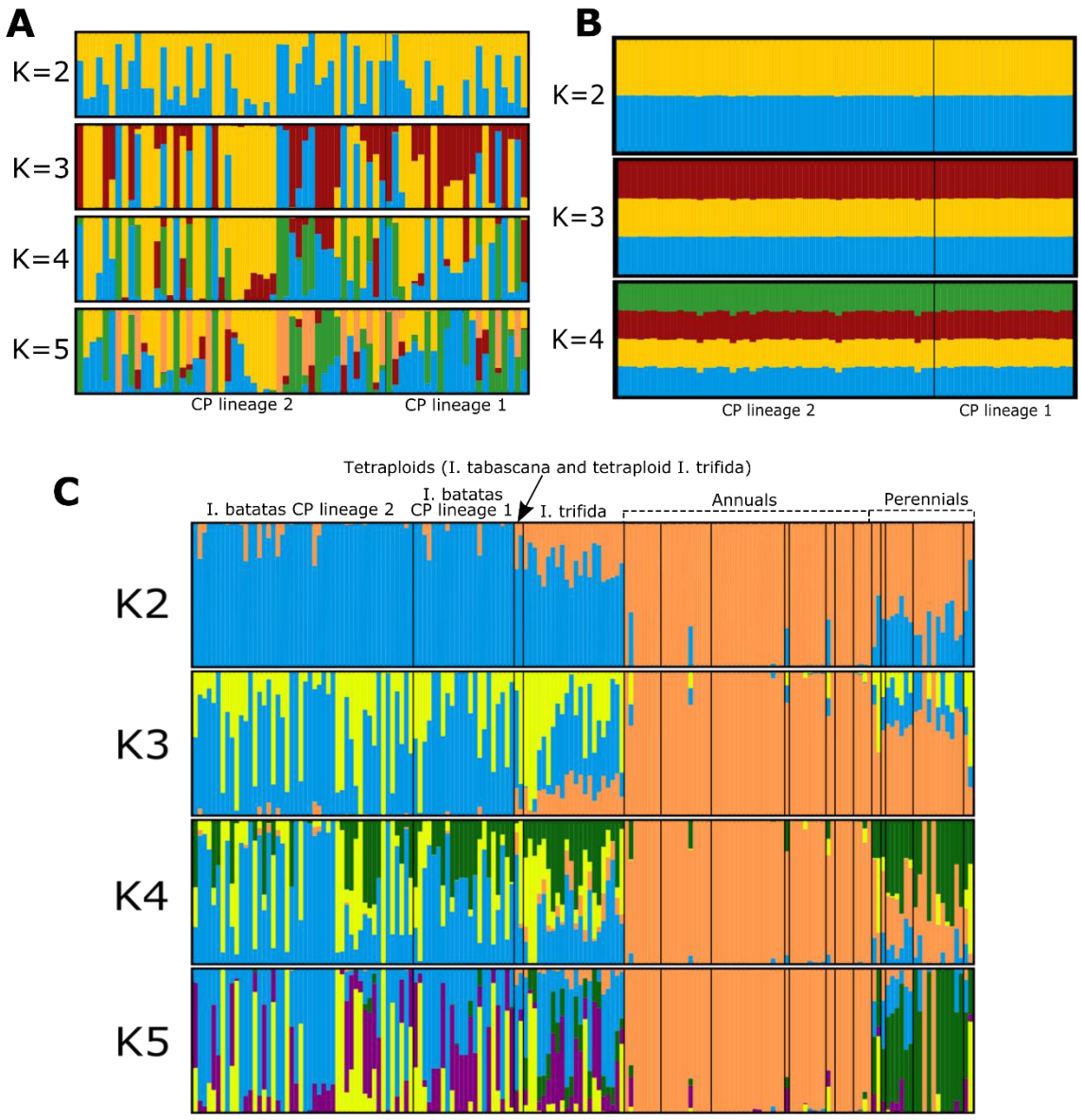


Figure S4. Additional population structure analyses, Related to Figures 2 and 3.

K is the number of assumed ancestral populations.
(A) Sweet potato only using nuclear coding regions.
(B) Sweet potato only using the nuclear ribosomal non-coding *ITS* DNA region.
(C) *Ipomoea* series *Batatas* using nuclear coding regions.

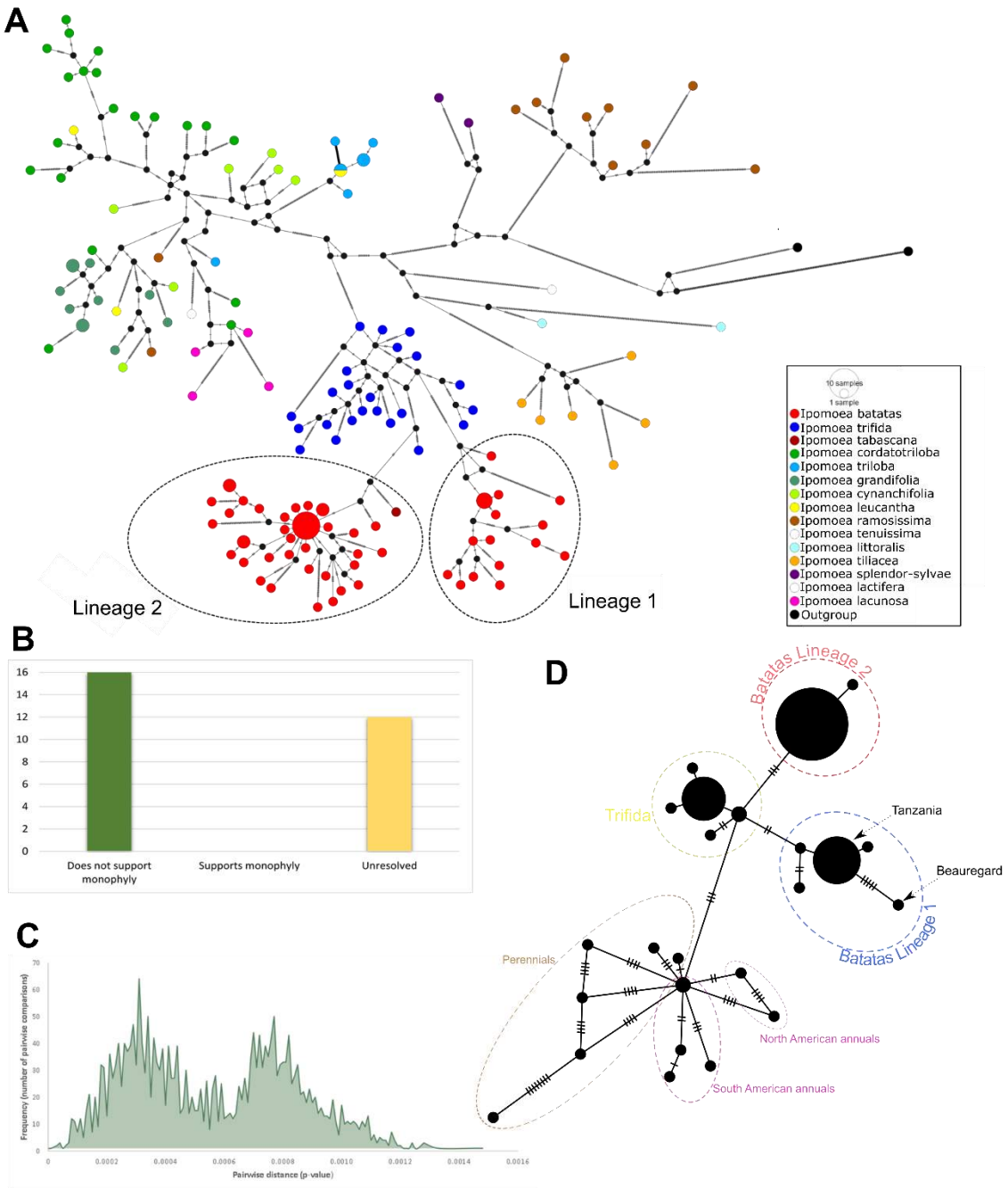


Figure S5. Additional analyses of chloroplast data, Related to Figure 3.

(A) Integer Neighbor-Joining chloroplast network of all species in *Ipomoea* series *Batatas*.
(B) None of the 28 most variable regions in the chloroplast genome supports monophyletic sweet potato.
(C) Pairwise distance between all sweet potato chloroplast genomes.
(D) Median-Joining *rpl32-trnL* network of all species in *Ipomoea* series *Batatas*, showing the position of the two commercial sweet potato varieties Beauregard and Tanzania.

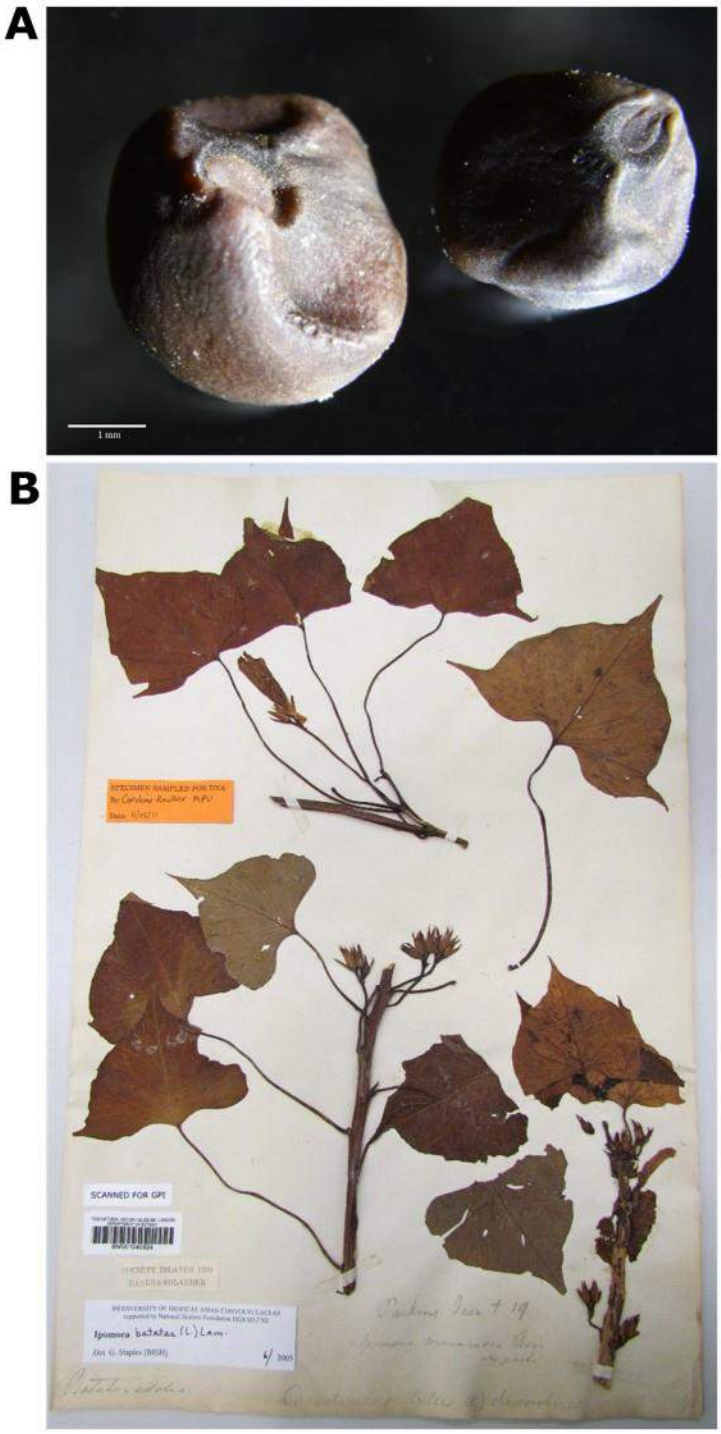


Figure S6. Pictures of *Ipomoea seeds* and a historic specimen, Related to Figure 7.

(A) Seeds of *Ipomoea littoralis* (left; *L.J. Brass 13940* [BM]) and *Ipomoea batatas* (right; *S9 55* [USDA]).
(B) *Ipomoea batatas* specimen collected in 1769 by Banks and Solander in the Society Islands.

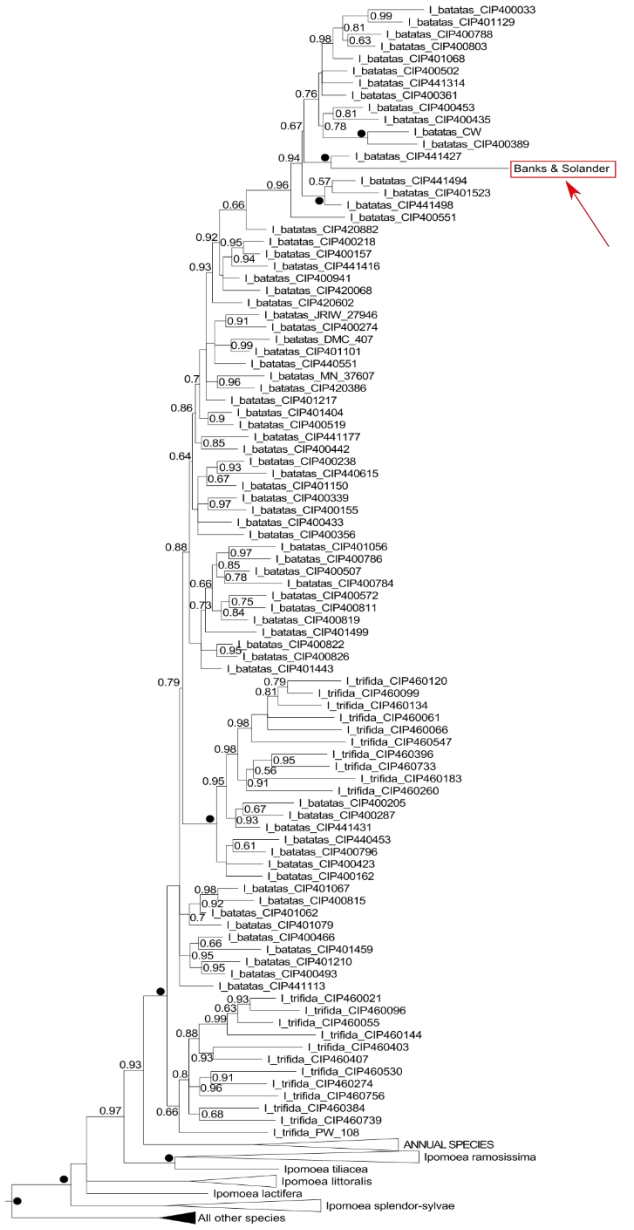


Figure S7. Nuclear phylogeny including Banks and Solander specimen, Related to Figure 7.

Position of the sweet potato specimen collected by Banks and Solander in a phylogenetic tree based on 12,905 nuclear variable positions, inferred using FastTree 2. Values at the nodes indicate local support values with the Shimodaira-Hasegawa test (1,000 resamples). Black dots indicate 100% support.

Table S1. Relative population sizes for *Ipomoea batatas* and *I. trifida* inferred from our plastome dataset, Related to Figure 5.

Population	Ne*	Minimum age (years)
TB_1B_2	0.07	~500,000
TB_2	0.06	~250,000
T	0.28	Extant population
B_1	0.27	Extant population
B_2	1	Extant population

T = *I. trifida* lineage; B_1 = *I. batatas* CL1; B_2 = *I. batatas* CL. TB_1B_2 = lineage ancestral to all three lineages. *TB2* = lineage ancestral to *T* and B_2. *Effective population sizes (Ne) are expressed as a proportion of that of $_{B2}$.