Identifying the closest wild relative of sweet potato (*Ipomoea batatas* (L.) Lam.)

Tom Wells

Doctoral Training Centre & Department of Biology

Worcester College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2023



"explanation does not follow from description;
it is description taken that much further."

Bruno Latour (1999) *Technology Is Society Made Durable*.
In Law, J., ed., Sociology of Monsters.

CONTENTS

THESIS ABSTRACT **5**

ACKNOWLEDGEMENTS **7**

CHAPTER 1. (THESIS INTRODUCTION) Tetraploid plants are crucial to resolving outstanding questions about the evolutionary origin and contemporary diversity of sweet potato (*Ipomoea batatas* (L.) Lam.) **9**

CHAPTER 2. A systematic review of tetraploid germplasm material shows taxonomic confusion has impeded sweet potato research & crop improvement **37**

CHAPTER 3. Distinguishing autotetraploids from modern hybrids identifies the progenitor of sweet potato **100**

CHAPTER 4. *Ipomoea aequatoriensis* T.Wells & P.Muñoz: a new autotetraploid species of *Ipomoea* from Ecuador & Colombia **157**

CHAPTER 5. (THESIS CONCLUSION) The identification and description of sweet potato's closest wild relative paves the way for new research on its evolution and diversity **206**

APPENDICES **220**

THESIS ABSTRACT

Thanks to its relative resilience, nutritional value, and high dry-matter content, sweet potato (*Ipomoea batatas* (L.) Lam.) is a vital staple food around the world, especially for small-scale farmers in drought-prone areas. The development and distribution of Vitamin A-rich, orange-fleshed varieties has proved a particularly valuable tool in combatting malnutrition and mineral deficiency, and made sweet potato an increasingly popular health food. Despite this agricultural and dietary importance however, sweet potato's origin and domestication are still relatively poorly understood. Several outstanding questions remain about how sweet potato evolved, including *whether sweet potato is an auto or allohexaploid*, *why it possesses two independent chloroplast lineages,* and *how it developed its thickened storage roots*. Answering these inter-related questions is integral to assembling sweet potato's complex hexaploid genome; understanding the development of key traits; identifying, classifying and utilising wild and cultivated diversity; and assessing how geneflow and introgression have shaped all three.

Central to answering these questions is the need for an improved understanding of the systematics and taxonomy of sweet potato's immediate crop wild relatives (CWRs), and in particular the identification of its closest wild ancestor, or "progenitor". Given current knowledge of the likely genomic structure of hexaploid *I. batatas*, this progenitor is thought to be autotetraploid – with two identical subgenomes – but no such taxon has yet been identified. The lack of previous progress in identifying this tetraploid progenitor of cultivated sweet potato is largely the result of extensive taxonomic confusion over nomenclature and species delimitation for sweet potato CWRs, which has been exacerbated by a complex evolutionary history for *I. batatas* involving polyploidy and potential hybridisation and introgression.

In this thesis, I provide a comprehensive review of all putatively wild tetraploid material resembling or previously associated with *I. batatas,* before placing them within the phylogenetic and taxonomic framework established by the recent Foundation Monograph of the genus *Ipomoea* L. in the Americas. I identify and describe sweet potato's closest CWR *I. aequatoriensis* T.Wells & P.Muñoz, providing a complete systematic account of its diversity and exploring its relationship to the cultivated crop. I have been able to achieve this by integrating material from germplasm and natural history (herbarium) collections – two data sources that are generally treated separately – and by the application of a combination of analytic approaches that include the latest in high-throughput sequencing technologies, bio-informatics, and rigorous morphological analysis. Doing so provides vital new insights into the origin of one of the world's most important food crops.

ACKNOWLEDGEMENTS

My thanks go to my supervisor Prof Robert Scotland, my colleagues John Wood, Dr Pablo Muñoz-Rodriguez, Dr Tom Carruthers, and Dr Alex Sumadijaya, as well as to my former supervisors Dr David Harris and Dr Alex Monro for all their combined guidance, support, suggestions and inspiration. I also thank my family and friends, and most importantly my partner, Ngaio, for her patience, advice and encouragement.

The work would not have been possible without the access to collections and materials provided by herbarium curators at University of Oxford Department of Plant Sciences, Royal Botanic Gardens Kew, Royal Botanic Garden Edinburgh, Natural History Museum London, Muséum national d'Histoire naturelle Paris, Pontificia Universidad Católica del Ecuador Quito, Museo Ecuatoriano de Ciencias Naturales del Instituto Nacional de Biodiversidad Quito, Instituto de Ecología Xalapa, Universidad Nacional Autónoma de México, and Naturhistoriska Riksmuseet Stockholm; germplasm managers at the USDA, CIP and NARO; librarians at the Bodleian, Worcester College, Pitt Rivers Museum, and Kyoto University; nor the laboratory help and support provided by Helen Prescott; the horticultural facilities provided by Simon Bagnall and Tim Wroe; fieldwork assistance provided by David Espinel-Ortiz, Dr Katya Romoleroux, Carlos Ceron, Xavier Cornejo, and the people of Agua Blanca Reserve; and the funding provided by the British Taxpayer through UKRI and BBSRC's Interdisciplinary Biosciences Doctoral Training Programme at Oxford.

Thank you all.

CHAPTER 1 (THESIS INTRODUCTION): Tetraploid plants are crucial to resolving outstanding questions about the evolutionary origin and contemporary diversity of sweet potato (*Ipomoea batatas* (L.) Lam.)

INTRODUCTION

Sweet potato, (*Ipomoea batatas* (L.) Lam.) is a member of the large pantropical genus *Ipomoea*, which contains an estimated 800 species globally (Wood et al. 2020). Within the genus, *I. batatas* has long been recognised as one of a smaller group of closely related species of slender climbers, predominantly from Central and South America, known as the 'Batatas group' (Austin 1988). The most comprehensive phylogenetic and taxonomic analyses of *Ipomoea* to date were recently carried out by the Scotland group at University of Oxford Department of Plant Sciences, confirming the Batatas group to be monophyletic and composed of 15 currently recognised species (Muñoz-Rodríguez et al. 2018, 2019; Wood et al. 2020) (see Appendix 1). These analyses have greatly improved our understanding of sweet potato's crop wild relatives (CWRs), providing a stable framework for further analysis. Most importantly, they confirm *I. trifida* (Kunth) G. Don, a weedy, diploid (2n=2x=30), circum-Caribbean species, as the closest known CWR of *I. batatas*, rejecting a role for any other currently known species in its origin. Nonetheless, outstanding questions remain about how cultivated sweet potato might have arisen from a diploid ancestor shared with *I. trifida*, which is necessarily complicated by sweet potato's hexaploid genome (2n=6x=90).

Whole-genome duplication (WGD) events result in an increase in the number of sets of chromosomes possessed by an organism. This condition, known as polyploidy, is recognised as an important and relatively common mode of diversification, especially in plants (Otto and Whitton 2000; Clark and Donoghue 2018). Rapid or even almost instantaneous changes in physiology, pathogen

resistance and environmental robustness that may arise from WGD make polyploids an important topic of research for evolutionary biologists, conservationists and crop breeders interested in understanding potential adaptation to predation, disease, and new environmental conditions such as those resulting from climate change (Soltis et al. 2015; Van De Peer et al. 2017). Many important crops, including wheat, potato, cotton, tobacco, peanut and banana are known to be polyploids, with recent research mapping the various evolutionary routes they have taken to polyploidy (e.g. Marcussen et al. 2014; Fang et al. 2017; Gordon et al. 2019).

Sweet potato is another important polyploid crop, with a hexaploid genome composed of three subgenomes, but its route to hexaploidy remains unclear (Yang et al. 2017). Along with most other species in the Batatas group, sweet potato's closest known relative *I. trifida* is diploid (2n=2x=30) and cannot therefore have given rise to hexaploid *I. batatas* directly, without either a triploid (2n=3x=45), or more likely a tetraploid (2n=4x=60) intermediary (Ramsey and Schemske 1998; Tayalé and Parisod 2013). For this reason, wild-collected tetraploid plants that resemble hexaploid *I. batatas* are potentially crucial to attempts to understand the origin, evolution and contemporary diversity of sweet potato, since they may represent the extant descendants of its immediate wild ancestor, or "progenitor". However, these tetraploids may alternatively be the product of more recent hybridization between hexaploid *I. batatas* and diploid *I. trifida* (Fig. 1) (Orjeda et al. 1991), but distinguishing between these two possibilities has proven extremely difficult (Oracion et al. 1990; Roullier et al. 2013). As a result, there has been considerable disagreement and confusion over the identity of known tetraploid plants resembling wild forms of hexaploid sweet potato (*I. batatas* (L.) Lam.), and their potential role in its origin. Definitively resolving the relationship of these different tetraploid accessions to hexaploid sweet potato is key to answering several important research questions outlined below.

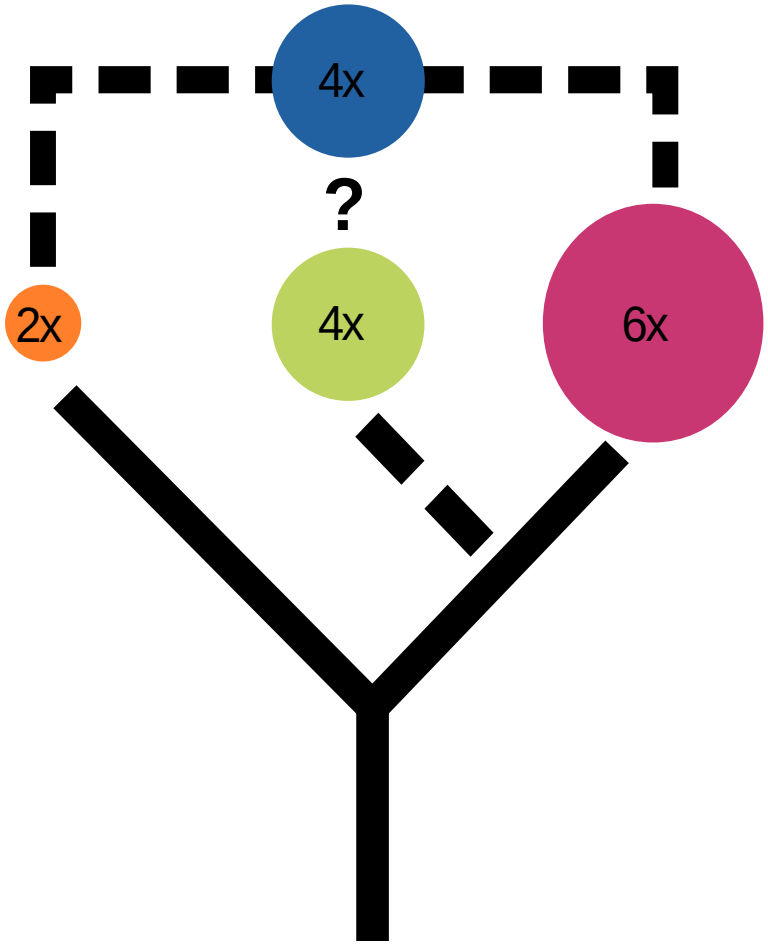


Figure 1: Illustrating the two possible evolutionary relationships of wild-collected tetraploid plants resembling *I. batatas* to hexaploid sweet potato. They could either be the tetraploid progenitor of hexaploid *I. batatas* (green circle) or recent hybrids between hexaploid *I. batatas* and diploid *I. trifida* (blue circle).

1. Is Sweet Potato nn Auto- or Allohexaploid?

The ancestor of *I. batatas* is believed to have diverged from its closest known relative *I. trifida* and begun the process of polyploidization at least 800,000 years ago (Yang et al. 2017; Muñoz-Rodríguez et al. 2018) but various possibilities exist for both the route to hexaploidy in *I. batatas,* and the taxa involved, with each likely to have differing implications for the composition of its resulting hexaploid genome (Fig. 2). As a result, several competing theories have been suggested for how hexaploid *I. batatas* might have arisen from its diploid ancestor (e.g. Nishiyama 1971; Kobayashi 1984; Roullier et al. 2013; Yang et al. 2017; Muñoz-Rodríguez et al. 2018), but none has been definitively confirmed or accepted as correct. These

theories have traditionally been distinguished by whether they suggest sweet potato is an autopolyploid (arising by multiple rounds of WGD from a single ancestral species) or an allopolyploid (originating through hybridization between two distinct species).

In either case, polyploidy most frequently occurs in plants through the production of unreduced gametes (Ramsey and Schemske 1998; Tayalé and Parisod 2013), which has been shown to occur around 20% of diploid *I. trifida* individuals (Orjeda et al. 1990). The union of a reduced and an unreduced gamete from diploid parents will produce triploid offspring (n + 2n = 3x), and the union of two unreduced gametes a tetraploid one (2n +2n = 4x) (Fig. 2). Triploid plants – which are likely to be sterile – can in theory undergo somatic doubling to directly produce an autohexaploid with three identical copies of the original diploid genome (Fig. 2A) but will more commonly produce fertile autotetraploids in a process known as the "triploid bridge" (Ramsey and Schemske, 1998) (Fig. 2). Once a population of autotetraploids is established, they are likely to diverge from the diploid progenitor, from which they should have a degree of reproductive isolation (Ramsey and Schemske 1998). There are then two possible routes to hexaploidy from this autotetraploid (Fig. 2B-C), again involving non-reduction of gametes (Ramsey and Schemske 1998), examples of which have been observed in autotetraploid populations of other crops such as Sugar Beet, *Beta vulgaris* L. (Hornsey 1973), and Alfalfa, *Medicago sativa* L. (Bingham 1968). For instance, the union of a reduced and an unreduced tetraploid gamete will result in a hexaploid (2n + 4n = 6x) (Fig. 2B). This new autohexaploid will have a genome composed of three identical subgenomes, initially identical to those present in the tetraploid progenitor, which themselves originally arose from two identical copies of its diploid progenitor's genome. All three (diploid, tetraploid and hexaploid) should show a degree of reproductive isolation from each other and are likely to diverge over time.

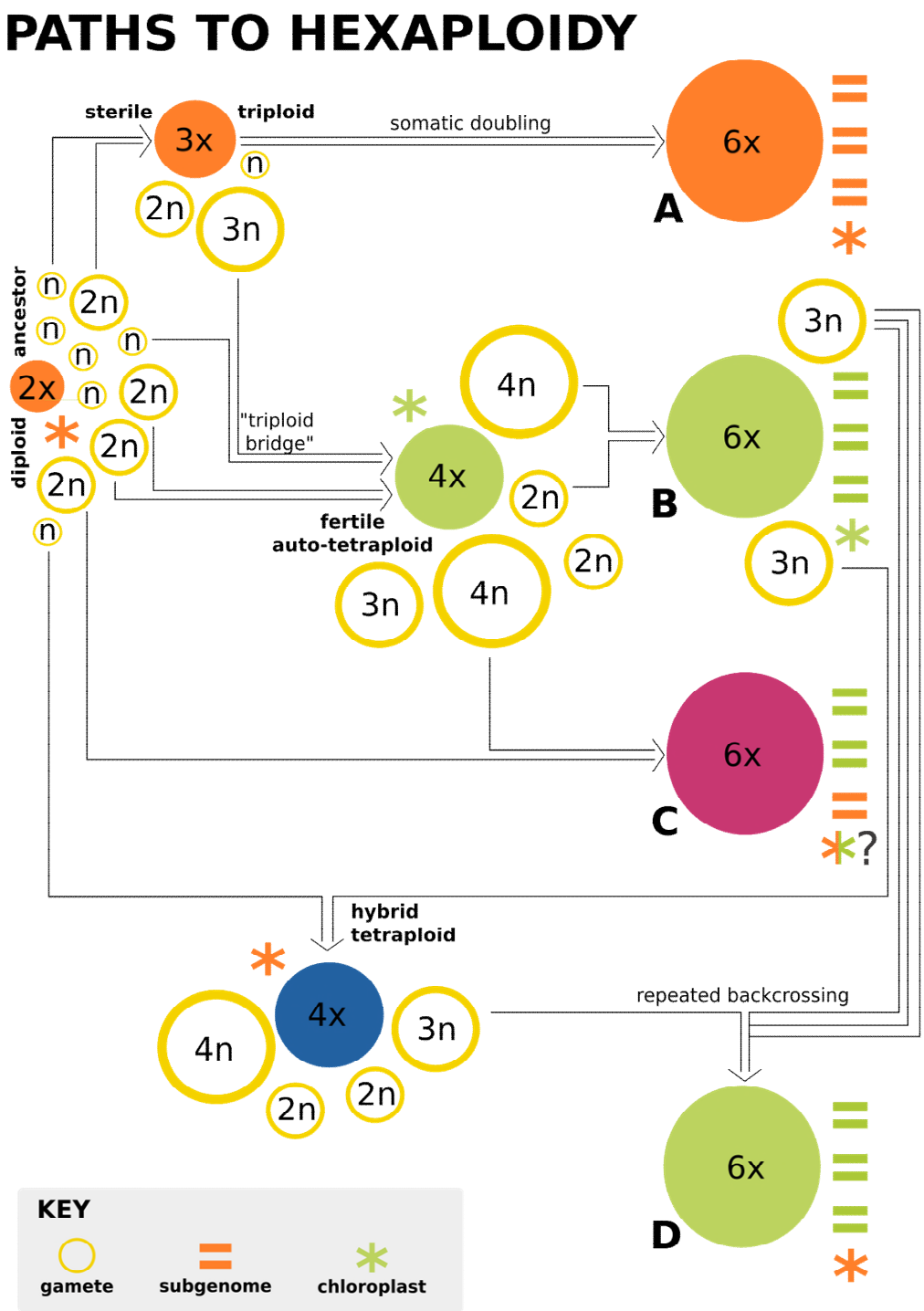


Figure 2: Diagram of possible routes to hexaploidy from a diploid ancestor. Solid circles represent sporophyte generation, hollow circles represent gametes, with size of both proportionate to ploidy. Coloured dashes and asterisks represent origin of subgenomes and chloroplast respectively. **A:** direct autohexaploid with three subgenomes and a chloroplast identical to those of diploid ancestor, derived by somatic doubling of a sterile triploid arising from union of reduced and unreduced gametes. **B:** indirect autohexaploid derived by union of unreduced and reduced gamete of an autotetraploid ancestor, itself arising from diploid ancestor, and with three subgenomes and a chloroplast identical to those of the tetraploid. **C:** allohexaploid derived by union of unreduced gametes from diploid and tetraploid ancestors, with one subgenome inherited from the former and two from the latter, and the chloroplast from either, depending on parental inheritance pattern. **D:** hypothesized route to new chloroplast lineage through initial formation of a tetraploid hybrid between hexaploid and diploid, and subsequent backcrossing to remove evidence from nuclear subgenomes.

This scenario would appear to fit the findings of Muñoz-Rodríguez et al. (2018), who suggested based on phylogenetic analysis of sweet potato and its known CWRs, that *I. batatas* is an autohexaploid, which arose solely from its closest relative, diploid *I. trifida*. By comparing allelic variation in 605 low-copy nuclear loci for multiple sweet potato accessions, they found that in the majority of cases, all six putative alleles were more closely related to each other than to those from any of the other species in the Batatas group, ruling out the involvement of any of these species other than *I. trifida* in the origin of sweet potato's hexaploid genome (Muñoz-Rodríguez et al. 2018).

By contrast however, in the most comprehensive attempt to assemble the genome of hexaploid *I. batatas* to date, Yang et al. (2017) showed that successfully phased haplotypes grouped into clades of two or four in phylogenetic analyses, with the latter in turn forming a further two pairs. They concluded that this pattern suggests hexaploid *I. batatas* possesses a pair of identical subgenomes, and one other that differs slightly (AABBBB), as in the configuration illustrated in Fig. 2C. This finding in turn appeared to support earlier hypotheses that *I. batatas* is an allohexaploid that arose through hybridization between a diploid (2n=2x=30) and a tetraploid (2n=4x=60) progenitor. This theory was initially advanced by Japanese researchers, based on synthetic crossing experiments (Nishiyama et al. 1975; Shiotani and Kawase 1987), with further support coming from observations of chromosome interactions during meiosis (Jones, 1965; Magoon et al., 1970; Ting et al., 1957; Ting & Kehr, 1953), inconsistent numbers of copies of the 18S region (Srisuwan et al. 2006), and Wx-Intron variations (Gao et al. 2011). An allohexaploid origin such as this could either have occurred through the union of unreduced gametes from both the autotetraploid and its diploid progenitor (4n + 2n = 6x), or of a reduced gamete of each followed by somatic doubling (2n + n = 3x => 6x) (Fig. 2C). If there had been sufficient divergence between the diploid and tetraploid

genomes prior to this hybridization event, it should in theory leave a discernible footprint in the composition of the resulting allohexaploid genome, corresponding to that described above, with one subgenome (AA) inherited from the diploid, and two (BBBB) from the tetraploid.

Evidence for these divergent subgenomes is based on observations of a relatively limited number of specimens of *I. batatas* however, and so it is not known whether all accessions of *I. batatas* have the same genomic configuration. Evidence has also repeatedly pointed to apparent inconsistencies in the genome composition of different sweet potato accessions (Srisuwan et al. 2006; Roullier et al. 2013; Kyndt et al. 2015; Quispe-Huamanquispe et al. 2019) (see following chapter). These may be the result of ongoing diploidization (Srisuwan et al. 2006; Soltis et al. 2015) but could also indicate multiple, separate rounds of polyploidization by different routes. What is more, chromosome pairing is not always a reliable means of assessing genome composition, particularly in groups of closely related taxa (Bomblies 2022); and the accuracy and completeness of the haplotype-resolved draft genome for hexaploid *I. batatas* has been the subject of some criticism (Wu et al. 2018).

Interpreting these apparently contradictory results is also complicated by the fact that terms such as auto- and allopolyploid are frequently insufficient to describe the complex processes leading to WGD in plants, thanks to the often close relationship between the taxa involved (Ramsey and Schemske 1998). For example, Kobayashi (1984) agreed with Nishiyama et al. (1975) and later authors that *I. batatas* likely arose from a diploid and a tetraploid progenitor, but he also classified both as *I. trifida*, which would make *I. batatas* an autopolyploid and conform with the hypothesis put forward by Muñoz-Rodríguez et al. (2018).

More importantly, without the identification of the as-yet unidentified tetraploid progenitor of hexaploid sweet potato, it has not been possible to phase and assemble the different subgenomes of hexaploid *I. batatas*, or properly test these competing hypotheses about their origin. The lack of this tetraploid progenitor from the sampling of Muñoz-Rodríguez et al. (2018) for example, could explain the apparent evidence for an autopolyploid origin for *I. batatas*. Until this tetraploid progenitor is identified, its genome cannot be assembled and compared to that of *I. trifida* or *I. batatas*, and progress cannot therefore be made on resolving how exactly hexaploid sweet potato evolved from a diploid ancestor.

2. How & When Did Sweet Potato Develop Its Enlarged Storage Roots?

A particular distinction that is often muddled or evaded in discussions of the origin of hexaploid *I. batatas* is whether the transition to hexaploidy took place as part of the domestication process, or prior to it. Polyploidy has certainly played a role in the domestication of some crops (e.g. Wendel and Clark Cronn 2003; Emshwiller et al. 2009), but it is also increasingly recognised as an important mechanism in evolution more generally, and does not appear to be any more common in domesticated plants than it is in wild ones (Soltis et al. 2010; Zielinski and Scheid 2012). In the case of sweet potato, the potential role of polyploidy in the evolution and development of thickened storage roots has not been investigated but WGD is known to result in changes in physiology and the enlargement of particular organs (Soltis et al. 2015; Van De Peer et al. 2017).

Given its apparently long history of cultivation (Ugent and Peterson 1988), and the extensive variation in colour, flavour and starch-content observed across the many extant cultivars (Loebenstein and Thottappilly 2009), human-mediated selection is almost certain to have had a dramatic impact on the storage roots of *I. batatas*. Nevertheless, recent work as part of the monograph of the genus *Ipomoea*

in the Americas revealed over 60 other undomesticated species with enlarged storage roots similar to those of *I. batatas*, some of which are eaten, though none are widely cultivated (Muñoz-Rodríguez et al. 2019). Given that This suggests that *I. batatas* could already have possessed a storage root of some kind prior to its cultivation by humans, perhaps as an adaptation to drought or fire.

Although there are some accounts of apparently thickened pencil-like roots, and even small storage roots in a few individuals of diploid *I. trifida* (Nishiyama et al. 1975; Srisuwan et al. 2018; Li et al. 2019), none of the other 15 currently known species that belong to the Batatas group possess root structures comparable to those found in cultivated sweet potato. This would suggest that the dramatically thickened storage roots we see today evolved sometime after *I. batatas* diverged from an ancestor shared with *I. trifida* and began the process of polyploidization. An increase in ploidy is known to have an impact on gene expression, the formation and size of different plant organs, and ecological tolerance (Parisod et al. 2010), and its potential role in the development of thickened storage roots in *I. batatas* therefore warrants investigation. The confusion about the origin of hexaploidy in *I. batatas* complicates this research.

For example, Muñoz-Rodríguez et al. (2018) dated the divergence of *I. batatas* and diploid *I. trifida* to 800,000ya, claiming that this suggests sweet potato must have evolved its thickened storage root prior to human-mediated selection. Since their phylogeny did not include the tetraploid progenitor of sweet potato however, this date does not in fact represent the origin of hexaploid *I. batatas*, but the earliest possible beginning of its route to hexaploidy. The ability to form thickened storage roots could have evolved at any point after this date, for example following the transition to tetraploidy or hexaploidy, but both these events currently remain undated.

Until the tetraploid progenitor of hexaploid sweet potato has been definitively identified, included in dated phylogenetic analyses, and its own roots characterized, the question of whether sweet potato's edible storage roots evolved naturally through an increase in ploidy, or purely through the process of human domestication cannot therefore be satisfactorily addressed. The result is continuing uncertainty about the origin of the most agriculturally and economically important trait in a major global food crop.

3. How Did Sweet Potato Evolve Two Separate Chloroplast Lineages?

Understanding the evolutionary origin of sweet potato is further complicated by the existence of two distinct chloroplast lineages within hexaploid *I. batatas*, one of which appears to be more closely related to its closest diploid relative *I. trifida* in phylogenetic analysis of chloroplast sequence data (Roullier et al. 2011, 2013; Muñoz-Rodríguez et al. 2018). Muñoz-Rodríguez et al. (2018) attributed this structure to chloroplast capture resulting from probable hybridization and subsequent backcrossing between *I. batatas* and *I. trifida* (Fig. 2:D). This conclusion was in direct contrast to Roullier et al.'s (2013) interpretation of the same phylogenetic structure, in which they had instead suggested that the two lineages may indicate multiple origins for hexaploid *I. batatas*, one in Central America and the other in South America. Roullier et al.'s (2013) studies were limited by a lack of comprehensive nuclear data however, whereas Muñoz-Rodríguez et al. (2018) used their analysis of 605 low-copy nuclear loci to show that there was no evidence for a similar split in the nuclear genome of the *I. batatas* samples analysed, making the dual-origin hypothesis less likely.

If one of sweet potato's two chloroplast lineages did arise via chloroplast capture as suggested by Muñoz-Rodríguez et al. (2018), then tetraploids are again key to understanding this process. This is because hexaploid plants will normally

produce 3n gametes, such that hybridization between hexaploid *I. batatas* and its closet diploid relative *I. trifida* (which normally produces 1n gametes) will most likely produce tetraploid offspring (3n + 1n = 4x) (Fig.1D), which could then provide the crucial link for the required introgression to take place between the two (Orjeda et al. 1991). Since the second chloroplast lineage in hexaploid *I. batatas* is more closely related to diploid *I. trifida* than it is to the other *I. batatas* lineage, this implies that the initial hybrid must have inherited a chloroplast from *I. trifida* or a recent ancestor of it (Muñoz-Rodríguez et al. 2018). For chloroplast capture to take place, the F1 tetraploid hybrid would then have to have repeatedly backcrossed with hexaploid *I. batatas*, such that the new chloroplast was retained, and all evidence of hybridization in the nuclear genome was erased (Folk et al. 2017).

Recently diverged sister taxa such as *I. trifida* and *I. batatas* are known to be susceptible to hybridisation, gene-flow, and introgression in areas of secondary contact (Rieseberg et al. 2007). This is the case for many crops (Ellstrand et al. 1999, 2010; Janzen et al. 2019), and has been demonstrated for two other species in the Batatas group – *I. cordatotriloba* Dennst. and *I. lacunosa* L. (Abel and Austin 1981; Rifkin et al. 2018). Introgression is also known to have played a prominent role in the evolution of a number of other plant species (Soltis 1991; Rieseberg and Wendell 1993; Abbott et al. 2013; Mallet et al. 2016) and domesticated crops (Palmer et al. 1985; van Heerwaarden et al. 2011; Hufford et al. 2013; Marcussen et al. 2014; Fang et al. 2017; Flowers et al. 2019), and forms a useful tool for the incorporation of beneficial traits from wild relatives in contemporary crop improvement breeding programmes (Frisch and Melchinger 2005).

While chloroplast capture is thought to occur relatively frequently in plants however (Soltis 1991; Folk et al. 2017), the difference in ploidy between *I. batatas* and *I. trifida* makes matters more complicated in this particular case. Crosses between parents of different ploidy are most often successful when the maternal

parent is of a higher ploidy, which in this case would suggest *I. batatas* is the most likely maternal parent of any F1 hybrids with diploid *I. trifida*. This is supported by evidence from experimental crosses between *I. batatas* and lower ploidy CWRs, where hybridization was only successful with hexaploid *I. batatas* as the maternal parent (e.g. Kobayashi and Miyazaki 1975). Since the chloroplast is generally accepted to be maternally-inherited in most plants (Greiner et al. 2015), this would appear to complicate Muñoz-Rodríguez et al.'s (2018) chloroplast capture theory, which requires a diploid ancestor of *I. trifida* to be the maternal parent in order to pass on the correct chloroplast to Lineage 2. Nonetheless, maternal inheritance of the chloroplast is far from universal, with biparental inheritance thought to occur in around 20% of flowering plants (Corriveau and Coleman 1988; Greiner et al. 2015), and likely paternal inheritance recorded in the sweet potato CWR *I. nil* (L.) Roth (Corriveau and Coleman 1988) and many hybrids (Hansen et al. 2007).

In addition to experimental evidence for the possibility of hybridization between hexaploid *I. batatas* and diploid *I. trifida* (Oracion et al. 1990; Orjeda et al. 1991), an apparently wild tetraploid hybrid between the two (Srisuwan et al. 2006) was discovered in the Mexican state of Tabasco and named *I. tabascana* J.A. McDonald & D.F. Austin (McDonald and Austin 1990; Austin et al. 1991). Later phylogenetic analysis of Sanger barcode marker *rpl32-trnL(UAG)* showed that *I. tabascana* possesses an *I. batatas* Lineage 2 chloroplast haplotype (Roullier et al. 2013), suggesting that it had inherited its chloroplast from its likely maternal parent. Identifying and characterizing other potential tetraploid hybrids like *I. tabascana* would allow us to confirm whether hybridization and gene-flow between sweet potato and its closest diploid relative is a widespread occurrence, what the predominant mode of chloroplast inheritance is, and what impact it might have had on the contemporary diversity of both the crop and the wild species.

It also remains possible however, that the two chloroplast lineages may have arisen as the result of two independent origins for hexaploid *I. batatas*, if one lineage arose as an autopolyploid from its tetraploid progenitor (Fig.2B), and the other by allohexaploidy, via a cross between the same tetraploid progenitor and its diploid ancestor, with the chloroplast donated by the diploid (Fig.2:C); though the same caveats about chloroplast inheritance in interploidy crosses apply here. The apparent lack of support for this scenario in the analyses of Muñoz-Rodríguez et al. (2018) could potentially be explained by both the absence of the tetraploid progenitor from their sampling, and a methodological artefact in their sequencing and bio-informatic pipeline, whereby the uneven composition of the hexaploid nuclear genome leads to the two subgenomes inherited from the tetraploid progenitor obscuring any phylogenetic signal from the subgenome inherited from the diploid (see chapter 3 for discussion of the challenges of consensus sequence assembly in allopolyploids).

Under these circumstances, evidence of any diploid contribution to the hexaploid genome would be missing from any resulting phylogeny based on the reconstruction of consensus sequences, giving the appearance of a single, autohexaploid origin for a monophyletic *I. batatas*. Resolving this issue is once again contingent on the definitive identification of the tetraploid progenitor of *I. batatas*, the assembly of its genome and its inclusion in phylogenetic and genomic analyses.

Definitively resolving the relationship of wild-collected tetraploid plants to cultivated hexaploid *I. batatas* is therefore integral to understanding the origin of sweet potato's two chloroplast lineages, both in terms of better understanding the prevalence and outcomes of hybridization with *I. trifida* and verifying the topology of its nuclear phylogeny.

Existing Tetraploid Germplasm Collections Resembling I. batatas

Within the Batatas group, both the robust coastal specialist *I. tiliacea* (Willd.) Choisy, and some individuals of the weedy North American species *I. cordatotriloba* have been recorded as tetraploid (2n=4x=60) (Ozias-Akins and Jarret 1994), but phylogenetic analysis has shown neither of these species to be sufficiently closely related to *I. batatas* to be plausible candidates for its tetraploid progenitor (Muñoz-Rodríguez et al., 2018). Meanwhile, since the 1950s, a number of other apparently wild tetraploid plants resembling *I. batatas* have been collected across Latin America, and deposited in global germplasm repositories (Nishiyama 1959; Martin et al. 1974; Kobayashi 1984; Austin et al. 1993; Contreras et al. 1994; Ozias-Akins and Jarret 1994) (See following chapter for comprehensive discussion of this material).

While most of these tetraploid collections have at one time or another been suggested as potential candidates for the progenitor of hexaploid sweet potato, their taxonomic identity and phylogenetic relationship to *I. batatas* have caused much dispute and confusion but have never been properly determined. When Roullier et al. (2013) analysed multiple examples of such tetraploid material, they found the majority of them grouped with *I. batatas* in their phylogenetic and network analyses based on nuclear and chloroplast barcode markers. These markers were not able to provide sufficient resolution to definitively assess whether they might be the missing tetraploid progenitor however, rather than feral or hybrid material derived from crosses between cultivated *I. batatas* and diploid *I. trifida* (Roullier et al. 2013).

Multiple herbarium specimens of apparently wild plants resembling *I. batatas* have also been collected across Latin America (Fig. 3), but their ploidy is not recorded, and chromosome counts cannot be undertaken for dried material. For example, Wood et al. (2020) recorded “apparently wild, fertile plants from various countries in tropical America including Colombia, Ecuador, Mexico, Panama and

Venezuela" resembling *I. batatas*, which they stated may either be escapes from cultivation or truly wild populations (Fig. 3).

These specimens have not been analysed alongside tetraploid germplasm collections, and are generally identified as either *I. batatas* or *I. trifida*, as well as a series of names synonymized in the recent foundation monograph of the genus *Ipomoea* in the Americas (Wood et al. 2020). Collections of wild plants resembling *I. batatas* are also recorded by Contreras et al. (1994) in southern Mexico, which were described as "feral" *I. batatas.* Some of these specimens were later identified as tetraploids, but the ploidy of others remains unknown (Fig. 3).

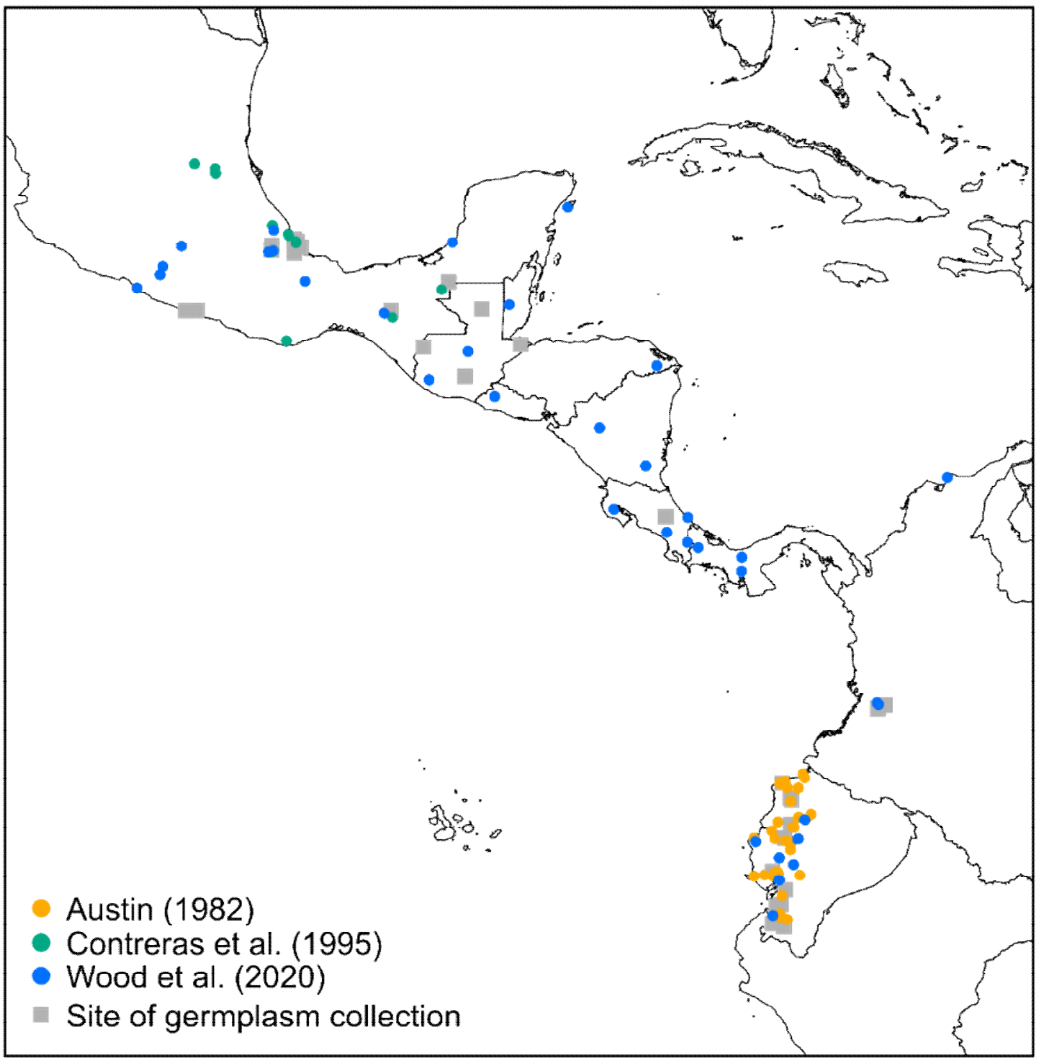


Figure 3: Map of collection sites of herbarium specimens resembling or identified as *I. batatas,* found growing wild but of unknown ploidy. Coloured dots indicate collection sites of herbarium material highlighted by different authors. Grey squares indicate collection sites of confirmed tetraploid germplasm material.

Meanwhile, as part of projects to document national and regional Flora in Ecuador, herbarium collections of unknown ploidy were made in some of the same areas as later tetraploid germplasm collections (Dodson and Gentry 1978; Austin 1982; Dodson et al. 1985) (Fig. 3) (see Chapter 4 for further discussion of some of this material).

Together these studies show that in addition to germplasm collections of known ploidy, plants resembling *I. batatas* but of unknown ploidy have been found growing outside cultivation across Latin America, between central Mexico in the north and southern Ecuador in the south. Understanding the relationship of these collections to cultivated sweet potato and tetraploid germplasm material offers an additional source of evidence for efforts to resolve questions about the crop's evolution and contemporary diversity.

CONCLUSION:

Recent developments in high-throughput sequencing and genomics have allowed researchers to overcome obstacles in reconstructing the complicated evolutionary histories of a number of polyploid crops, including bread wheat (Marcussen et al. 2014), strawberries (Kamneva et al. 2017), cotton (Fang et al. 2017), and maize (Hufford et al. 2013). The completion of draft genomes for *I. batatas* (Yang et al. 2017) and four of its CWRs – *I. trifida* (Hirakawa et al. 2015; Wu et al. 2018; Li et al. 2019), *I. triloba* L. (Wu et al. 2018), *I. lacunosa* (Rifkin et al. 2018) and *I. nil* (Hoshino et al. 2016) – provides the potential opportunity to use similar methods to resolve the route to hexaploidy, development of storage roots, and origin of separate chloroplast lineages in *I. batatas*.

But utilising the latest bioinformatics techniques to understand the evolutionary history of hexaploid *I. batatas* will first require the accurate identification of its tetraploid progenitor. While knowledge of sweet potato's CWRs has been

significantly improved by recent taxonomic and phylogenetic research, its immediate tetraploid progenitor has remained unidentified. Accurate knowledge about this tetraploid progenitor and its extant descendants will be essential to answering questions about the evolution of sweet potato, but difficulties in distinguishing them from more recent tetraploid hybrids with diploid *I. trifida*, are a major obstacle to progress. Overcoming this issue will require the integrative analysis of all available tetraploid material to resolve their likely relationships to hexaploid *I. batatas*.

To date, there has been no comprehensive study that places all previously collected germplasm material of apparently wild plants resembling *I. batatas* in the context of a well-sampled and resolved phylogeny. As such, it remains unclear what this material represents, and how it relates to cultivated hexaploid sweet potato in terms of its origin and its contemporary diversity. What is more, the existence of similar specimens in herbarium collections has never been properly investigated, resulting in the neglect of a large source of potential data, obscuring the diversity present in these collections, and most likely causing confusion in species delimitation and identification. In this context, distinguishing between true CWRs and merely feral plants remains a matter of subjective decision, while research into the origin of sweet potato is reliant on a limited number of germplasm collections of questionable provenance, equivalence, and identity.

In the course of this thesis, I set out to overcome these challenges. And in doing so, to create a resource for researchers wishing to better answer the fundamental questions about sweet potato outlined here, or to produce more productive and resilient varieties through the incorporation of existing wild diversity. This introduction forms the **first chapter**, outlining the outstanding questions in sweet potato research, and the centrality of tetraploid CWR material to answering them. The **second chapter** provides a systematic review of the various polyploid germplasm collections previously identified as wild forms of sweet potato and

suggested to potentially be involved in its origin, summarising previous research undertaken and what is and is not known about these collections. The **third chapter** builds on that review and the recent Foundation Monograph of the genus *Ipomoea* in the Americas (Wood et al. 2020) to distinguish wild autotetraploid plants that are the closest CWR of hexaploid *I. batatas*, from likely modern hybrids between *I. batatas* and *I. trifida*. The **fourth chapter** provides a review of the decision to describe this material as a new species from Ecuador and Colombia (*I. aequatoriensis* T.Wells & P.Muñoz), which is the closest known relative of sweet potato, and most likely directly descended from its tetraploid progenitor. The **fifth and final chapter** forms the conclusion, with a summary of my work and its implications.

REFERENCES:

Abbott R., Albach D., Ansell S., Arntzen J.W., Baird S.J.E., Bierne N., Boughman J., Brelsford A., Buerkle C.A., Buggs R., Butlin R.K., Dieckmann U., Eroukhmanoff F., Grill A., Cahan S.H., Hermansen J.S., Hewitt G., Hudson A.G., Jiggins C., Jones J., Keller B., Marczewski T., Mallet J., Martinez-Rodriguez P., Möst M., Mullen S., Nichols R., Nolte A.W., Parisod C., Pfennig K., Rice A.M., Ritchie M.G., Seifert B., Smadja C.M., Stelkens R., Szymura J.M., Väinölä R., Wolf J.B.W., Zinner D. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Abel W.E., Austin D.F. 1981. Introgressive Hybridization between Ipomoea trichocarpa and Ipomoea lacunosa. Bull. Torrey Bot. Club. 108:231–239.

Austin D.F. 1982. Convolulaceae. In: Harling G., Sparre B., editors. Flora of Ecuador. Stockholm: Swedish Research Councils. p. 30–69.

Austin D.F. 1988. The taxonomy, evolution and genetic diversity of sweetpotatoes and related wild species. Exploration, maintenance, and utilization of sweet potato genetic resources. Lima, Peru: CIP. p. 27–59.

Austin D.F., Jarret R.L., Tapia C., de la Puente F. 1993. Collecting tetraploid *I. batatas* (L.) Lam. in Ecuador. FAO/IBPGR Plant Genet. Resour. Newsl. 91/92:33–35.

Austin D.F., de la Puente F., Contreras J. 1991. Ipomoea tabascana , an Endangered Tropical Species. Econ. Bot. 45:435.

Bingham E.T. 1968. Aneuploids in seedling populations of tetraploid alfalfa, *Medicago sativa* L. Crop Sci. 8:191–201.

Bomblies K. 2022. Learning to tango with four (or more): the molecular basis of adaptation to polyploid meiosis. Plant Reprod. 36:107–124.

Clark J.W., Donoghue P.C.J. 2018. Whole-Genome Duplication and Plant Macroevolution. Trends Plant Sci. 23:933–945.

Contreras J., de la Puente F., Austin D.F., Diaz J. 1994. Biodiversity of sweetpotato (I*pomoea batatas*, Convolvulaceae) in Southern Mexico. Econ. Bot. 49:286–

296.

Corriveau J.L., Coleman A.W. 1988. Rapid Screening Method to Detect Potential Biparental Inheritance of Plastid DNA and Results for Over 200 Angiosperm Species. 75:1443–1458.

Dodson C.H., Gentry A.H. 1978. FLORA OF THE RIO PALENQUE SCIENCE CENTER: Los Rios Province ECUADOR. Selbyana. 4:301–302.

Dodson C.H., Gentry A.H., Valverde F.M. 1985. La Flora de Jauneche, Los Rios, Ecuador. Quito: Banco Central del Ecuador.

Ellstrand N.C., Heredia S.M., Leak-Garcia J.A., Heraty J.M., Burger J.C., Yao L., Nohzadeh-Malakshah S., Ridley C.E. 2010. Crops gone wild: Evolution of weeds and invasives from domesticated ancestors. Evol. Appl. 3:494–504.

Ellstrand N.C., Prentice H.C., James F. 1999. Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics. p. 539–63.

Emshwiller E., Theim T., Grau A., Nina V., Terrazas F. 2009. Origins of domestication and polyploidy in oca (*Oxalis tuberosa*; Oxalidaceae). 3. AFLP data of oca and four wild, tuber-bearing taxa. Am. J. Bot. 96:1839–1848.

Fang L., Gong H., Hu Y., Liu C., Zhou B., Huang T., Wang Y., Chen S., Fang D.D., Du X., Chen H., Chen J., Wang S., Wang Q., Wan Q., Liu B., Pan M., Chang L., Wu H., Mei G., Xiang D., Li X., Cai C., Zhu X., Chen Z.J., Han B., Chen X., Guo W., Zhang T., Huang X. 2017. Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. Genome Biology. Genome Biology. p. 1–13.

Flowers J.M., Hazzouri K.M., Gros-Balthazard M., Mo Z., Koutroumpa K., Perrakis A., Ferrand S., Khierallah H.S.M., Fuller D.Q., Aberlenc F., Fournaraki C., Purugganan M.D. 2019. Cross-species hybridization and the origin of North African date palms. Proceedings of the National Academy of Sciences. p. 1651–1658.

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

Frisch M., Melchinger A.E. 2005. Selection theory for marker-assisted backcrossing. Genetics. 170:909–917.

Gao M., Ashu G.M., Stewart L., Akwe W.A., Njiti V., Barnes S. 2011. Wx intron variations support an allohexaploid origin of the sweetpotato [*Ipomoea batatas* (L.) Lam]. Euphytica. 177:111–133.

Gordon S.P., Levy J.J., Vogel J.P. 2019. PolyCRACKER, a robust method for the unsupervised partitioning of polyploid subgenomes by signatures of repetitive DNA evolution. BMC Genomics. 20:1–14.

Greiner S., Sobanski J., Bock R. 2015. Why are most organelle genomes transmitted maternally? BioEssays. 37:80–94.

Hansen A.K., Escobar L.K., Gilbert L.E., Jansen R.K. 2007. Paternal, maternal, and biparental inheritance of the chloroplast genome in Passiflora (Passifloraceae): Implications for phylogenetic studies. Am. J. Bot. 94:42–46.

van Heerwaarden J., Doebley J., Briggs W.H., Glaubitz J.C., Goodman M.M., de Jesus Sanchez Gonzalez J., Ross-Ibarra J. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc. Natl. Acad. Sci. 108:1088–1092.

Hirakawa H., Okada Y., Tabuchi H., Shirasawa K., Watanabe A., Tsuruoka H., Minami C., Nakayama S., Sasamoto S., Kohara M., Kishida Y., Fujishiro T., Kato M., Nanri K., Komaki A., Yoshinaga M., Takahata Y., Tanaka M., Tabata S., Isobe S.N. 2015. Survey of genome sequences in a wild sweet potato, *Ipomoea trifida* (H. B. K.) G. Don. DNA Res. 22:171–179.

Hornsey K.G. 1973. The occurrence of hexaploid plants among autotetraploid populations of sugar beet, (*Beta vulgaris* l.) and the production of tetraploid progeny using a diploid pollinator. Caryologia. p. 225–228.

Hoshino A., Jayakumar V., Nitasaka E., Toyoda A., Noguchi H., Itoh T., Shin T., Minakuchi Y., Koda Y., Nagano A.J., Yasugi M., Honjo M.N., Kudoh H., Seki M., Kamiya A., Shiraki T., Carninci P., Asamizu E., Nishide H., Tanaka S., Park K. Il, Morita Y., Yokoyama K., Uchiyama I., Tanaka Y., Tabata S., Shinozaki K., Hayashizaki Y., Kohara Y., Suzuki Y., Sugano S., Fujiyama A., Iida S., Sakakibara Y. 2016. Genome sequence and analysis of the Japanese morning glory *Ipomoea nil*. Nat. Commun. 7:1–10.

Hufford M.B., Lubinksy P., Pyhäjärvi T., Devengenzo M.T., Ellstrand N.C., Ross-Ibarra J. 2013. The Genomic Signature of Crop-Wild Introgression in Maize. PLoS Genet. 9.

Janzen G.M., Wang L., Hufford M.B. 2019. The extent of adaptive wild introgression in crops. New Phytol. 221:1279–1288.

Jones A. 1965. Cytological observations and fertility measurements of sweet potato (*Ipomoea batatas* (L.) Lam.). Proc. Am. Soc. Hortic. Sci. 86:527–537.

Kamneva O.K., Syring J., Liston A., Rosenberg N.A. 2017. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. BMC Evol. Biol. 17:1–19.

Kobayashi M. 1984. The Ipomoea trifida Complex Closely Related to Sweet Potato. Proceedings, Sixth Symposium of the International Society for Tropical Root Crops, hosted by CIP in Lima, Peru. p. 561–568.

Kobayashi M., Miyazaki T. 1975. Sweet Potato Breeding Using Wild Related Species. .

Kyndt T., Quispe D., Zhai H., Jarret R., Ghislain M., Liu Q., Gheysen G. 2015. The genome of cultivated sweet potato contains Agrobacterium T-DNAs with expressed genes : An example of a naturally transgenic food crop. .

Li M., Yang S., Xu W., Pu Z., Feng J., Wang Z., Zhang C., Peng M., Du C., Lin F., Wei C., Qiao S., Zou H., Zhang L., Li Y., Yang H., Liao A., Song W., Zhang Z., Li J., Wang K., Zhang Y., Lin H., Zhang J., Tan W. 2019. The wild sweetpotato

(*Ipomoea trifida*) genome provides insights into storage root development. BMC Plant Biol. 19:1–17.

Loebenstein G., Thottappilly G. 2009. The Sweetpotato. Springer.

Magoon M.L., Krishnan R., Vijaya Bai K. 1970. Cytological evidence on the origin of sweet potato. Theor. Appl. Genet. 40:360–366.

Mallet J., Besansky N., Hahn M.W. 2016. How reticulated are species? BioEssays. 38:140–149.

Marcussen T., Sandve S.R., Heier L., Spannagl M., Pfeifer M., The International Wheat Genome Sequencing Consortium, Jakobsen K.S., Wulff B.B.H., Steuernagel B., Mayer K.F.X., Olsen O.-A. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. Science (80). 345:1285–1288.

Martin F.W., Jones A., Ruberte R.M. 1974. A Wild *Ipomoea* Species Closely Related to the Sweet Potato. Econ. Bot. 28:287–292.

McDonald J.A., Austin D.F. 1990. Changes and Additions in *Ipomoea* Section Batatas (Convolvulaceae). Brittonia. p. 116–120.

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., Rausher M.D., Kelly S., Liston A., Scotland R.W., Williams B.R.M., Kronmiller B., Harris S.A., Anglin N.L., Ellis D., Rausher M.D., Kelly S., Liston A., Scotland R.W. 2018. Reconciling Conflicting Phylogenies in the Origin of Sweet Potato and Dispersal to Polynesia. Curr. Biol. 28:1246-1256.e12.

Muñoz-Rodríguez P., Carruthers T., Wood J.R.I., Williams B.R.M., Weitemier K., Kronmiller B., Goodwin Z., Sumadijaya A., Anglin N.L., Filer D., Harris D., Rausher M.D., Kelly S., Liston A., Scotland R.W. 2019. A taxonomic monograph of *Ipomoea* integrated across phylogenetic scales. Nat. Plants. 5:1136–1144.

Nishiyama I. 1959. Collecting the sweet potato and its allied species in USA and Mexico. Japanese Journal of Breeding. p. 73–78.

Nishiyama I. 1971. Evolution and Domestication of the Sweet Potato. Bot. Maj. Tokyo. 84:377–387.

Nishiyama I., Miyazaki T., Sakamoto S. 1975. Evolutionary autoploidy in the sweet potato (*Ipomoea batatas* (L.) Lam.) and its progenitors. Euphytica. 24:197–208.

Oracion M.Z., Niwa K., Shiotani I. 1990. Cytological analysis of tetraploid hybrids between sweet potato and diploid *Ipomoea trifida* (H. B. K.) Don. Theoretical and Applied Genetics. p. 617–624.

Orjeda G., Freyre R., Iwanaga M. 1990. Production of 2n pollen in diploid Ipomoea trifida, a putative wild ancestor of sweet potato. J. Hered. 81:462–467.

Orjeda G., Freyre R., 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.

Otto S.P., Whitton J. 2000. Polyploid Incidence and Evolution. Annu. Rev. Genet. 34:401–37.

Ozias-Akins P., Jarret R.L. 1994. Nuclear DNA content and ploidy levels in genus *Ipomoea.* J. Am. Soc. Hortic. Sci. 119:110–115.

Palmer J.D., Jorgensen R.A., Thompson W.F. 1985. Chloroplast Dna Variation and Evolution in Pisum: Patterns of Change and Phylogenetic Analysis. Genetics. 109:195–213.

Parisod C., Holderegger R., Brochmann C. 2010. Evolutionary consequences of autopolyploidy. New Phytol. 186:5–17.

Van De Peer Y., Mizrachi E., Marchal K. 2017. The evolutionary significance of polyploidy. Nat. Rev. Genet. 18:411–424.

Quispe-Huamanquispe D.G., Gheysen G., Yang J., Jarret R., Rossel G., Kreuze J.F. 2019. The horizontal gene transfer of *Agrobacterium* T-DNAs into the series Batatas (Genus *Ipomoea*) genome is not confined to hexaploid

sweetpotato. Sci. Rep. 9:1–13.

Ramsey J., Schemske D.W. 1998. Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. Annu. Rev. Ecol. Syst. 29:467–501.

Rieseberg L., Wendell J.F. 1993. Gene flow and its consequences in plants. In: Harrison R.G., editor. Hybrid Zones and the Evolutionary Process. Oxford University Press. p. 70–109.

Rieseberg L.H., Kim S., Randell R.A., Whitney K.D., Gross L., Lexer C., Clay K. 2007. Hybridization and the colonization of novel habitats by annual Sunflowers. Genetica. p. 149–165.

Rifkin J.L., Castillo A.S., Liao I.T., Rausher M.D. 2018. Gene flow, divergent selection and resistance to introgression in two species of morning glories (*Ipomoea*). Mol. Ecol.

Roullier C., Duputié A., Wennekes P., Benoit L., Fernández Bringas V.M., Rossel G., Tay D., McKey D., Lebot V. 2013. Disentangling the Origins of Cultivated Sweet Potato (*Ipomoea batatas* (L.) Lam.). PLoS One. 8.

Roullier C., Rossel G., Tay D., McKey D., Lebot V. 2011. Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. Mol. Ecol. 20:3963–3977.

Shiotani I., Kawase T. 1987. Synthetic Hexaploids Derived from Wild Species Related to Sweet Potato. Japanese J. Breed. 37:367–376.

Soltis D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants. 5:65–84.

Soltis D.E., Buggs R.J.A., Doyle J.J., Soltis P.S., Soltis D.E., Buggs R.J.A., Doyle J.J., Soltis P.S. 2010. What we still don't know about polyploidy. Taxon. 59:1387–1403.

Soltis P.S., Marchant D.B., Van de Peer Y., Soltis D.E. 2015. Polyploidy and genome evolution in plants. Curr. Opin. Genet. Dev. 35:119–125.

Srisuwan S., Sihachakr D., Martín J., Vallès J., Ressayre A., Brown S.C., Siljak-

Yakovlev S. 2018. Change in nuclear DNA content and pollen size with polyploidisation in the sweet potato (*Ipomoea batatas*, Convolvulaceae) complex. Plant Biol.:0–2.

Srisuwan S., Sihachakr D., Siljak-Yakovlev S. 2006. The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. Plant Sci. 171:424–433.

Tayalé A., Parisod C. 2013. Natural pathways to polyploidy in plants and consequences for genome reorganization. Cytogenet. Genome Res. 140:79–96.

Ting Y.C., Kehr A.E. 1953. Meiotic studies in the sweet potato: (*Ipomoea batatas* Lam.). J. Hered. 44:207–211.

Ting Y.C., Kehr A.E., Miller J.C. 1957. A Cytological Study of the Sweet Potato Plant *Ipomoea batatas* (L.) Lam. and its Related Species. Am. Nat. 91:197–203.

Ugent D., Peterson L.W. 1988. Archaeological Remains of Potato and Sweet Potato in Peru. CIP Circ. 16:1–10.

Wendel J.F., Clark Cronn R. 2003. Polyploidy and the Evolutionary History of Cotton. Botany Publication and Papers. .

Wood J.R.I., Muñoz-Rodríguez P., Williams B.R.M., Scotland R.W. 2020. A foundation monograph of *Ipomoea* (Convolvulaceae) in the New World. PhytoKeys. 143:1–823.

Wu S., Lau K.H., Cao Q., Hamilton J.P., Sun H., Zhou C., Eserman L., Gemenet D.C., Olukolu B.A., Wang H., Crisovan E., Godden G.T., Jiao C., Wang X., Kitavi M., Manrique-carpintero N., Vaillancourt B., Wiegert-rininger K., Yang X., Bao K., Schaff J., Kreuze J., Gruneberg W., Khan A., Ghislain M., Ma D., Jiang J., Mwanga R.O.M. 2018. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. Nature Communications. p. 1–12.

Yang J., Moeinzadeh M.-H.H., Kuhl H., Helmuth J., Xiao P., Haas S., Liu G., Zheng

J., Sun Z., Fan W., Deng G., Wang H., Hu F., Zhao S., Fernie A.R., Boerno S., Timmermann B., Zhang P., Vingron M. 2017. Haplotype-resolved sweet potato genome traces back its hexaploidization history. Nat. Plants. 3:696–703.

Zielinski M.-L., Scheid O.M. 2012. Polyploidy and Genome Evolution. Polyploidy and Genome Evolution. p. 33–55.

CHAPTER 2: A systematic review of tetraploid germplasm material shows taxonomic confusion has impeded sweet potato research & crop improvement

CHAPTER ABSTRACT

The discovery and identification of the closest wild relatives of domesticated crops, known as Crop Wild Relatives (CWRs), is an important step in understanding a crop's domestication history and facilitating potential crop improvement. Germplasm collections of CWRs are often poorly documented and frequently misidentified however, leading to confusion over their true taxonomic identity and phylogenetic relationship to the crop in question. This lack of clarity limits the potential power and impact of new research techniques such as high-throughput sequencing and genomics, thereby impeding or misleading progress in understanding the crop's domestication history, evolutionary relationships and contemporary diversity. Sweet potato (*Ipomoea batatas* (L.) Lam.) is an example of a major global crop for which potentially important CWR germplasm collections have been beset by disagreements and misunderstandings over their identity and potential role in the crop's origin. This is particularly true in the context of the search for the immediate ancestor, or "progenitor", of sweet potato, which has remained unidentified despite over 80 years of research. The recent publication of a well-resolved phylogeny and taxonomic monograph for the genus *Ipomoea* provides the opportunity to review sweet potato CWR germplasm collections previously suggested as candidates for this progenitor or otherwise potentially involved in sweet potato's origin. Doing so illustrates the negative impact past taxonomic confusion has had on sweet potato research and efforts to breed new varieties, while also highlighting opportunities for potential progress in resolving these issues to definitively identify the progenitor of hexaploid *I. batatas*. In this chapter, I provide a comprehensive account of all known germplasm collections previously identified

as potential candidates for the progenitor of sweet potato (those suggested to have been potentially involved in its origin), encompassing their collection history and geographic distribution, previous determinations and citations, use in research and crop improvement breeding programmes, and current availability in global repositories. I show how misidentifications, the lack of passport data and vouchered specimens, poor citations, and a failure to discriminate among heterogeneous material have all contributed to slow progress in sweet potato research and breeding. I then highlight the most pressing questions and promising paths for making progress on this key area of sweet potato research.

INTRODUCTION

The Role of CWR Germplasm Collections in Domestication Research

The selection and breeding of certain wild plants to meet humanity's needs brought about a fundamental shift in the development of both *Homo sapiens* L. and the rest of life on planet Earth. As such, studying crops and their wild relatives can improve our understanding of the evolutionary and cultural history of our species, while also providing tractable systems for testing hypotheses about evolution itself. This research has become increasingly urgent in light of dramatic contemporary biodiversity loss and environmental change, since it can help to illuminate how crop biodiversity has been generated and maintained in the past, while also identifying and documenting locally adapted landraces and crop wild relatives (CWRs) that are potential repositories of beneficial traits often lost in modern cultivated varieties.

The last decade has seen significant advances in our understanding of the evolutionary origin, domestication history and contemporary diversity of many crops (e.g. Cacao (Cornejo et al. 2018); Cotton (Fang et al. 2017); Dates (Gros-Balthazard et al. 2018; Flowers et al. 2019; Pérez-Escobar et al. 2021); Maize (Lorant et al. 2017; Stitzer and Ross-Ibarra 2018); Melon (Renner et al. 2021); Potato (Gutaker et al. 2019); Rice (Huang et al. 2012); Tomato (Razifard et al. 2020); Soy (Liu et al. 2020); Wheat (Marcussen et al. 2014); and Zuccini (Montero-Pau et al. 2018)). While much of this progress has been facilitated by technological innovations such as high-throughput sequencing and genomics (Zeder et al. 2006; Vaughan et al. 2007; Gross and Olsen 2010; Meyer and Purugganan 2013; Larson et al. 2014; Guerra-García and Piñero 2017; Gaut et al. 2018; Purugganan 2019), the majority of these studies have also been dependent on access to well-curated collections of germplasm material of the closest *crop wild relatives* (CWRs) of the crops in question (e.g. Huang et al. 2012; Marcussen et al. 2014; Lorant et al. 2017; Montero-Pau et al. 2018; Flowers et al. 2019; Gutaker et al. 2019; Liu et al. 2020;

Razifard et al. 2020; Pérez-Escobar et al. 2021; Renner et al. 2021). These collections of living plant material are essential for providing the context necessary to investigate questions such as the timing and location of probable domestication events (Huang et al. 2012; Marcussen et al. 2014; Cornejo et al. 2018; Gros-Balthazard et al. 2018; Gutaker et al. 2019; Razifard et al. 2020), evidence of historical hybridization, geneflow or introgression between crops and their CWRs (Huang et al. 2012; Marcussen et al. 2014; Fang et al. 2017; Gros-Balthazard et al. 2018; Flowers et al. 2019; Razifard et al. 2020; Pérez-Escobar et al. 2021), and the source and development of valuable traits and adaptations (Huang et al. 2012; Fang et al. 2017; Cornejo et al. 2018; Gutaker et al. 2019; Razifard et al. 2020; Renner et al. 2021).

Germplasm collections of CWRs also represent important sources of diversity and adaptive traits potentially useful in crop improvement breeding programmes (Hajjar and Hodgkin 2007; Khoury et al. 2014; Castañeda-Álvarez et al. 2016; Dempewolf et al. 2017), or even for domesticating entirely new crops (von Wettberg et al. 2020). This potential resource is particularly important given contemporary and predicted climate change, and wild and cultivated biodiversity loss (Khoury et al. 2014; Castañeda-Álvarez et al. 2016; Kilian et al. 2021).

Unfortunately, issues of misidentification and poor or erroneous documentation of germplasm collections are relatively common in many genebanks, with significant knock-on effects for their use in research and crop improvement breeding programmes (FAO 1998 p. 117, 2010 p. 77-78; Weise et al. 2020; van Bemmelen van der Plaat et al. 2021). If CWR germplasm material is not, or cannot be correctly identified, researchers cannot be sure of its true relationship to the crop, or whether other material with the same identification is really the same taxon. This makes it difficult to be confident of comparisons between studies based on separate material identified as the same taxon, potentially undermining any conclusions

drawn from them. In theory, unique institutional identifiers such as accession numbers should facilitate better comparisons, but only if accompanying data about collection history and phenotypic and genotypic characteristics are made publicly available and acknowledged by researchers (Dempewolf et al. 2017; Weise et al. 2020).

Institutions and researchers face many challenges in the management of CWR germplasm collections, including initial misidentification of material at collection or submission (Dempewolf et al. 2017; van Bemmelen van der Plaat et al. 2021); lack of basic passport data such as geographic origin, type of material, and phenotypic characteristics (Dempewolf et al. 2017; Weise et al. 2020); inconsistent or unstable identifiers within and among institutions (Weise et al. 2020); documentation errors resulting from institutional exchanges and material regeneration (van Bemmelen van der Plaat et al. 2021); and difficulty distinguishing true CWRs from feral or hybrid escapes from cultivation (Gepts 2004; Bagavathiannan and Van Acker 2008; Khoury et al. 2015; Gros-Balthazard et al. 2017; Renner et al. 2021). Most importantly, all of these issues are exacerbated if the taxonomy of the crop and its CWRs is difficult or poorly resolved (Dempewolf et al. 2017), which until recently has been the case for the globally important food crop, sweet potato (*Ipomoea batatas* (L.) Lam.) (Khoury et al. 2015).

Sweet Potato & Its CWRs

Despite its agricultural and dietary importance, sweet potato's evolution and domestication are still relatively poorly understood (see previous chapter). Most importantly, the means by which it evolved a hexaploid genome from a diploid one remains uncertain (Isobe et al. 2017; Yang et al. 2017). Progress on resolving outstanding questions about sweet potato's origin and contemporary diversity has been impeded by an incomplete knowledge of its closest CWRs, and considerable

confusion over the identification of both wild plants and existing germplasm collections held in scientific institutions around the world (Khoury et al. 2015).

Crop wild relative is a itself a relative term, with no fixed definition of how closely related to the crop they should be. One definition might include all or some of the species in the same genus, while another might cover only the immediate progenitor (wild ancestor) of the cultivated crop, or else any wild species that can be effectively crossed with it. For sweet potato, the genus *Ipomoea* contains over 800 species from around the world (Wood et al. 2020) but a small group of weedy species predominantly from Latin America and known as the "Batatas group" have long been effectively treated as its CWRs (Austin 1988).

The most recent monographic treatment of *Ipomoea* lists 15 species within a monophyletic Batatas group (Wood et al. 2020), greatly improving our understanding of sweet potato's currently known CWRs, and providing a stable framework for further research. Several batatas group species had previously been suggested as potential progenitors of sweet potato, but molecular phylogenetics has confirmed *I. trifida* (Kunth) G. Don, a weedy species from Mexico, Central America and the Caribbean coasts of Colombia and Venezuela, as the closest known CWR of *I. batatas*, rejecting a role for any other currently described species in its origin (Muñoz-Rodríguez et al. 2018). Following the "gene pool" system of Harlan and De Wet (1971), *I. trifida* would be considered GP-2, since fertile hybrids are possible (e.g. Orjeda et al. 1991) but the difference in ploidy complicates breeding. This is also likely the case of most other species in the Batatas group, but further crossing studies are required to confirm this. Meanwhile non-Batatas group species of Ipomoea would fall within GP-3, whereby gene transfer requires more radical interventions than conventional crossing (Harlan and De Wet 1971).

Despite this recent progress, it remains unclear how hexaploid *I. batatas* (2n=6x=90) might have evolved from a diploid ancestor shared with *I. trifida*

(2n=2x=30), given that the evolution of a hexaploid genome from a diploid ancestor most likely involves a tetraploid intermediary (Ramsey and Schemske 1998; Tayalé and Parisod 2013). The search for this tetraploid progenitor (2n=4x=60) of hexaploid sweet potato has therefore been an important research priority ever since *I. batatas* was first shown to be hexaploid (Ting and Kehr 1953), and yet it remains unresolved (Yang et al. 2017).

Existing Polyploid Germplasm Collections

In the course of the search for a tetraploid progenitor of sweet potato, a number of apparently wild polyploid plants closely resembling cultivated *I. batatas* or other Batatas group species have been collected across Latin America and suggested to be potentially involved in its origin (Table 1 & Fig. 1). To date, the relationship of these diverse collections to cultivated hexaploid sweet potato has remained unclear (Roullier et al. 2013), and there has been no comprehensive review that describes and compares all tetraploid germplasm collections. In this chapter I set out to summarise and review the existing evidence on the likely relationships of all these collections to cultivated hexaploid *I. batatas* and its closest known relative, diploid *I. trifida*. I begin by outlining the collection history of each collection, summarise what is known about its morphological characteristics, assess its past use in scientific research, and provide determination histories with reference to the most up to date taxonomic treatment for *Ipomoea* (Wood et al. 2020). I then use this information to assess what is and is not known about each collection and their relationship to cultivated *I. batatas*, highlighting the most urgent unresolved questions and potential avenues for progress.

MATERIALS & METHODS

I conducted a comprehensive literature review, cross-referencing and tracing all published material that cites collections of wild-collected polyploid germplasm

material previously associated with *I. batatas* (Table 1). I then used those citations and the online database genesys-PGR to establish in which repositories material was housed, and under what accession numbers. Where possible I contacted these institutions, requesting and receiving seeds of available accessions for cultivation and study at University of Oxford Department of Plant Sciences, as well as images of preserved herbarium specimens taken at the time of collection. I divided the works citing tetraploid germplasm material into four categories based on whether they were (i) describing an initial collection event, (ii) providing details on morphology, (iii) conducting research based on cytology, crossing experiments and palynology, (iv) using DNA-based molecular phylogenetics and genomics to explore evolutionary relationships, or (v) attempting to use these germplasm collections in crop improvement breeding programmes. In some cases, a single citation fell into more than one of these categories. By tracing the use of each individual collection through these five research categories, I aimed to compile a comprehensive summary of what is known about that collection, and to look for consistencies and differences among the different collections. This information allowed me to assess the different taxonomic determinations that had previously been applied to each collection, comparing them to the latest taxonomic treatment for the genus *Ipomoea* (Wood et al. 2020), and phylogenetic analysis of the Batatas group of sweet potato's closest CWRs (Muñoz-Rodríguez et al. 2018). Reviewing the likely accuracy of these past determinations and its influence on the conclusions drawn in different studies that cite them in turn demonstrates how disagreements, misidentifications and inconsistent sampling have impacted research on the origin of sweet potato and attempts to use wild germplasm material in crop improvement breeding programmes. Finally, I summarised the most important priorities and opportunities for future research aiming to use these collections to understand the origin of cultivated hexaploid sweet potato.

RESULTS & DISCUSSION:

Collection History of Polyploid Sweet Potato CWR Germplasm

In 1955 Ichizo Nishiyama, a Japanese researcher from Kyoto University, made the first in a series of important collections of sweet potato and CWR germplasm in Mexico (Nishiyama 1959; Nishiyama et al. 1961a; Nishiyama and Teramura 1962). The majority of these collections were of cultivated sweet potatoes gathered from fields and markets, while others were identified as wild species such as *I. trifida*, *I. triloba* L., *I. tiliacea* (Wild.) Choisy, *I. gracilis* R.Br.[1] or *I. caroliniana* Lam.[2] following van Ooststroom's (1953) classification of the genus *Ipomoea* (see Nishiyama and Teramura (1962) Table 1).

Two initially undetermined specimens from the state of Veracruz were given the accession numbers K134 and K123 (Figure 1A), and shown to be tetraploid (2n=4x=60) and hexaploid (2n=6x=90) respectively (Nishiyama 1959). The hexaploid (K123) was at first tentatively suggested to resemble *I. commutata* Roem. et Schulz[3] and shown to be self-incompatible but to cross readily with cultivated *I. batatas* (Nishiyama 1959; Nishiyama et al. 1961b). Thanks to its lack of noticeably thickened storage roots and its thinner, twining stems, K123 was later assigned to *I. trifida* (Nishiyama 1961; Nishiyama et al. 1961a; Nishiyama and Teramura 1962) and suggested to be the “ancestral wild species of the cultivated sweet potato” (Nishiyama 1971). Meanwhile, the tetraploid specimen (K134) failed to produce viable seeds when crossed with sweet potato (Nishiyama 1959), and was identified as either *I. gracilis* or *I. denticulata* Choisy[4], but was also said to resemble both *I. batatas* and *I. tiliacea* (Nishiyama et al. 1961a).

[1] Nom. illeg. Often (incorrectly) treated as a synonym of *I.littoralis* (Austin, 1978)
[2] Nom. illeg. superfl. = *I. carolina* L. (Wood et al., 2020)
[3] Nom. illeg. superfl. = *I. cordatotriloba* Dennst.
[4] another synonym of *I. littoralis* (Wood et al., 2020)

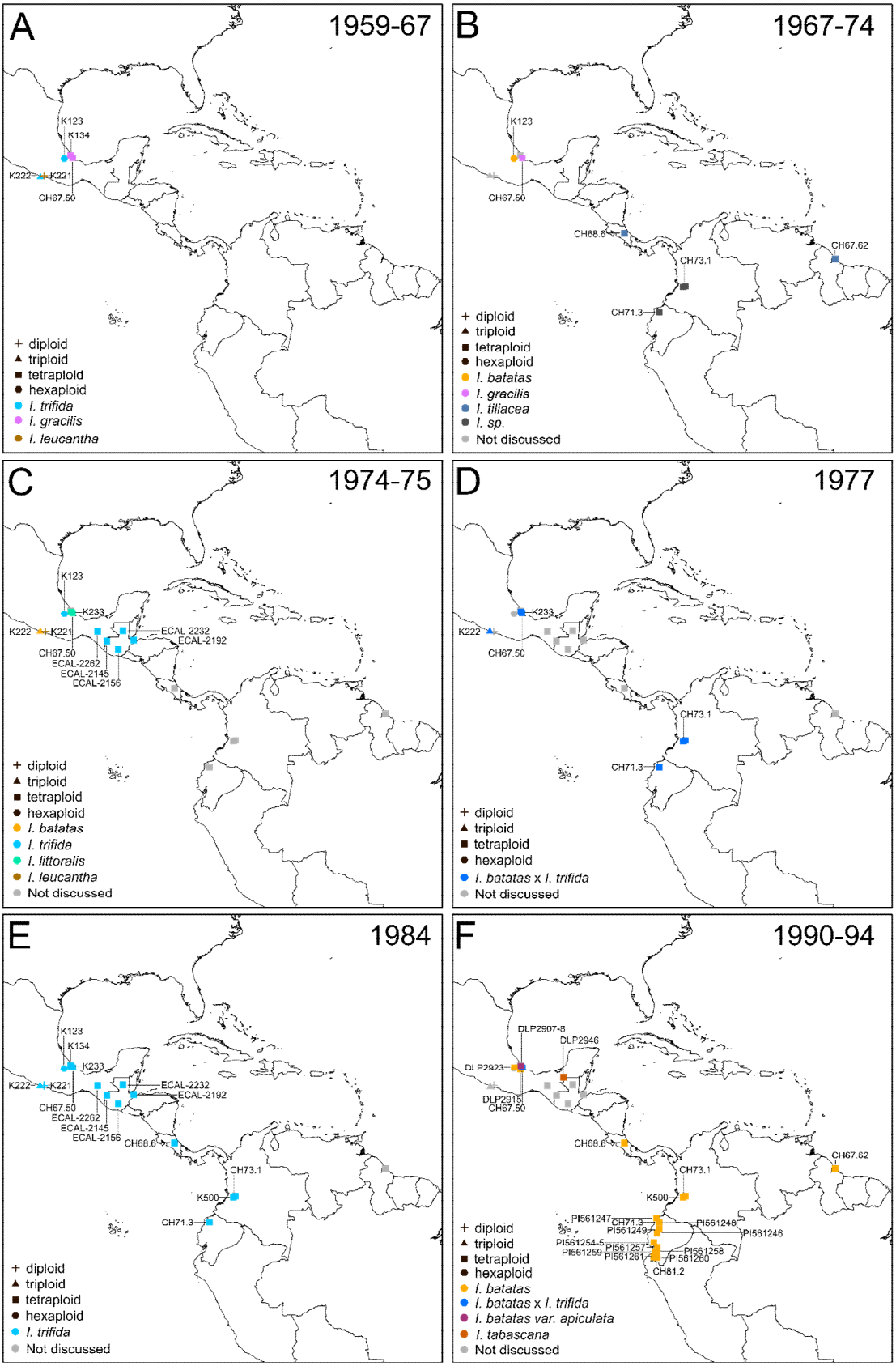


Figure 1: Maps showing locations of germplasm collections and their taxonomic determinations through time. Shape indicates ploidy, colour corresponds to determination. Relevant publications as follows: **A:** Nishiyama et al., 1961a; Teramura et al. 1967. **B:** Jones 1967, 1970; Martin & Jones 1972; Martin et al. 1974. **C:** Muramatsu & Shiotani, 1974; Kobayashi & Miyazaki, 1975. **D:** Austin, 1977. **E:** Kobayashi, 1984. **F:** Jones, 1990; McDonald & Austin,

In 1959, a colleague of Nishiyama named Mikio Muramatsu collected another tetraploid plant in the Mexican state of Veracruz, which was given the accession number K233 (Muramatsu and Shiotani 1974; Kobayashi and Miyazaki 1975) (Figure 1A). He identified this collection as *I. littoralis* Blume and noted its similarity to a separate tetraploid specimen collected in the same area by G. Chapman (CH67.50) (Muramatsu and Shiotani 1974)(Figure 1A), which had been identified as *I. gracilis* by Alfred Jones of the USDA (Jones 1970; Martin and Jones 1972).

A year later, Masahi Kobayashi of the Kagoshima Agricultural Experiment Station collected diploid (K221) and triploid (K222) plants from around the city of Acapulco (Figure 1A), designating them *I. leucantha* Jacq. and *I. trifida* respectively (Kobayashi and Miyazaki 1975). Together with colleague Itaru Shiotani, Muramatsu later collected a further 27 tetraploid plants from 17 disturbed sites in Mexico and Guatemala (Muramatsu and Shiotani 1974) including the accessions labelled ECAL 2145-2148, ECAL 2156-2157, ECAL 2191-2195, ECAL 2232 and ECAL 2262, (Figure 1C) all of which they identified as *I. trifida*. While these were all promising and potentially important collections of polyploid sweet potato CWRs, the taxonomic determinations of the relevant specimens were unfortunately highly confused. In fact, Nishiyama (1971) found differentiating between species in the Batatas group so challenging that he chose to re-circumscribe most of them into either *I. batatas* or *I. triloba*[5] – a decision not followed by subsequent authors. The name first applied to the diploid K221 for example, *I. leucantha*, in fact refers to sporadic hybrids between *I. cordatotriloba* and *I. lacunosa* in Southeastern USA (Austin 1978; Wood et al. 2020), neither of which is an immediate CWR of *I. batatas* (Muñoz-Rodríguez et al., 2018; Wood et al., 2020).

[5]*I. trifida*, *I. littoralis* and *I. leucantha* as forms or varieties of *I. batatas*, and *I. lacunosa*, *I. trichocarpa* and *I. ramoni* as forms of *I. triloba*.

Of the names given to the tetraploid specimens, *I. gracilis* (applied to K134 and CH67.50) is often treated as a synonym of *I. littoralis* (Austin 1978; Wood et al. 2020), though Robert Brown's original collection was in fact distinct from the type of *I. littoralis* and has been collected only once, in Australia (Austin 1978). Furthermore, *I. littoralis* (the determination applied to tetraploid K233) is itself known only from islands and sea shores in the Indian Ocean and western Pacific (Austin 1978; Wood et al. 2020), being completely absent from the Americas. Finally, *I. trifida* – the name most frequently applied to Nishiyama's hexaploid collection (K123) – is in fact a slender-stemmed vine with a circum-Caribbean distribution, and distinctively acute and chartaceous sepals (Bohac et al. 1993; Wood et al. 2020). Following concerns over the applicability of this name to K123 and other polyploid material (Jones 1967; Martin and Jones 1972; Austin 1977, 1978), Kobayashi (1984) introduced the notion of an "*Ipomoea trifida* complex" to cover all diploid, triploid, tetraploid, pentaploid and hexaploid plants that were the hypothesised ancestors of cultivated sweet potato (*I. batatas*) (Figure 1E). More recent phylogenetic (Roullier et al. 2013; Muñoz-Rodríguez et al. 2018) and taxonomic (Wood et al. 2020) research have shown *I. trifida* to be a well-circumscribed, purely diploid species, monophyletic and morphologically distinct from all these various polyploid collections (see below). As we shall see, the subsequent confusion caused by these inaccurate species determinations has been only somewhat offset by the provision and use of Japanese National Agriculture and Food Research Organization (NARO) accession numbers.

Despite this taxonomic and nomenclatural confusion however, the discovery of tetraploid and hexaploid plants superficially resembling sweet potato and apparently growing wild in Mexico and Central America prompted further searching for similar material in the same areas. Perhaps most notably, the Peruvian botanist Fermin De La Puente collected hundreds of specimens of sweet potato and its

CWRs from the wider Batatas Group across Latin America, depositing them as seeds and germplasm at the International Potato Centre (CIP), Peru.

As part of these efforts, and in collaboration with Dan Austin from Florida Atlantic University and others, he collected a number of *I. batatas*-like plants in Southern Mexico (Contreras et al. 1994), five of which (DLP2908, DLP2915, DLP2918, DLP2923, and DLP2946 (Figure 1F)) were subsequently shown to be tetraploid (Ozias-Akins and Jarret 1994). In 1990, DLP2946 was described as a new species (*I. tabascana* McDonald & Austin), while DLP2908 was designated a variety of sweet potato (*I. batatas* var. *apiculata* (Martins & Galeotti) McDonald & Austin) (McDonald and Austin 1990). Both were observed to be restricted to very small geographic areas, with *I. tabascana* found in a single marshy location near the Guatemalan border, and *I. batatas* var. *apiculata* on coastal sand dunes in the Mexican Gulf port of Veracruz.

The majority of early sweet potato CWR germplasm collecting had been focused on Mexico, but subsequent efforts also uncovered tetraploid material growing in northwest South America (Fig. 1B-F). In 1970, the leading expert on wild tomato germplasm, Charles M. Rick of UC Davis, collected a tetraploid plant in Esmeraldas province, northwest Ecuador (PI355830/CH71.3/K300) (Figure 1B), which he described as a “common weed with all features” of cultivated sweet potato (Martin et al. 1974). He followed this up with two further collections in 1972 (both CH73.1) from the Cali and Palmira areas of south-central Colombia (Figure 1B), which Martin et al. (1974) considered to be identical to his earlier Ecuadorean collection. Kobayashi (1984) also observed many similar plants in the area around Cali, a specimen of which (K500/CH80.3) (Figure 1E) he collected in 1976 (Kobayashi 1978). In August 1980, Rick made another tetraploid collection in southern Ecuador (SAM5316/CH81.2) (Figure 1F), which was later found to produce unreduced pollen (Jones 1990). Although Martin et al. (1974) believed these South

American tetraploids to all be close CWRs of sweet potato, Austin (1977, 1982) initially suggested that along with the material from Mexico and Guatemala, these collections were all most likely hybrids between cultivated sweet potato and a diploid relative (probably *I. trifida*) (Figure 1D). Nevertheless, in February 1992 Austin and De la Puente led an expedition to Ecuador to collect more germplasm of this "tetraploid *I. batatas*", depositing 15 accessions at CIP and the USDA (Figure 1F), with accompanying herbarium vouchers sent to FAU (Austin et al. 1993).

Table 1 at the end of this chapter lists the accession numbers, collection details and past determinations of all known polyploid potential sweet potato CWR germplasm accessions – that is polyploid accessions previously hypothesised as potentially involved in the origin of hexaploid *I. batatas* – and the following section summarises the results of the various studies that have utilised this material to investigate different aspects of its evolution and biology.

Notes on Morphological Attributes

While the majority of the initial taxonomic determinations outlined above were based primarily on analysis of morphological characters, in most cases these were not discussed systematically, explicitly or in detail. High-resolution photographs, illustrations or measurements of key traits were not generally provided, and in the majority of cases, no reference was made to the justification for determinations, nor to preserved herbarium vouchers that might be requested or viewed for comparison. This severely limits reproducibility in these studies and subsequent research that uses polyploid CWR germplasm accessions, making comparison among them extremely challenging.

Below I outline and compare all available information on the morphology of these collections in an effort to summarise what is known about their phenotypic characteristics. A key issue when considering the morphology of these specimens in

relation to cultivated sweet potato however, is that the crop is itself highly variable, making it difficult to assess the validity of comparisons and inferences based solely on morphological similarity or difference.

Martin et al. (1974) for example, describe leaf size, shape and pubescence in sweet potato to be so variable as to offer little utility as taxonomic characters, though they also rely on these characters to a certain extent in drawing their own conclusions about the relationship of tetraploid collections to cultivated hexaploid sweet potato. Perhaps the key taxonomic character in *Ipomoea* is the size and shape of its sepals (calyx) (Austin 1978; Wood et al. 2020) but few studies on polyploid germplasm collections give details for this trait. The following notes on morphology should therefore be treated with a degree of caution.

The only tetraploid germplasm collections to receive formal taxonomic descriptions based on detailed analysis of their morphology are DLP2946 (*I. tabascana*), and DLP2908 (*I. batatas* var*. apiculata*) (McDonald and Austin 1990). The former is described as a slender, herbaceous, low-growing and twining vine, rooting at the nodes, with narrow sagittate leaves, and 2-5-flowered inflorescences of 3-3.5cm long funnelform corollas and subtly unequal sepals 7-9mm long, obtuse to acute at the tip and completely glabrous (Fig. 2).

In contrast, *I. batatas* var. *apiculata* is described as more or less completely prostrate, also rooting at the nodes, but with small, somewhat fleshy leaves, broadly ovate to cordate and mostly with obtuse lobes, producing 1-2-flowered cymes of 3-3.5cm long, funnelform, lavender corollas, and pronouncedly unequal, obtuse, glabrous sepals, 3-9mm long (Fig. 2). Both were described from very few specimens[6] and their ploidy was not discussed until four years later (Ozias-Akins and Jarret 1994).

[6] One in the case of *I. tabascana*, and three for *I. batatas var. apiculata*

In his initial description of K123 as the potential wild ancestor of cultivated sweet potato, Nishiyama and Teramura (1962) did give relatively detailed observations of this apparently wild hexaploid collection, contrasting in particular its thin, twining, and densely hairy stems and lack of pronouncedly thickened roots with the more robust, glabrous stems and thickened roots of a series of sweet potato cultivars. He made no mention of floral characters such as sepal dimensions or thickness however, which are key to the diagnosis of *I. trifida* (Wood et al. 2020). It was this reliance on vegetative characters that caused so much subsequent confusion, and Jones (1967) later showed that the characters highlighted by Nishiyama (1971) were in fact common in first-generation offspring of sweet potato – as were its flowers – falling well within acceptable variation for the crop, all of which he believed pointed to K123 in fact being a feral escape rather than its wild progenitor.

Nishiyama (1959) had also described tetraploid K134 as twining, with glabrous, lobed or unlobed leaves and few-flowered cymes of pink funnel-shaped flowers with unequal sepals, but no description of the morphology of diploid K221, triploid K222 or tetraploid K233 was given at collection. Later authors suggested that K233 strongly resembled K134 (Martin et al. 1974; Bohac et al. 1993), as well as Chapman's CH67.50 (Martin et al. 1974; Muramatsu and Shiotani 1974; Oracion et al. 1990) and *I. batatas* var. *apiculata* (DLP2908) (Jarret et al. 1992; Bohac et al. 1993). Martin et al. (1974) described diploid K221 ("*I. leucantha*") as having acuminate sepals with a short (25mm) "salver form" corolla and constantly flowering 1-9-flowered inflorescences, while tetraploid K233 ("*I. littoralis*") had short (20mm) but campanulate corollas in few-flowered (1-6) inflorescences, with glabrous and mucronate sepals, and a short-day flowering cycle. Meanwhile, Muramatsu and Shiotani (1974) described their tetraploid collections from southern Mexico and Guatemala (ECAL 2145-2148, ECAL 2156-2157, ECAL 2191-2195, ECAL 2232 and

ECAL 2262) as variable but resembling K123 in being frequently pubescent, with whitish-pink corollas and slightly thickened, but never tuberous roots.

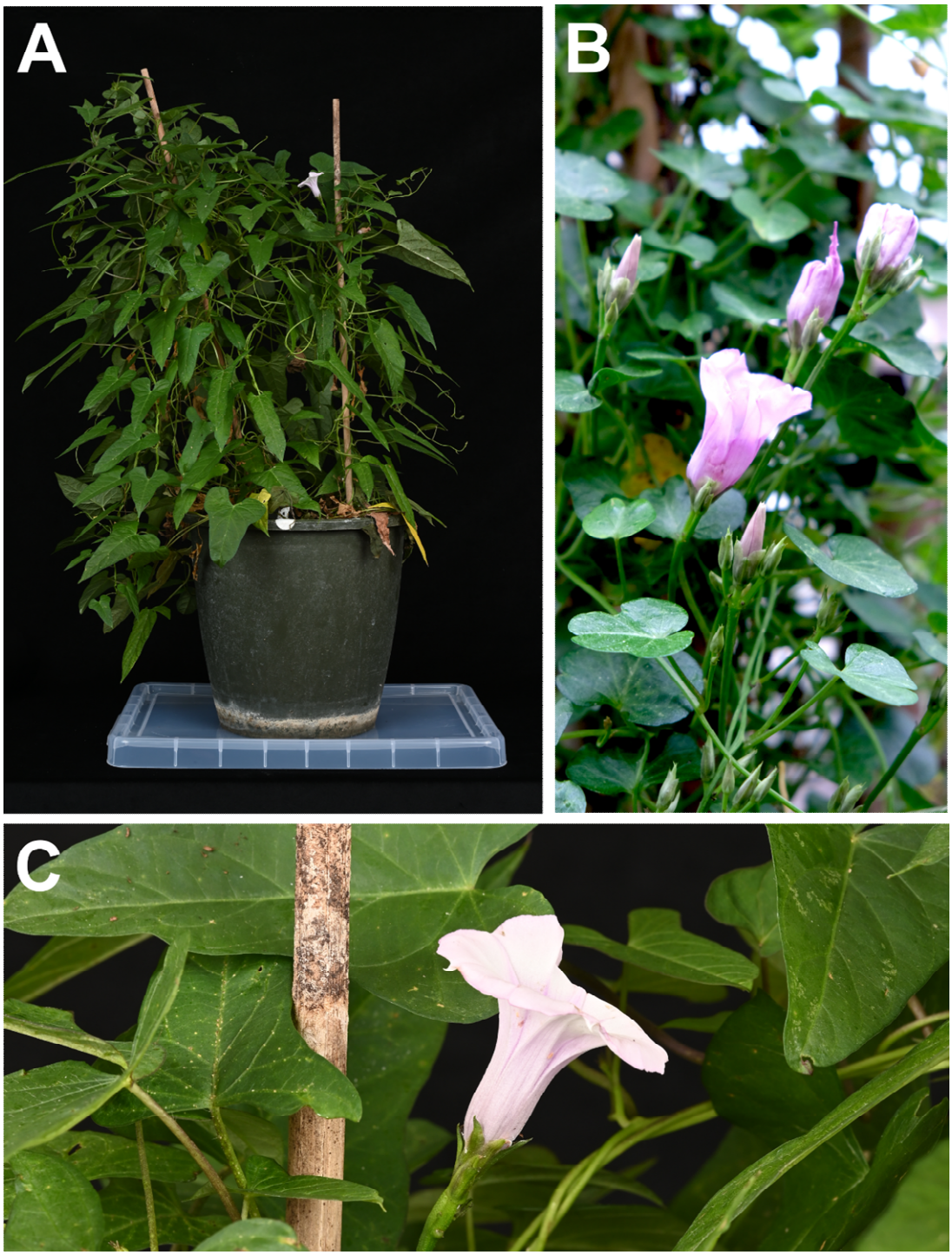


Figure 2: Accessions of **A&C:** DLP2946 (*I. tabascana*), and **B:** DLP2908 (*I. batatas* var. *apiculata*) grown from seed in greenhouses at University of Oxford Department of Biology. Showing sagittate leaves (**A**) and acuminate sepals (**C**) in *I. tabascana*, and small, multi-lobed, and leathery leaves and broadly campanulate corollas with obtuse sepals (**B**).

Martin et al. (1974) did not consider any of the tetraploid collections from Mexico and Central America to strongly resemble cultivated sweet potato, or by extension to be plausible candidates for its progenitor, describing their "leathery foliage, small leaves" and lack of tuberous roots as difficult to confuse with those of *I. batatas*, though they also noted the cultivated crop's pronounced variability. They found a closer resemblance to *I. batatas* in the tetraploid collections made by Rick in Ecuador (PI355830/CH71.3) and Colombia (Martin et al. 1974), which had 35-40mm long, campanulate flowers "that could be easily confused with those of I. batatas", held in inflorescences of up to 15 flowers with a short-day flowering cycle, though these plants were also reported to lack any form of tuberous root.

In one of the few comprehensive morphological studies to include a broad range of tetraploid collections and use quantitative statistical analysis, Rick's two Ecuadorean tetraploids (CH71.3, CH81.2), Kobayashi's Colombian one (CH80.3), two tetraploid accessions from Costa Rica and Guyana previously identified by Martin and Jones (1972) as *I. tiliacea* (CH67.62 and CH68.6), and three Mexican collections that included CH67.50, CH73.2[7] and an un-identified specimen collected by Austin, were all compared to seven specimens of diploid *I. trifida* and 25 sweet potato cultivars in order to assess their morphological similarity (Bohac et al. 1993). Based on corolla tube and sepal dimensions, the authors suggested that all this tetraploid material with the exception of CH67.50 (which they believed to be a hybrid) represented a tetraploid race of *I. batatas* (Figure 1F). The tetraploids shared with hexaploid *I. batatas* a significantly more obtuse mean sepal apex than that of diploid *I. trifida* (100.5 º and 95.4 º vs 55.1º) and a narrower corolla tube (11.4mm and 11.8mm vs 7.8mm), while their sepals were on average consistently shorter than in either hexaploid *I. batatas* or diploid *I. trifida* (6.6mm vs 8.5mm and 8.7mm).

[7] No reference could be found of this accession elsewhere in the literature and neither it nor any details on it were available on request from the USDA.

These characters had been noted previously (e.g. Austin 1978, 1988), but this was the first study to use quantitative measurement and statistical analysis to assess, describe and differentiate them. The sole exception in Bohac et al.'s (1993) results was Chapman's tetraploid collection from Veracruz, CH67.50, which did not match the other tetraploid material they analysed, particularly in terms of its sepal shape, leading them to agree with the earlier assessment of Austin (1977) that it was a likely hybrid. Exact details of the differences between CH67.50 and the other tetraploid material were not given, but Martin and Jones (1972) describe CH67.50 as a climbing and trailing perennial, rooting at the nodes, with lobed or entire, leathery, glabrous leaves, and inflorescences of one or two to five flowers, each with acuminate sepals and a pink to purplish corolla, and producing somewhat thickened roots, much of which appears to match Nishiyama (1959) description of K134, which had also previously been identified as *I. gracilis*. Interestingly, Bohac et al. (1993) also discussed (though they did not analyse) the recently described tetraploid variety *I. batatas* var. *apiculata* (DLP2908), noting its similarity to both K134 and K233, but failing to mention any resemblance to CH67.50, despite Muramatsu and Shiotani (1974) having described CH67.50 and K233 as almost identical.

As further evidence of the likely hybrid origin of CH67.50, Bohac et al. (1993) cited its failure to produce seed in over 20 years of cultivation. This had also been noted by Martin and Jones (1972), who ascribed it to previously reported abnormalities at meiosis (Jones 1970). In contrast, Bohac et al. (1993) believed all the other tetraploid material they had analysed to represent a "tetraploid race" of *I. batatas*, resembling "primitive sweet potatoes", and producing thickened "pencil-shaped" storage roots, which would re-sprout when cut. Finally, they could find no evidence for the existence of any polyploid accessions that matched the morphological circumscription of *I. trifida*, despite an active and extensive search,

believing material identified as such by other authors to instead fit their definition of tetraploid *I. batatas* (Bohac et al. 1993).

Further tetraploid collections made in western Ecuador were also described as matching this "tetraploid *I. batatas*", and said to be identifiable by "outer sepals that were shorter than those of hexaploid *I. batatas*" (Austin et al. 1993). Examination of plants grown from seed at Oxford, and herbarium specimens taken at the time of collection, showed these later collections as well as Rick's Ecuadorean (CH71.3, CH81.2) and Kobayashi's Colombian (K500/CH80.3) specimens to produce tall vines (>3m), with slender twining stems and large broadly cordate (occasionally trilobed) leaves, and many-flowered inflorescences of funnelform corollas and short, blunt, unequal sepals.[8]

In summary, by piecing together the patchy and inconsistent descriptions and comparisons of recorded tetraploid CWR germplasm collections with new observations from available living and preserved material taken at collection, it can be seen that there appears to be considerable morphological variation among them, but with an apparent geographical split. Mexican and Central American collections appear to form a heterogeneous group of few-flowered, trailing or low-growing vines with small, variably lobed leaves, and acute or obtuse sepals of varying lengths. In contrast, Colombian and Ecuadorian material are more consistently tall, simple-leaved twining vines, with many-flowered inflorescences and blunt, unequal, hairy sepals.

While this split was first hypothesised by Martin et al. (1974) however, Bohac et al. (1993) do not appear to have recognised it, instead treating all but one of the collections (CH67.50) they analysed from Mexico and Costa Rica as tetraploid *I. batatas*, alongside Ecuadorian and Colombian material. Unfortunately, the Mexican

[8] for more detailed and quantitative analysis of this and other material, please see following chapter.

and Central American accessions analysed by Bohac et al. (1993) were not available on request from the USDA and may no longer exist, ruling out re-evaluation.

What is more, on their own, these potential differences in morphology provide no definitive evidence as to the exact relationship of either group of specimens to cultivated hexaploid sweet potato, resolving which question will likely require evidence based on observations of genomic structure and molecular phylogenetic analysis.

Cytology, Palynology & Crossing experiments

Prior to the advent of DNA sequencing and molecular phylogenetics, cytological analysis and crossing experiments formed the primary means of assessing relatedness in hypothesised sweet potato CWRs, and their potential role in the crop's origin. Meanwhile, palynology has been used to investigate changes in genome size, the incidence of unreduced gametes and their potential role in the transition to hexaploidy and subsequent geneflow between *I. batatas* and its diploid relatives.

When Ting and Kehr (1953) first confirmed sweet potato to be hexaploid, they also noted a high frequency of secondary association of bivalents at metaphase I, which they attributed to a likely allopolyploid origin for *I. batatas* through hybridization of a diploid and a tetraploid ancestor. As a result, they highlighted the importance of searching for potential tetraploid progenitors in Central and South America. This led them to expand their subsequent cytological studies to include potential CWRs, which they also attempted to hybridize with hexaploid *I. batatas* in over 3000 controlled crosses (Ting et al. 1957). Two wild accessions originally collected in Cuba and Puerto Rico were found to be tetraploid (2n=4x=60). The Puerto Rican tetraploid was identified as *I. tiliacea*, and the Cuban tetraploid as

either *I. gracilis* or *I. fastigiata*[9], while remaining diploid (2n=2x=30) accessions were all identified as more distantly related species of *Ipomoea* from outside the Batatas group (Ting et al. 1957).

Unfortunately, no accession numbers, vouchered specimens or detailed morphological descriptions were given for any of this material, so it is not possible to confirm their taxonomic determinations, or whether they resembled other tetraploid collections that received similar determinations, such as K134 or CH67.50. However, the failure to produce any successful hybrids between either of their tetraploid accessions and multiple individuals of hexaploid sweet potato led Ting et al. (1957) to rule both out as potential progenitors of hexaploid *I. batatas*.

In a series of crossing and cytological studies that attempted to reconstruct the evolution of sweet potato's hexaploid genome from the polyploid collections made in Mexico, Japanese researchers showed that crosses between K221, K233, K123 and *I. batatas* were possible (Nishiyama 1971; Kobayashi and Miyazaki 1975; Nishiyama et al. 1975; Shiotani and Kawase 1987), which had not been the case for previous attempts to cross sweet potato with other species of *Ipomoea* (e.g. Ting et al. 1957; Wedderburn 1967). They also noted that in contrast to other members of the Batatas group they analysed[10], these accessions were all highly self-incompatible, in common with *I. batatas*. In combination, these results led them to designate these collections as the "B group" of immediate CWRs closely related to the cultivated crop (Nishiyama et al. 1975; Shiotani and Kawase 1987, 1989), which would later be re-named the "*I. trifida* complex" (Kobayashi 1984; Oracion et al. 1990).

[9] As described above, *I. gracilis* was the name initially applied to Nishiyama's K134 and Chapaman's CH67.50 (both from Mexico and also tetraploid) but is often (incorrectly) treated as a synonym of *I. littoralis*, which is likely restricted to Australasia and the Indian Ocean. Meanwhile *I. fastigiata* is now recognized as a synonym of *I. tiliacea* (Wood et al., 2020).

[10] (*I. lacunosa, I. triloba, I. trichocarpa* (=*I.cordatotriloba*), and I. ramoni (=*I. trifida/I. triloba*))

Nishiyama (1971) went on to hypothesize that the triploid collected from Acapulco (K222) had arisen from hybridization between diploid K221 ("*I. leucantha*") and tetraploid K233 ("*I. littoralis*"). He believed K222, which was shown to have extremely poor seed set (0.14%), but to be capable of producing diploid, tetraploid and hexaploid progeny (Nishiyama 1971), had itself undergone somatic doubling to become the supposedly wild hexaploid K123 ("*I. trifida*"), which had in turn been taken into cultivation to become cultivated *I. batatas* (sweet potato). As support for this hypothesis, he cited the development of a synthetic hexaploid through the crossing of tetraploid K233 with diploid K221, (Nishiyama 1971; Shiotani and Kawase 1987), which was said to have a similar chromosome arrangement to cultivated *I. batatas*[11].

If correct, this theory would make K233 the most likely tetraploid progenitor of hexaploid sweet potato. However, the later crossing experiments of Oracion et al. (1990) complicated this conclusion, by demonstrating that two tetraploid specimens collected wild in Guatemala (ECAL2232-33, ECAL2232-34) – and which they regarded as identical to K233, Jones' *I. gracilis* (CH67.50) and two wild tetraploids from Ecuador and Colombia (CH71.3 & CH73.1) – showed no noticeable cytological differences from artificially generated tetraploid F1 hybrids between hexaploid *I. batatas* cultivars and diploid *I. trifida*. These conflicting results serve to underline the extreme difficulty of differentiating between potential wild progenitors and more modern hybrids in such closely related and recently diverged plants of different ploidies.

As part of an attempt to generate a storage root-producing tetraploid breeding population, controlled crosses between five sweet potato cultivars and 41 accessions of diploid *I. trifida* also provided insights into cross-compatibility and

[11] A second artificial hexaploid was also bred from a cross between tetraploid K134 ("*I. gracilis*") with a diploid *I. lacunosa* L. specimen (K61) obtained from Baton Rouge, Louisiana, but this does not appear to have born any resemblance to *I. batatas*.

potential changes in ploidy in *I. batatas* and its closest known CWR (Orjeda et al. 1991). Firstly, that these crosses were possible, if somewhat difficult, with a success rate of one successfully germinated seedling for every 114 cross-pollinations, and secondly that all of the successful progeny were described as possessing morphological growth habits intermediate between the two parents, matching the description of most of the wild-collected tetraploid germplasm collections from Mexico and Guatemala (see above).

The results also showed that crosses between hexaploid *I. batatas* and diploid *I. trifida* would normally produce the expected tetraploid F1 offspring, but could also produce both hexaploid (21% of successful crosses) and pentaploid (2%) progeny. One implication of these results is that contact between cultivated sweet potato and wild growing diploid *I. trifida* is relatively likely to produce tetraploid (and sometimes pentaploid or hexaploid) hybrids that would resemble “weedy” sweet potatoes difficult to distinguish from any potential wild tetraploid progenitor. The second was that unreduced (2n) gametes previously reported for *I. trifida* (Orjeda et al. 1990) can produce successful polyploid crosses, reinforcing their likely role in the origin of sweet potato’s hexaploid genome.

The production of unreduced gametes represents the most common route to polyploidy in plants, whereby meiosis fails to generate haploid cells, and gametes instead possess a full diploid set of chromosomes (Ramsey and Schemske 1998). This has made establishing whether sweet potato CWRs produce unreduced gametes an important area of study, which can be achieved by looking for bimodal distributions in pollen grain diameter. Using this method, unreduced pollen production has been observed and confirmed cytologically in both diploid *I. trifida* (Orjeda et al. 1990), and a tetraploid accession from southern Ecuador (CH81.2) (Jones 1990; Becerra Lopez-Lavalle and Orjeda 2002). In the case of diploid *I. trifida*, (Orjeda et al. 1990) found that 10% of the 429 individuals studied produced

unreduced pollen, representing 55% of the 31 accessions those individuals were derived from.

These results demonstrate that non-reduction of pollen is a widespread phenomenon in sweet potato's closest known diploid relative, supporting its likely role in the origin of hexaploid *I. batatas*. In the case of tetraploid accession CH81.2, around 90% of individuals produced unreduced pollen, most likely as a result of abnormal spindle configurations (Becerra Lopez-Lavalle and Orjeda 2002), and Jones (1990) showed that non-reduction was also likely heritable. Nonetheless, Becerra Lopez-Lavalle and Orjeda (2002) failed to produce hexaploid progeny through controlled crosses between tetraploid accessions with reduced and unreduced gametes, which they suggested indicated the existence of a potential barrier to this route to hexaploidy in sweet potato, though it seems possible this could also be a result of strong self-incompatibility in *I. batatas* and its closest relatives.

Given that pollen size correlates strongly with nuclear DNA content in this manner, it also represents a valuable way of rapidly assessing genome size and likely ploidy in sweet potato CWRs. Srisuwan et al. (2018) for example, found that pollen grains of diploid *I. trifida* had a mean diameter of 76.9(±3.8)µm, while those of hexaploid *I. batatas* were 102.9(±14.8)µm (Fig. 3). Six tetraploid accessions were also assessed, including *I. tabascana* (PI518479) and five artificially bred tetraploids identified as "*I. trifida* 4x"[12]. Pollen grain diameter in *I. tabascana* was shown to be 83.2(±3.2)µm, while in the hybrid tetraploids ("*I. trifida* 4x") it was 91.2(±8.9)µm, placing both intermediate in diameter between diploid *I. trifida* and hexaploid *I. batatas* (Srisuwan et al., 2018).

[12]No collection or passport data was given for these five specimens (CIP695141, CIP695143, CIP695146, CIP695152, CIP694436), and no reference could be found to them in the literature or through CIP/Genesys-PGR online germplasm databases. Upon enquiry to CIP, I was informed that these accessions are artificial tetraploid hybrids created as part of a crop improvement breeding program (pers. comm. CIP/Rosario Falcon Meza).

Pollen diameters have also been recorded for diploid K221 (97µm), tetraploids CH67.50 (107µm), CH71.3 (132 µm), CH73.1 (111µm), K233 (112µm), and an unknown specimen of hexaploid *I. batatas* (186 µm) (Martin et al. 1972, 1974). While these results mirror those of Srisuwan et al. (2018) in terms of the correlation between increases in ploidy and pollen diameter (Fig. 3), their sizes for each category are considerably larger. This may be the result of differences in methodology and pollen preparation, since Martin et al. (1972, 1974) make no mention of the use of acetolysis to remove the outer layer of proteins, while Srisuwan et al. (2018) did so.

In cytogenetic analysis of hexaploid *I. batatas* and CWRs, the same set of tetraploid accessions including *I. tabascana* and the five artificial tetraploids labelled "*I. trifida* 4x" were assessed using fluorescence in situ hybridization (FISH) in an attempt to resolve their relationship to sweet potato (Srisuwan et al. 2006). Variation in copy numbers of the 18S and 5S regions, as well as karyotypic characters were taken to suggest *I. tabascana* arose through hybridization between hexaploid *I. batatas* and diploid *I. trifida*. The authors did not reach the same conclusion for their five "*I. trifida* 4x" accessions however, despite these having been generated through controlled-crosses at CIP, which suggests they were unaware how these accessions had originated.

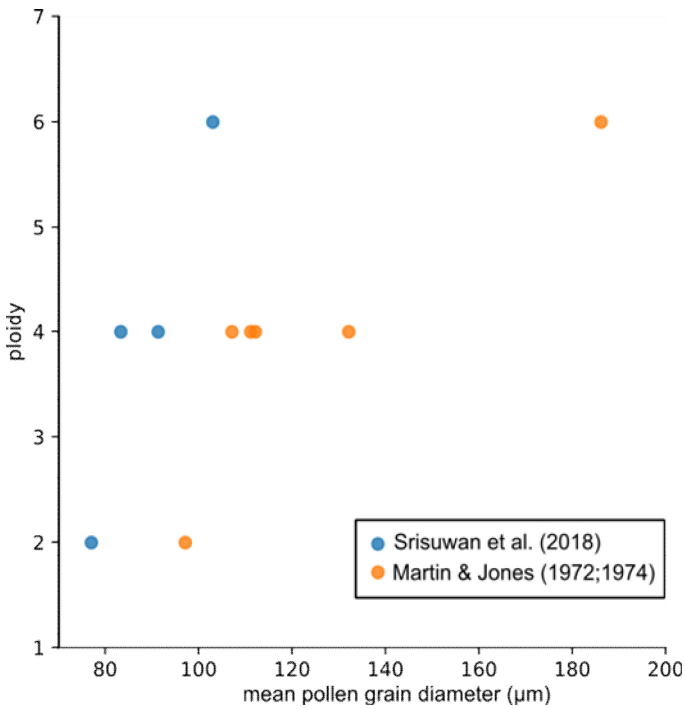


Figure 3: Graph comparing mean pollen diameter to ploidy in previous studies. Blue dots showing samples from Srisuwan et al. (2018), orange dots from Martin & Jones (1972,1974)

These artificial tetraploids shared with *I. tabascana* identical numbers of copies of the 18S and 5S regions (10 and four respectively), and a similar number of CMA+ bands (10 in *I. tabascana* vs 12 in "*I. trifida* 4x"), but differed in the number of secondary constrictions (SAT chromosomes), with two observed in *I. tabascana*, versus eight in "*I. trifida* 4x". Numbers of copies of 18S and numbers of CMA+ bands were also shown to be variable in hexaploid *I. batatas*, with the majority of accessions possessing 12 of each, but one having 16 and another 18. This suggests there is some variation among the genomes of different hexaploid *I. batatas* accessions. Unfortunately, the number of secondary constrictions in hexaploid *I. batatas* could not be recorded, due to the difficulty of identifying centromeres in the small and densely clustered chromosomes. Meanwhile, Martin et al. (1974) also observed cytological differences between Colombian and Ecuadorian tetraploid collections, specifically that Colombian accession CH73.1 showed frequent secondary associations at meiosis, similar to those observed in hexaploid *I. batatas*, while these were not seen in Ecuadorian accession CH71.3. Both Ecuadorian and Colombian accessions showed frequent multivalent formation however, potentially indicating a high level of homology between subgenomes, which was described as suggestive of an autopolyploid origin.

The combined evidence from cytological and palynological analyses reveals considerable variation both within hexaploid *I. batatas* and across different potential tetraploid CWR accessions. The potentially close relationship among these taxa, and the somewhat unpredictable outcomes of changes in ploidy and potential hybridization make drawing inferences based purely on observations of chromosome arrangements at meiosis difficult however. In the past, the formation of bivalents at meiosis was taken as evidence of allopolyploidy for example, while multivalents were considered evidence for autopolyploidy, but this distinction is no longer widely accepted (Bomblies 2022). Once again therefore, the major issue

remains distinguishing between candidates for the autotetraploid progenitor of hexaploid *I. batatas* and potential hybrid offspring of hexaploid *I. batatas* and diploid *I. trifida*.

Molecular Phylogenetics & Genomics

With the advent of molecular phylogenetics, came the opportunity to put some of the hypothesised relationships of CWR germplasm to cultivated sweet potato to a new test. Unfortunately, the majority of molecular phylogenetic analyses of sweet potato and its immediate CWRs have tended to be stymied by limited taxonomic or genomic coverage, or more often than not, by both. In most cases, these studies have been able to recover the majority of tetraploid CWR accessions in a close phylogenetic relationship to hexaploid *I. batatas* and/or diploid *I. trifida* (Fig. 4), but not to provide any greater clarity on what exactly the nature of that relationship is. Analyses based on RFLPs for example, recovered *I. tabascana* (DLP2946) sister to a monophyletic hexaploid *I. batatas*, and K233 nested within diploid *I. trifida*, with the two respective clades sister to each other (Figure 4A) (Jarret et al. 1992). Later RAPD analysis instead recovered *I. tabascana, I. batatas* var. *apiculata* (DLP2907, DLP2908) and one other Mexican tetraploid (DLP2923) in a single clade, sister to a monophyletic hexaploid *I. batatas*, and with an Ecuadorean tetraploid (DLP5285) and two diploid *I. trifida* accessions in a grade below them (Figure 4B) (Jarret and Austin 1994).

Elsewhere, Rajapakse et al. (2004) also recovered *I. tabascana* (DLP2946) in a clade with *I. batatas* and *I. trifida* in their analysis of β-amylase gene sequences, though the exact topology differed between analyses of introns and exons (Fig. 4C). This was again interpreted as evidence for a hybrid origin for *I. tabascana* through a cross between hexaploid *I. batatas* and diploid *I. trifida* (Rajapakse et al. 2004; Srisuwan et al. 2006), which appeared to be supported by separate phylogenetic analysis of three artificially generated tetraploid or pentaploid hybrids between *I.*

batatas and other diploid CWRs, which had shown them to share a similar topology with *I. tabascana* (Fig. 4D), intermediate between *I. trifida* and *I. batatas* (Cao et al. 2014). The lack of sampling, poor resolution, low or non-existent support, and inconsistent topologies in all these studies make it difficult to draw any firm conclusions from their results however, other than that the majority of these tetraploid collections appeared to be closely related to both diploid *I. trifida* and hexaploid *I. batatas*. Two additional samples in these early molecular phylogenetic studies provide striking examples of the confusion potentially caused by poor or inconsistent identification of germplasm material.

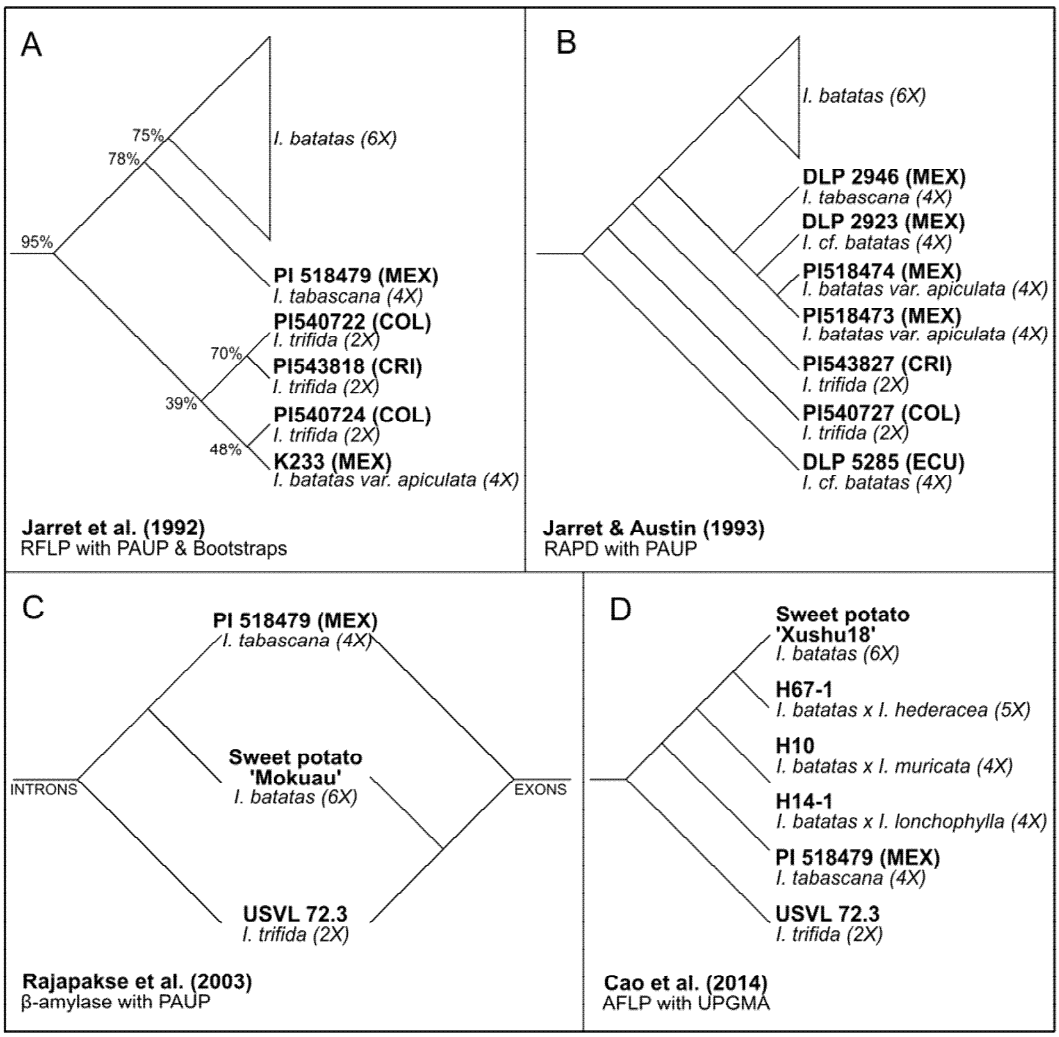


Figure 4: Representative topologies recovered in early phylogenetic analyses of molecular sequence data, based on **A:** Restriction Fragment Length Polymorphism, **B:** Random Amplification of polymorphic DNA, **C:** β-amylase; and **D:** Amplified Fragment Length Polymorphism**.** Showing the low resolution and inconsistencies in topology resulting from poor taxonomic and genomic sampling. Support values were only calculated by Jarret et al. (1992), with no support values provided by any other authors.

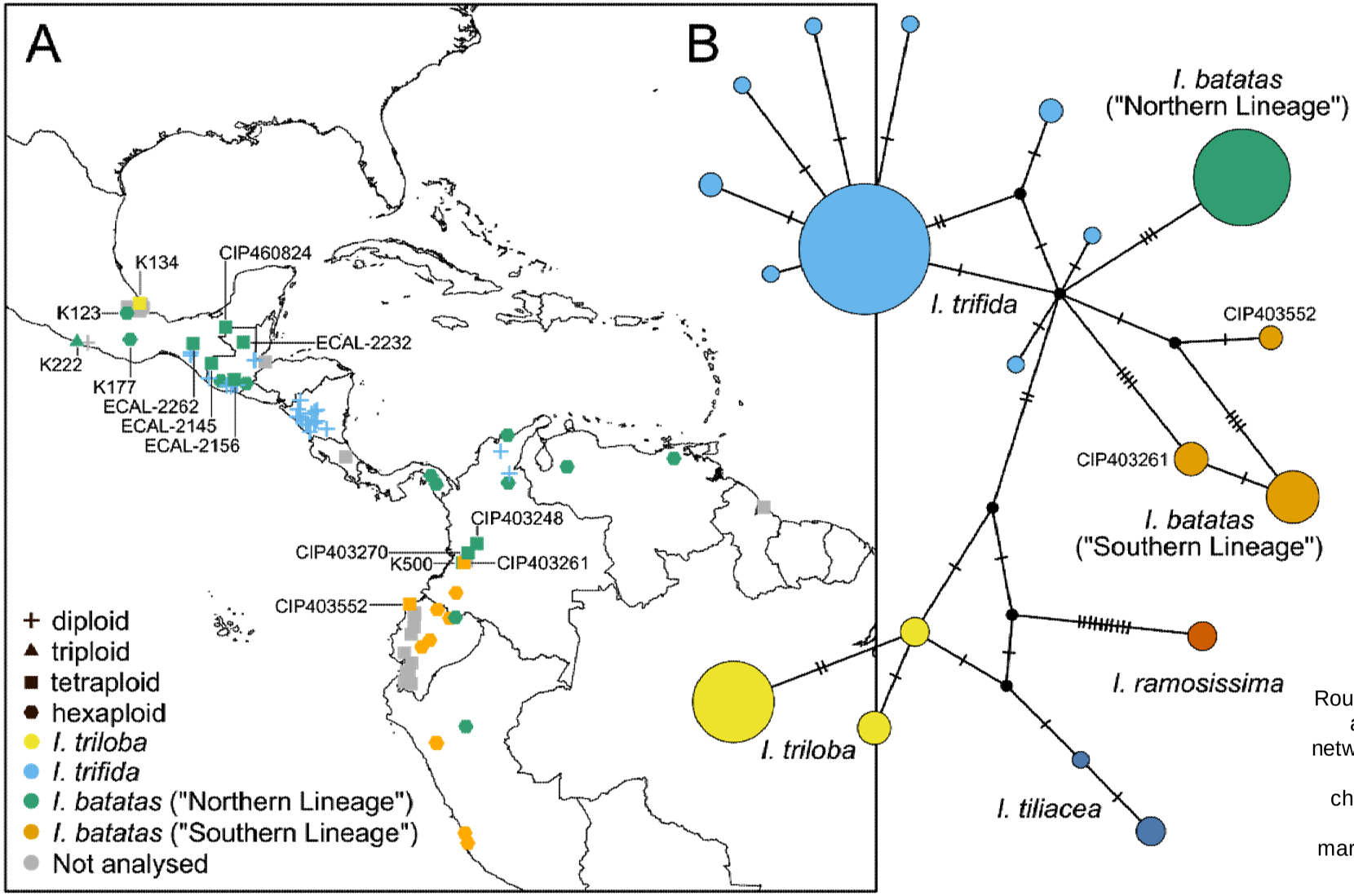


Figure 5: **A:** Collection locations of accessions sampled by Roullier et al. (2013); and **B:** haplotype network based on re-analysis of their chloroplast data for *rpl32-TrnL*(UAG) marker using TCA in PopART.

Rajapakse et al. (2004) for example, list one of their sampled specimens as *I. littoralis* under the accession number USVL79.5, which is recovered sister to the batatas-trifida-tabascana clade in the majority of their results. It is possible that this sample represents K233, which was originally misidentified as *I. littoralis* by (Muramatsu and Shiotani 1974), but no record appears to exist for it under the given accession number.

Separately, Huang and Sun (2000) list a specimen of unknown ploidy identified as *I. tiliacea* under Fermin De La Puente's collection number DLP2917, which they recovered sister to a clade containing *I. trifida* and *I. batatas* in analysis of ISSR chloroplast markers. A genesys-PGR search revealed this accession (CIP460070) to have been collected beside a field just south of Veracruz, close to the collection sites of other polyploid germplasm accessions such as K123, K134, K233, DLP2923, DLP2907, DLP2908 and CH67.50. In both cases the misidentification of the material, and failure to provide more detailed passport data such as collector, collection site, or relationship to other material have almost certainly served to perpetuate misconceptions about which taxa are likely to have been involved in the origin of *I. batatas*, while also obscuring potentially important additional CWR germplasm collections otherwise not discussed in the wider literature.

In the only molecular phylogenetic analysis to date to include most of the recorded wild-collected polyploid germplasm material (Fig. 5), Roullier et al. (2013) also found the majority of accessions to form a well-supported monophyletic group with diploid *I. trifida* and hexaploid *I. batatas.* In a statistical parsimony haplotype network analysis of the 1077bp *rpl32-trnL(UAG)* chloroplast region, and in line with earlier results (Roullier et al. 2011), they again recovered two chloroplast lineages within *I. batatas*, each more divergent from the other than from diploid *I. trifida* (Roullier et al. 2013). Based on the geographic origins of the majority of samples in

each lineage they designated these *I. batatas* chloroplast lineages as “Northern” (accessions predominantly from Mexico, Central America, northern Colombia and Venezuela) and “Southern” (southern Colombia, Ecuador and Peru) (Fig. 5A), and hypothesised that each might have arisen through separate auto-polyploidization events from distinct populations of the same (unidentified) wild progenitor (Roullier et al. 2013).

The majority of tetraploid germplasm accessions shared a chloroplast haplotype with the “Northern” *I. batatas* lineage, as did hexaploids K123 and K177, and triploid K222. Meanwhile Roullier et al. (2013) stated that five other tetraploid samples carried the “Southern” *I. batatas* haplotype. A complicating factor in interpreting these results is that they included multiple duplicate accessions of some collections (e.g. CIP403248_2 & CIP403248_3, or K500_1, K500_2, & K500_3) and in two cases different individuals of the same accession were recovered in different chloroplast lineages. This is the case for Ecuadorean tetraploid K300 (Northern Lineage: K300_6; Southern Lineage: K300_7) and Mexican tetraploid CIP403959[13] (Northern Lineage: 403959_2; Southern Lineage: 403959_ 1 & 403959_3). This polyphyly among duplicates of supposedly identical accessions suggests a labelling error either within the source institution (NARO/CIP) or during sequencing and analysis. That it is not mentioned in the results or discussions, suggests the authors were unaware of the collection history and origin of these crucial germplasm collections, and/or their own error. Meanwhile, re-analysis of the sequences deposited in genbank by Roullier et al. (2013) using the same methodology, showed that tetraploid Colombian accession CIP403261 also possessed the Southern Lineage haplotype (Fig. 5B), contrary to the results listed in the main text of their publication.

[13] No record of this accession could be found in online database genesis-pgr, or through CIP, but CIP403958 refers to another tetraploid accession: DLP2923, collected in Veracruz.

If we exclude the three collections with ambiguous results (K300, CIP403959, CIP403261), we find that all remaining tetraploid germplasm collections from Mexico, Guatemala and Colombia possess a “Northern” *I. batatas* chloroplast lineage, while the sole remaining Ecuadorian collection (CIP403552) has one similar, but not identical to that of the “Southern” *I. batatas* lineage (Fig. 5B).

Roullier et al.'s (2013) phylogenetic analysis of the nuclear ribosomal ITS region (701bp) for the same samples failed to produce a resolved topology, as did a NeighbourNet analysis of nuclear SSR microsatellites. A DAPC clustering analysis of the same nuclear SSR data however, did produce results somewhat similar to those for the chloroplast data, with most “Southern” lineage *I. batatas* (48 of 56) plus Ecuadorean tetraploid CIP403553 resolved within cluster K1, while most “Northern” chloroplast lineage *I. batatas* (50 of 83) fell within K2, or else in a separate cluster K3 (19 of 83) with around half (23 of 42) the other tetraploid accessions. The remaining tetraploids shared cluster K4 with diploid *I. trifida*, while a minority (5 of 83) of “Northern” chloroplast lineage *I. batatas* fell within “Southern” cluster K1, and vice versa for 8 “Southern” samples, which were found in clusters K2 or K3. Roullier et al. (2013) put this apparent discrepancy between results from nuclear and chloroplast data for a minority of specimens down to hypothesised secondary contact among the two lineages of domesticated sweet potatoes having led to introgression.

Muñoz-Rodríguez et al. (2018) later recovered a well-supported monophyletic *I. batatas* in their phylogenetic analysis of 605 low-copy nuclear regions however, which they concluded indicated that hexaploid sweet potato arose only once, with subsequent hybridization and backcrossing with diploid *I. trifida* resulting in the development of a second chloroplast lineage (Roullier et al.'s (2013) “Northern” lineage) through “chloroplast capture”.

The only tetraploid accession included by Muñoz-Rodríguez et al. (2018) was the probable hybrid (Srisuwan et al. 2006) *I. tabascana* (PI518479), which as in previous analyses (Fig. 4), was again recovered intermediate between diploid *I. trifida* and hexaploid *I. batatas* in nuclear phylogenies, and possessed a Lineage 2/"Southern" chloroplast haplotype.

If we re-assess Roullier et al.'s (2013) results based on Muñoz-Rodríguez et al.'s (2018) hypothesis for the origin of the two chloroplast lineages, the only tetraploid accession to share a chloroplast similar to the "ancestral" (Roullier et al.'s (2013) "Southern") chloroplast lineage of hexaploid *I. batatas* was Ecuadorian collection CIP403553. This would make it the most plausible candidate for the tetraploid progenitor of *I. batatas* under Muñoz-Rodríguez et al.'s (2018) hypothesis of a single origin for hexaploid *I. batatas* followed by subsequent chloroplast capture. Unpublished analysis at Oxford has shown that several other Ecuadorian collections also possess the same chloroplast haplotype (pers. comm. Munoz Rodriguez). The possession of a Lineage 2/"Northern" chloroplast haplotype by Mexican, Guatemalan and Colombian tetraploid germplasm accessions by contrast, would be consistent with these accessions having originated through hybridization between hexaploid *I. batatas* with a Lineage 2 chloroplast and a CWR, most likely *I. trifida*. Both Roullier et al.'s (2013) and Muñoz-Rodríguez et al.'s (2018) results do agree in confirming that *I. trifida* is a monophyletic and well-circumscribed species, known only from diploid material. It is also clear that the tetraploid CWR accessions (with the exception of K134, which was recovered with *I. triloba*) are not related to this or any other described species in the Batatas group, such as *I. littoralis*, *I. tiliacea* or *I. triloba*[14], as had been suggested in the past.

[14]*I. leucantha* was applied by Nishiyama et al. (1975) to a diploid collection from Acapulco, but is thought to refer to sporadic hybrids of *I. cordatotriloba* and *I. lacunosa*, and appears to have been frequently misapplied to collections of *I. triloba*, *I. cordatotriloba* and *I. australis* (Munoz et al., 2018).

These species were all recovered as monophyletic and more distantly related than to *I. trifida* and *I. batatas* (Muñoz-Rodríguez et al. 2018). The use of these names for many of the wild-collected polyploid plants resembling *I. batatas* appears to have stuck however, and continues to be incorrectly applied to polyploid germplasm material (see below), causing considerable ongoing confusion about the taxa and individual accessions that are most likely to be involved in the origin of sweet potato.

At the most basic level, many recent studies of sweet potato simply continue to use this inaccurate nomenclature in their discussion of sweet potato's possible origin, repeating Nishiyama's (1971) theory of an original hybridization event between "diploid *I. leucantha*" and "tetraploid *I. littoralis*" leading to a triploid and then "hexaploid *I. trifida*", which was finally domesticated to become *I. batatas* (e.g. Gao et al. 2011, 2020; Feng et al. 2018; Li et al. 2019). This is despite the fact that this theory is based on nothing more than outdated assumptions about chromosome behaviour at meiosis (Bomblies 2022), that the hexaploid *I. trifida* accession (K123) strongly resembles a feral sweet potato (Jones 1967), and that all the originally cited material have been shown to possess an *I. batatas* chloroplast haplotype (Roullier et al. 2013).

Worse, this inaccurate nomenclature is also repeated in new analyses using often unidentifiable accessions with no collection history or passport data that are simply designated "polyploid *I. trifida*" and treated as a homogenous taxon. One example is the five artificial tetraploids generated at CIP and used in cytological and palynological studies on the origin of sweet potato (Srisuwan et al. 2006, 2018) (see above).

Similarly, Feng et al. (2018) cited three tetraploid and one hexaploid accessions of "*I. trifida*" in their phylogenetic analysis of Restriction-site Associated DNA Sequencing (RAD-seq). The accession numbers given do not match those

from any other publication or institutional database, and their origin is listed as "Unkown". The results showed these accessions grouped closely with sweet potato, though the use of unrooted trees and lack of support values make it difficult to assess exact relationships. No other species from the Batatas group were included and all the included sweet potato accessions were modern cultivars from China, Japan, Africa, Australia and the USA. Nonetheless the authors went on to conclude that:

"These results provided new evidence that cultivated sweetpotato originated from *I. trifida* 6X, and that *I. trifida* 6X evolved from *I. trifida* 4X and *I. trifida* 2X. Therefore, using *I. trifida* 6X as the model plant of sweetpotato research should be more practical than using *I. trifida* 2X in the future."

Consistent with this assertion, these flawed results were then used as the basis of a study into storage root development in the "wild sweetpotato (*Ipomoea trifida*)" (Li et al. 2019).

In an even more striking example of the impact of this taxonomic confusion on sweet potato genomics research, Gao et al. (2011, 2020) have repeatedly used an Australian accession of *I. littoralis* (PI 573335) in their studies of sweet potato evolution, presumably on the basis of the initial misidentification of K233 as that species. Drawing useful conclusions from these studies is therefore all but impossible.

Meanwhile, an alternative, but equally confusing approach has been to treat all tetraploid germplasm collections as a tetraploid form of sweet potato, "*I. batatas* 4X", in spite of the considerable heterogeneity among these different accessions outlined above. Not only does treating all collections with potentially different origins and relationships to hexaploid *I. batatas* as the same taxon obscure their considerable diversity, it also risks drawing erroneous conclusions about the origin of sweet potato's hexaploid genome and which taxa were potentially involved.

In 2015 for example, two separate insertions of A*grobacterium* T-DNA were discovered in the genome of hexaploid *I. batatas*, representing an apparent example of natural horizontal gene transfer (HGT) arising from an ancient infection (Kyndt et al. 2015; Quispe-Huamanquispe et al. 2019). One of these insertions (IbT-DNA1) was found in all analysed accessions of hexaploid *I. batatas*, while the other (IbT-DNA2) was found in around 21% of them. No IbT-DNA1 was found by PCR amplification in any accessions of diploid *I. trifida*, but evidence of it was found in three tetraploid accessions: *I. batatas* var. *apiculata* (PI518474), a southern Colombian collection (CIP403270) and one from Ecuador (CIP460577)[15], all of which were identified as "*I. batatas* 4X".

Meanwhile, IbT-DNA2 was found in nine out of 28 accessions of diploid *I. trifida*, as well as eight "*I. batatas* 4X" accessions, all from Ecuador. No evidence of either insertion was found in *I. tabascana* (PI518479), two further Colombian accessions (CIP403248 & CIP403261), one from Mexico (CIP403969), and another from Ecuador (CIP460576), which were again all described as "*I. batatas* 4X".

While the discovery of these insertions was a major development in crop science and evolutionary biology, their utility as an effective or reliable marker for trying to reconstruct phylogenetic relationships and evolutionary history is less clear. This is partly because the possibility of false negatives when trying to amplify these regions is hard to exclude (Quispe-Huamanquispe et al. 2019), but also because similar insertions were also found in the more distant CWR *I. tenuissima* and two unidentified germplasm collections from Mexico (CIP460814) and the USA (CIP460250) (Kyndt et al. 2015; Quispe-Huamanquispe et al. 2019), implying either that this insertion entered a wild common ancestor of the Batatas group and has

[15] CIP403270 was collected by Fermin De La Puente near Cali in 1986, but no collection details could be found in the CIP online database or Genesys -PGR for CIP460577, which is also given the alternative accession number UME 32. The same applies to CIP460576 (UME 31) and CIP460578 (UME 33).

since been repeatedly lost in at least 8 other species, or else that it has introgressed into sweet potato and its wild relatives on multiple separate occasions. Either of these scenarios would fundamentally undermine the utility of these insertions as phylogenetic markers.

More importantly, treating tetraploid germplasm collections as a homogenous group makes interpreting these results even more difficult, since the presence or absence of these two insertions is not uniform across them. For example, the absence of IbT-DNA1 from any samples of diploid *I. trifida* and its presence in only a minority of tetraploid accessions may suggest that it originated in these tetraploids before being inherited by hexaploid *I. batatas*. But, if the tetraploid samples that possess IbT-DNA1 are in fact modern hybrids between *I. batatas* and *I. trifida*, then they would still be expected to possess the insertion, which they could have inherited from their *I. batatas* parent.

Meanwhile IbT-DNA2 is distributed in at least some accessions of diploid *I. trifida*, suggesting this is where it may have originated. It is also present in the majority of Ecuadorian tetraploids, which suggests that they could have passed it to hexaploid *I. batatas*, potentially supporting their role as the tetraploid progenitor of sweet potato, but the same caveat applies. IbT-DNA2 is also not present in the majority of hexaploid *I. batatas* specimens, which potentially undermines this theory, or else potentially supports multiple origins for sweet potato.

What is clear is that with a few exceptions, there does appear to be a similar pattern to before, of considerable heterogeneity across tetraploid collections from Mexico and Central America (with or without IbT-DNA1), and a more homogenous grouping from Ecuador, where almost all accessions possess IbT-DNA2 but not IbT-DNA1. A further point of interest is the uneven distribution of IbT-DNA2 in hexaploid *I. batatas*. This is another example of apparent inconsistencies in the genomic structure of different sweet potato accessions.

No analysis has compared the distribution of IbT-DNA2 in accessions of hexaploid *I. batatas* with the possession of either of the chloroplast lineages identified by Roullier et al. (2013) and Muñoz-Rodríguez et al. (2018) for example, or the differing numbers of copies of 18S uncovered by (Srisuwan et al. 2006). Different sampling was used in each case, making such a comparison impossible, but future studies might consider assessing whether these different patterns of heterogeneity in the hexaploid genome of *I. batatas* are correlated.

Use in Crop Improvement Breeding

The use of CWRs in sweet potato crop improvement breeding programmes has been surprisingly limited to date. Orjeda et al. (1991) listed two possible reasons for this as (i) the lack of significantly thickened roots in most CWRs, and (ii) the challenge created by the difference in ploidy between hexaploid *I. batatas* and its predominantly diploid CWRs, which requires that a tetraploid intermediary be used in controlled crossing. On this basis, they attempted over 28,000 pollinations between hexaploid *I. batatas* and diploid *I. trifida* in an effort to establish a tetraploid "initiator" population capable of producing storage roots that could then be used to evaluate potentially beneficial traits and incorporate them into hexaploid sweet potato varieties (Orjeda et al. 1991). Eight tetraploid progeny produced yields of 0.8-1.5kg, and 70% of them produced a higher yield than their cultivated parent.

Komaki and Katayama (1999) had similar success crossing *I. batatas* and *I. trifida*, with their tetraploid progeny producing yields of up to 1.8kg per plant. Both sets of authors highlighted the benefits of these tetraploid hybrids, both as potential intermediaries via which to incorporate wild diversity in hexaploid cultivars, but also as potential cultivars in their own right, with a simpler tetrasomic inheritance pattern. No such tetraploid cultivars appear to have been developed however, and little if any further use appears to have been made of these promising tetraploid hybrids.

In fact, the only commercially successful cultivars derived from wild-collected germplasm material have been developed using Mexican hexaploid accession, K123 (Iwanaga 1988). Nishiyama et al. (1961a) found that this accession showed good resistance to both root knot nematodes (*Meloidogyne incognita*) and root legion nematodes (*Pratylenchus spp.*) and black rot disease (*Ceratocystis fimbriata*), and its later incorporation in a Japanese breeding program gave rise to the *Kyukei 15-2120* cultivar (Oracion et al. 1990), as well as *Minamiyutaka* (Kobayashi and Miyazaki 1975; Ahmad et al. 2021), which became the second most important variety in Japan, cultivated in over 5,000ha (Iwanaga 1988). K123 has apparently also been used successfully as a parent in Chinese breeding programmes, again leading to increased yield and high dry matter content (Iwanaga 1988; Xie et al. 2003).

As Iwanaga (1988) points out, if K123 is in fact a feral escape from cultivation (Jones 1967a), then this is not in fact a true example of incorporating CWRs in crop improvement breeding, and instead might simply demonstrate the benefits of widening genetic diversity in an asexually propagated crop. K123 could also be an example of a hexaploid hybrid between cultivated sweet potato and diploid *I. trifida* however (as described by Orjeda et al. (1991)), in which case it could be acting as a natural intermediary for the introgression of wild material. This could also be the case for tetraploid K233, which was hybridized with sweet potato accessions in China, again producing high-yield and high dry-matter progeny, with resistance to black rot (Iwanaga 1988).

Similar results were also obtained for crosses from four further tetraploid accessions labelled *I. trifida* but without accession numbers (Iwanaga 1988), 10% of which produced yields similar to their cultivated parent, with dry matter content of up to 45% (a record for sweet potato), protein content of 3.1g/100g fresh weight, and apparent resistance to one of the most important sweet potato pests, weevils

(*Cyclas* spp.) (Iwanaga 1988). Histological examination shows that these crosses from potential CWR accessions possessed significant differences in parenchymatous cell division compared with other sweet potato cultivars, suggesting they may be introducing changes in development that are responsible for the increased dry matter content (Iwanaga 1988).

The five tetraploid specimens from CIP listed by Srisuwan et al. (2006, 2018) as "*I. trifida* 4X", also appear to have been derived from crosses between hexaploid *I. batatas* and wild diploid relatives as part of a crop improvement breeding program (pers. comm. CIP/Rosario Falcon Meza), but no further information on their exact parentage or use in that program could be obtained. Xie et al. (2003) also successfully crossed an un-accessioned tetraploid labelled *I. littoralis* and tetraploid and hexaploid specimens labelled *I. trifida* with sweet potato cultivars, some progeny of which were used to develop the new cultivars *Suyu 303* and *Yusu 297*.

Based on the differences in morphology and collection site of the material discussed in this chapter, several other accessions with potentially promising characteristics have not yet been incorporated in breeding trials. Accession DLP2908 (*I. batatas* var. *apiculata*) for example was collected from coastal sand dunes around the Mexican Gulf port of Veracruz, suggesting it may have increased drought- and salt-tolerance.

The suspected hybrid *I. tabascana* (PI518479) on the other hand was found growing in marshy conditions in southern Mexico, and several collections from Ecuador (CH81.2) and Colombia (K500, CH73.1, DLP1818, DLP 1843, DLP1864) have been collected at elevations of up to 2000m. The confusion around the taxonomic identity of these specimens appears to have obscured some of the details of this diversity, hindering the search for potentially beneficial traits in CWRs.

CONCLUSION:

The results of the comprehensive review carried out in this chapter reveal considerable heterogeneity among tetraploid and other polyploid collections of *Ipomoea* resembling wild sweet potatoes. The confusion surrounding the taxonomy and exact phylogenetic relationships of these specimens to cultivated hexaploid *I. batatas* has served to obscure details of much of this diversity, and to impede efforts to understand the origin of sweet potato and breed new cultivars incorporating diversity from wild species.

Specifically, the continued application of the names *I. leucantha*, *I. littoralis*, and *I. gracilis* to this material should be avoided, given that there is ample phylogenetic and taxonomic evidence that these names apply to distinct taxa with no involvement in the origin of hexaploid *I. batatas*. Similarly, *I. trifida* should be recognized as a purely diploid species, distinct from polyploid collections, which have instead been shown to share chloroplast haplotypes with the either of the two *I. batatas* lineages. Simply treating all tetraploid material as a “tetraploid race” of *I. batatas* is also unsatisfactory however, given the aforementioned heterogeneity of the material, and the lack of certainty around its exact origin resulting from the difficulty of distinguishing potential tetraploid progenitors of hexaploid *I. batatas* from likely hybrids between *I. batatas* and *I. trifida*.

Given this confusion, it is difficult to draw satisfactory conclusions about important aspects of sweet potato biology, such as the identity of its tetraploid progenitor, the origin of *Agrobacterium* insertions or independent chloroplast lineages, the evolution and development of thickened storage roots, and the utility of sweet potato CWRs in crop improvement breeding programmes. Particularly surprising is the failure to characterize and attempt to incorporate potentially promising traits from accessions such as coastal specialist *I. batatas var. apiculata*, or high-elevation collections from Ecuador and Colombia.

This review has also highlighted two distinct areas of past tetraploid collection, the first in Mexico and Guatemala, and the second in western Ecuador and southern Colombia. Sporadic mentions of tetraploid material collected elsewhere were difficult to verify, as the original material was either not given accession numbers, or else has since been lost.

The Mexican and Guatemalan material appears to be fairly heterogeneous, but with a generally sprawling habit, variously lobed leaves, and few-flowered inflorescences. This variability may be indicative of a potential hybrid origin, for which there is sporadic evidence in the literature, such as Guatemalan collections resembling artificial hybrids in terms of their morphology and chromosome pairing (Oracion et al. 1990), the karyology and phylogenetic position of *I. tabascana* (Rajapakse et al. 2004; Srisuwan et al. 2006; Cao et al. 2014; Muñoz-Rodríguez et al. 2018), the morphology and failure to produce seeds in CH67.50 (Bohac et al. 1993), and the possession by all these specimens of a chloroplast haplotype identical to that of the potentially derived sweet potato Lineage 2 (Roullier et al. 2013). All were collected from highly disturbed environments where cultivation is common and within the geographic range of diploid *I. trifida*, which has been shown to hybridize relatively easily with hexaploid *I. batatas* (Orjeda et al. 1991).

By contrast, the Ecuadorian and southern Colombian collections were made outside the geographic range of *I. trifida*, or any other Batatas group species except *I. triloba*, which is recorded from the Galapagos islands and occasionally the Ecuadorian coast (Wood et al., 2020). Though also sometimes collected in disturbed areas of cultivation (Austin et al. 1993), Ecuadorian collections have also been collected in primary and secondary tropical forest habitat. The Ecuadorian tetraploids have a chloroplast haplotype similar but not identical to the hypothesised ancestral lineage of sweet potato, and are recovered sister to it in phylogenetic analysis of the *rpl32-trnL*(UAG) barcode marker (Roullier et al. 2013).

The morphology of these specimens is also more consistently that of climbing vines with largely unlobed leaves, and many-flowered inflorescences with short, blunt calyxes. Austin (1977) initially suggested that along with Mexican material, these collections were also likely hybrids, though he provided no real evidence for this hypothesis and he later appears to have changed his mind, going on to describe them as a distinctive wild form of *I. batatas* (Austin et al. 1993). (Martin et al. 1974) also suggested that multivalent formation showed Colombian and Ecuadorian collections were likely autotetraploid in origin, though this line of evidence is somewhat unreliable. The evidence therefore points to these specimens potentially representing the most promising avenue for progress in identifying the tetraploid progenitor of hexaploid *I. batatas*, but it is by no means definitive, and the possibility exists that either group, both or neither may have had some role in the origin of cultivated sweet potato.

Beyond insights into its potential tetraploid CWRs, multiple lines of evidence suggest that hexaploid sweet potato itself possesses considerable variability in terms of its genomic structure. These differences include varying numbers of copies of 18S and CMA+ bands (Srisuwan et al. 2006), two independent chloroplast lineages (Roullier et al. 2011, 2013; Muñoz-Rodríguez et al. 2018), possession of one or both *Agrobacterium* insertions (Kyndt et al. 2015; Quispe-Huamanquispe et al. 2019) and instances of abnormal chromosome behaviour in some cultivars but not others (Ting and Kehr 1953), which all point to a high degree of variability in the hexaploid genome of sweet potato. This variability may be the result of uneven diploidization, the process by which relatively unstable polyploid genomes undergo re-arrangement to restore bivalent chromosome pairing and disomic inheritance (Li et al. 2021) but it could also point to a more complex evolutionary history for *I. batatas*, such as multiple separate transitions to hexaploidy by different routes

(Roullier et al. 2013). These differences therefore warrant further investigation, which will require the definitive identification of its likely tetraploid progenitor.

Specific questions requiring an answer include whether all sweet potato accessions share the subgenomic structure documented in the Taizhong cultivar by (Yang et al. 2017); whether *I. tabascana* is indeed a recent hybrid between hexaploid *I. batatas* and diploid *I. trifida*, as well as whether other tetraploid material from the same area might also be; and finally, whether Ecuadorian accessions that share a chloroplast haplotype with the hypothesized ancestral lineage of *I. batatas* might indeed be its tetraploid progenitor. Resolving these questions will require an integrative methodology that combines High-throughput sequencing of carefully selected samples with genomic analysis and consideration of the collection history and combined evidence outlined in this chapter.

TABLE 1: COLLECTION & DETERMINATION HISOTORY OF POLYPLOID SWEET POTATO CWR GERMPLASM COLLECTIONS									
Accession numbers			Ploidy	Collection Details					Past Determinations
CIP	**USDA**	**NARO**		**Collector**	**Number**	**Country**	**Year**		
		K123	6X	Nishiyama		MEX	1955		***I. trifida*** (Nishiyama et al., 1961a, 1961b; Teramura & Nishiyama, 1962; Kobayashi, 1978, 1984)*;* ***I. batatas*** (Jones 1967b)
		K134	4X	Nishiyama		MEX	1955		***I. gracilis*** (Nishiyama 1959; Nishiyama et al. 1961a)
		K233	4X	Muramatsu		MEX	1959		***I. gracilis*** (Ting et al., 1957; Nishiyama & Teramura, 1962; Jones, 1970; Martin & Jones, 1972)*;* ***I. littoralis*** (Nishiyama, 1971; Kobayashi & Miyazaki, 1975)*;* ***I. batatas*** (Austin, 1988; Jones, 1990; Mcdonald & Austin, 1990)*;* ***I. trifida*** (Kobayashi, 1984; Shiotani & Kawase, 1987, 1989; Oracion et al., 1990; Ozias-Akins & Jarret, 1994)*;* ***I. batatas var. apiculata*** (Austin 1977; Jarret et al. 1992)
		K222	3X	Kobayashi		MEX	1960		***I. trifida*** (Nishiyama, 1971; Kobayashi & Miyazaki, 1975)*;* ***I. batatas*** (Bohac et al. 1993)

Accession numbers			Ploidy	Collection Details					Past Determinations
CIP	**USDA**	**NARO**		**Collector**	**Number**	**Country**	**Year**		
		K177	6X	Muramatsu		MEX	1962		***I. trifida*** (Nishiyama 1961; Kobayashi 1978)
	CH67.62 †		4X	Martin		Guyana			***I. tiliacea*** (Martin et al. 1974)*;* ***I. batatas*** (Bohac et al. 1993)
	CH68.6 †		4X	Martin		CRI			***I. tiliacea*** (Martin et al. 1974)*;* ***I. batatas*** (Bohac et al. 1993)
		K270	4X			Puerto Rico			***I. tiliacea*** (Nishiyama, 1971; Kobayashi & Miyazaki, 1975; Roullier et al., 2013)
	CH71.3; PI355830	K300	4X	Rick	SAM2614	ECU	1970		***I. batatas*** (Martin et al. 1974; Bohac et al. 1993)*;* ***I. trifida*** (Kobayashi 1984; Oracion et al. 1990)*;* ***I. batatas x I. trifida*** (Austin 1977)*;*
	CH73.1 ‡		4X	Jones		COL	1972		***I. batatas*** (Martin et al. 1974)*;* ***I. batatas x I. trifida*** (Austin 1977)*;* ***I. trifida*** (Oracion et al. 1990)*;*

Accession numbers			Ploidy	Collection Details					Past Determinations
CIP	**USDA**	**NARO**		**Collector**	**Number**	**Country**	**Year**		
	CH73.2 †		4X	Martin		MEX			***I. batatas*** (Bohac et al. 1993)
		ECAL-2145	4X	Muramatsu & Shiotani		GTM	1973		***I. trifida*** (Muramatsu & Shiotani, 1974)*;* ***I. sp.***(Roullier et al. 2013)
		ECAL-2156	4X	Muramatsu & Shiotani		GTM	1973		***I. trifida*** (Muramatsu & Shiotani, 1974)*;* ***I. sp.*** (Roullier et al. 2013)
		ECAL-2192	4X	Muramatsu & Shiotani		GTM	1973		***I. trifida*** (Muramatsu & Shiotani, 1974)*;* ***I. sp.***(Roullier et al. 2013)
		ECAL-2232	4X	Muramatsu & Shiotani		GTM	1973		***I. trifida*** (Muramatsu & Shiotani, 1974; Kobayashi, 1984; Oracion et al., 1990)*;* ***I. sp.*** (Roullier et al. 2013)
		K400; ECAL-2262	4X	Muramatsu & Shiotani		MEX	1973		***I. trifida*** (Muramatsu & Shiotani, 1974)*;* ***I. sp.*** (Kobayashi & Miyazaki, 1975; Roullier et al., 2013)
	CH80.3	K500	4X	Kobayashi		COL	1976		***I. trifida*** (Kobayashi, 1984; Ozias-Akins & Jarret, 1994)*;* ***I. batatas*** (Bohac et al. 1993)

Accession numbers			Ploidy	Collection Details					Past Determinations
CIP	**USDA**	**NARO**		**Collector**	**Number**	**Country**	**Year**		
	CH81.1 †		4X	Rick	SAM5306	ECU			***I. batatas*** (Bohac et al. 1993)
	CH81.2		4X	Rick	SAM5316	ECU	1980		***I. batatas*** (Jones, 1990; Bohac et al., 1993; Ozias-Akins & Jarret, 1994)
	CH87.2 †		4X			ECU			***I. trifida*** (Ozias-Akins & Jarret, 1994)
	NA †		4X	McJunkin	1378	VEN			***I. batatas*** (Bohac et al. 1993)
403248			4X	De La Puente	1818	COL	1986		***I. sp.*** (Roullier et al. 2013)*;* ***I. batatas*** (Quispe-Huamanquispe et al. 2019)
403261			4X	De La Puente	1843	COL	1986		***I. sp.*** (Roullier et al. 2013)*;* ***I. batatas*** (Quispe-Huamanquispe et al. 2019)
403270			4X	De La Puente	1874	COL	1986		***I. batatas*** (Quispe-Huamanquispe et al. 2019; Yan et al. 2021)
460191	PI518473 †		4X	De La Puente	2907	MEX	1987		***I. batatas*** (Ozias-Akins & Jarret, 1994)*;* ***I. batatas var. apiculata*** (Jarret & Austin, 1994)

Accession numbers			Ploidy	Collection Details					Past Determinations
CIP	**USDA**	**NARO**		**Collector**	**Number**	**Country**	**Year**		
403953	PI518474		4X	De La Puente	2908	MEX	1987		***I. batatas*** (Ozias-Akins & Jarret, 1994; Yan et al., 2021)*;* ***I. batatas var. apiculata*** (Jarret & Austin, 1994; Quispe-Huamanquispe et al., 2019)
403955 †	PI518476†		4X	De La Puente	2915	MEX	1987		***I. batatas*** (Ozias-Akins & Jarret, 1994)
403958	PI518478		4X	De La Puente	2923	MEX	1987		***I. batatas*** (Jarret & Austin, 1994; Ozias-Akins & Jarret, 1994)
460824	PI518479		4X	De La Puente	2946	MEX	1990		***I. tabascana*** (Mcdonald & Austin, 1990; Ozias-Akins & Jarret, 1994)*;* ***I. batatas x I. trifida*** (Srisuwan et al. 2006)
403969			4X	De La Puente	2949	MEX	1990		***I. batatas*** (Quispe-Huamanquispe et al. 2019)
460576			4X	UME	31	ECU	1990		***I. batatas*** (Quispe-Huamanquispe et al. 2019)
460577			4X	UME	32	ECU	1990		***I. batatas*** (Quispe-Huamanquispe et al. 2019)

Accession numbers			Ploidy	Collection Details					Past Determinations
CIP	**USDA**	**NARO**		**Collector**	**Number**	**Country**	**Year**		
460578			4X	UME	33	ECU	1990		***I. batatas*** (Quispe-Huamanquispe et al. 2019)
403551	PI561246		4X	De La Puente /Austin	5282/7801	ECU	1992		***I. batatas*** (Austin et al. 1993; Quispe-Huamanquispe et al. 2019)
403552	PI561247		4X	De La Puente /Austin	5283/7802	ECU	1992		***I. batatas*** (Austin et al., 1993; Ozias-Akins & Jarret, 1994; Quispe-Huamanquispe et al., 2019)*;* ***I. sp.*** (Roullier et al. 2013)
403553	PI561248		4X	De La Puente /Austin	5284/7803	ECU	1992		***I. batatas*** (Austin et al. 1993; Quispe-Huamanquispe et al. 2019)
	PI561249		4X	De La Puente /Austin	5285/7804	ECU	1992		***I. batatas*** (Austin et al., 1993; Jarret & Austin, 1994)
	PI561250		4X	De La Puente /Austin	5287/7806	ECU	1992		***I. batatas*** (Austin et al. 1993)
	PI561251		4X	De La Puente /Austin	5290/7809	ECU	1992		***I. batatas*** (Austin et al. 1993)
	PI561252		4X	De La Puente /Austin	5291/7810	ECU	1992		***I. batatas*** (Austin et al. 1993)

Accession numbers			Ploidy	Collection Details					Past Determinations
CIP	**USDA**	**NARO**		**Collector**	**Number**	**Country**	**Year**		
	PI561253		4X	De La Puente /Austin	5292/7811	ECU	1992		***I. batatas*** (Austin et al. 1993)
	PI561254		4X	De La Puente /Austin	5293/7812	ECU	1992		***I. batatas*** (Austin et al. 1993)
403560	PI561255		4X	De La Puente /Austin	5296/7815	ECU	1992		***I. batatas*** (Austin et al. 1993; Quispe-Huamanquispe et al. 2019)
	PI561257		4X	De La Puente /Austin	5297/7816	ECU	1992		***I. batatas*** (Austin et al. 1993)
	PI561258		4X	De La Puente /Austin	5298/7817	ECU	1992		***I. batatas*** (Austin et al. 1993; Ozias-Akins and Jarret 1994; Quispe-Huamanquispe et al. 2019)
	PI561259		4X	De La Puente /Austin	5301/7820	ECU	1992		***I. batatas*** (Austin et al. 1993)
	PI561260 †		4X	De La Puente /Austin	5302/7821	ECU	1992		***I. batatas*** (Austin et al. 1993)
	PI561261 †		4X	De La Puente /Austin	5304/7823	ECU	1992		***I. batatas*** (Austin et al. 1993; Quispe-Huamanquispe et al. 2019)
694436 †			4X						***I. trifida*** (Srisuwan et al. 2006)

Accession numbers			Ploidy	Collection Details					Past Determinations
CIP	**USDA**	**NARO**		**Collector**	**Number**	**Country**	**Year**		
695141 †			4X						***I. trifida*** (Srisuwan et al. 2006, 2018)*;* ***I. batatas*** (Yan et al. 2021)
695143 †			4X						***I. trifida*** (Srisuwan et al. 2006)
695150 †			4X						***I. batatas*** (Yan et al. 2021)
695152 †			4X						***I. trifida*** (Srisuwan et al. 2006)
696146 †			4X						***I. trifida*** (Srisuwan et al. 2006)

TABLE 1: Accession numbers, Collection details, and past determinations of wild-collected polyploid germplasm material previously associated with *I. batatas* († unavailable on request)

REFERENCES:

Austin, D. F. (1977). Hybrid polyploids in *Ipomoea* section batatas. J. Hered. 62, 259–260.

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

Austin, D. F. (1982). “Convolulaceae,” in Flora of Ecuador, eds. G. Harling and B. Sparre (Stockholm: Swedish Research Councils), 30–69.

Austin, D. F. (1988). The taxonomy, evolution and genetic diversity of sweetpotatoes and related wild species. Explor. maintenance, Util. sweet potato Genet. Resour., 27–59.

Austin, D. F., Jarret, R. L., Tapia, C., and de la Puente, F. (1993). Collecting tetraploid *I. batatas* (L.) Lam. in Ecuador. FAO/IBPGR Plant Genet. Resour. Newsl. 91/92, 33–35.

Bagavathiannan, M. V., and Van Acker, R. C. (2008). Crop ferality: Implications for novel trait confinement. Agric. Ecosyst. Environ. 127, 1–6. doi:10.1016/j.agee.2008.03.009.

Bohac, J. R., Austin, D. F., and Jones, A. (1993). Discovery of Wild Tetraploid Sweetpotatoes. Econ. Bot. 47, 193–201.

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., et al. (2016). Global conservation priorities for crop wild relatives. Nat. Plants 2, 1–6. doi:10.1038/nplants.2016.22.

Cevallos-Casals, B. A., and Cisneros-Zevallos, L. (2003). Stoichiometric and kinetic studies of phenolic antioxidants from Andean purple corn and red-fleshed sweetpotato. J. Agric. Food Chem. 51, 3313–3319. doi:10.1021/jf034109c.

Contreras, J., de la Puente, F., Austin, D. F., and Diaz, J. (1994). Biodiversity of sweetpotato (*Ipomoea batatas*, Convolvulaceae) in Southern Mexico. Econ. Bot. 49, 286–296. Available at: https://www.jstor.org/stable/4255744 Accessed:

Cornejo, O. E., Yee, M.-C., Dominguez, V., Andrews, M., Sockell, A., Strandberg, E., et al. (2018). Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. Commun. Biol. 1, 1–12. doi:10.1038/s42003-018-0168-6.

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, 1070–1082. doi:10.2135/cropsci2016.10.0885.

Dodson, C. H., and Gentry, A. H. (1978). FLORA OF THE RIO PALENQUE SCIENCE CENTER: Los Rios Province ECUADOR. Selbyana 4, 301–302.

Dodson, C. H., Gentry, A. H., and Valverde, F. M. (1985). La Flora de Jauneche, Los Rios, Ecuador. Quito: Banco Central del Ecuador.

Fang, L., Gong, H., Hu, Y., Liu, C., Zhou, B., Huang, T., et al. (2017). Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons, in Genome Biology (Genome Biology), 1–13. doi:10.1186/s13059-017-1167-5.

FAO (1998). The State of the World ’s Plant Genetic Resources. Rome, Italy: Food and Agriculture Organization of the United Nations Available at: www.tcrecord.org.

FAO (2010). Second Report on the State of the World’s Plant Genetic Resources for Food and Agriculture. Rome, Italy: Commission on Genetic Resources for Food and Agriculture; Food and Agriculture Organization of the United Nations Available at: http://www.fao.org/docrep/013/i1500e/i1500e.pdf.

Flowers, J. M., Hazzouri, K. M., Gros-Balthazard, M., Mo, Z., Koutroumpa, K., Perrakis, A., et al. (2019). Cross-species hybridization and the origin of North African date palms, in Proceedings of the National Academy of Sciences, 1651–1658. doi:10.1073/pnas.1817453116.

Gao, M., Ashu, G. M., Stewart, L., Akwe, W. A., Njiti, V., and Barnes, S. (2011). Wx intron variations support an allohexaploid origin of the sweetpotato [*Ipomoea batatas* (L.) Lam]. Euphytica 177, 111–133. doi:10.1007/s10681-010-0275-z.

Gaut, B. S., Seymour, D. K., Liu, Q., and Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. Nat. Plants 4, 512–520. doi:10.1038/s41477-018-0210-1.

Gepts, P. (2004). Crop Domestication as a long-term selection experiment. Plant Breed. Rev. 24, 1–44. doi:0-471-46892-4.

Girard, A. W., Grant, F., Watkinson, M., Okuku, H. S., Wanjala, R., Cole, D., et al. (2017). Promotion of Orange-Fleshed Sweet Potato Increased Vitamin A Intakes and Reduced the Odds of Low Retinol-Binding Protein among Postpartum Kenyan Women. J. Nutr. 147, 955–963. doi:10.3945/jn.116.236406.

Gros-Balthazard, M., Galimberti, M., Kousathanas, A., Newton, C., Ivorra, S., Paradis, L., et al. (2017). The Discovery of Wild Date Palms in Oman Reveals a Complex Domestication History Involving Centers in the Middle East and Africa, in Current Biology, 2211-2218.e8. doi:10.1016/j.cub.2017.06.045.

Gros-Balthazard, M., Hazzouri, K. M., and Flowers, J. M. (2018). Genomic insights into date palm origins. Genes (Basel). 9, 1–14. doi:10.3390/genes9100502.

Gross, B. L., and Olsen, K. M. (2010). Genetic perspectives on crop domestication. Trends Plant Sci. 15, 529–537. doi:10.1016/j.tplants.2010.05.008.

Guerra-García, A., and Piñero, D. (2017). Current approaches and methods in plant domestication studies, in Botanical Sciences, 345–362. doi:10.17129/botsci.1209.

Gutaker, R. M., Weiß, C. L., Ellis, D., Anglin, N. L., Knapp, S., Luis Fernández-Alonso, J., et al. (2019). The origins and adaptation of European potatoes reconstructed from historical genomes. Nat. Ecol. Evol. 3. doi:10.1038/s41559-019-0921-3.

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. Euphytica 156, 1–13. doi:10.1007/s10681-007-9363-0.

Huang, X., Kurata, N., Wei, X., Wang, Z. X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. Nature 490, 497–501. doi:10.1038/nature11532.

Isobe, S., Shirasawa, K., and Hirakawa, H. (2017). Challenges to genome sequence dissection in sweetpotato. Breed. Sci. 67, 35–40. doi:10.1270/jsbbs.16186.

Isobe, S., Shirasawa, K., and Hirakawa, H. (2019). Current status in whole genome sequencing and analysis of *Ipomoea*. Plant Cell Rep. doi:10.1007/s00299-019-02464-4.

Iwanaga, M. (1988). “Use of wild germplasm for sweet potato breeding,” in Exploration, maintenance, and utilization of sweet potato genetic resources, ed. P. Gregory (Lima. Peru: CIP), 199–210.

Jarret, R. L., and Austin, D. F. (1994). Genetic diversity and systematic relationships in sweetpotato (*Ipomoea batatas* (L.) Lam.) and related species as revealed by RAPD analysis. Genet. Resour. Crop Evol. 41, 165–173. doi:10.1007/BF00051633.

Jarret, R. L., Gawel, N., and Whittemore, A. T. (1992). Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. J. Am. Soc. Hortic. Sci. 117, 633–637.

Jones, A. (1965). Cytological observations and fertility measurements of sweet potato (*Ipomoea batatas* (L.) Lam.). Proc. Am. Soc. Hortic. Sci. 86, 527–537.

Jones, A. (1967). Should Nishiyama's K123 (*Ipomoea trifida*) Be Designated *I. batatas*? Econ. Bot. 21, 163–166.

Jones, A. (1970). Asynapsis in *Ipomoea gracilis.* J. Hered. 61, 151–152. doi:10.1093/oxfordjournals.jhered.a108066.

Jones, A. (1990). Unreduced Pollen in a Wild Tetraploid Relative of Sweetpotato. J. Amer. Soc. Hort. Sci. 115, 512–516.

Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., et al. (2014). "Increasing homogeneity in global food supplies and the implications for food security," in Proceedings of the National Academy of Sciences, 4001–4006. doi:10.1073/pnas.1313490111.

Khoury, C. K., Heider, B., Castañeda-Álvarez, N. P., Achicanoy, H. A., Sosa, C. C., Miller, R. E., et al. (2015). Distributions, ex situ conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [ipomoea batatas (L.) lam., I. series batatas]. Front. Plant Sci. 6, 1–14. doi:10.3389/fpls.2015.00251.

Kilian, B., Dempewolf, H., Guarino, L., Werner, P., Coyne, C., and Warburton, M. L. (2021). Crop Science special issue: Adapting agriculture to climate change: A walk on the wild side. Crop Sci. 61, 32–36. doi:10.1002/csc2.20418.

Kobayashi, M. (1978). Sweetpotato breeding method using wild relatives in Japan. Trop. Agr. Res. 11, 1–8.

Kobayashi, M. (1984). "The Ipomoea trifida Complex Closely Related to Sweet Potato," in Proceedings, Sixth Symposium of the International Society for Tropical Root Crops, hosted by CIP in Lima, Peru, 561–568.

Kobayashi, M., and Miyazaki, T. (1975). Sweet Potato Breeding Using Wild Related Species.

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., et al. (2014). Current perspectives and the future of domestication studies. Proc. Natl. Acad. Sci. U. S. A. 111, 6139–6146. doi:10.1073/pnas.1323964111.

Laurie, S., Faber, M., Adebola, P., and Belete, A. (2015). “Biofortification of sweet potato for food and nutrition security in South Africa,” in FRIN (Elsevier Ltd), 962–970. doi:10.1016/j.foodres.2015.06.001.

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., et al. (2020). Pan-Genome of Wild and Cultivated Soybeans. Cell 0, 1–15. doi:10.1016/j.cell.2020.05.023.

Loebenstein, G., and Thottappilly, G. (2009). The Sweetpotato. Springer Netherlands doi:10.1007/978-1-4020-9475-0.

Lorant, A., Pedersen, S., Holst, I., Hufford, M. B., Winter, K., Piperno, D., et al. (2017). The potential role of genetic assimilation during maize domestication. PLoS One 12, 0–9. doi:10.1371/journal.pone.0184202.

Magoon, M. L., Krishnan, R., and Vijaya Bai, K. (1970). Cytological evidence on the origin of sweet potato. Theor. Appl. Genet. 40, 360–366. doi:10.1007/BF00285415.

Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., The International Wheat Genome Sequencing Consortium, et al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. Science (80). 345, 1285–1288. doi:10.1126/science.1250092.

Martin, F. W., and Jones, A. (1972). The Species of Ipomoea Closely Related to the Sweet Potato. Econ. Bot. 26, 201–215. Available at: https://www.jstor.org/stable/4253349.

Martin, F. W., Jones, A., and Ruberte, R. M. (1974). A Wild *Ipomoea* Species Closely Related to the Sweet Potato. Econ. Bot. 28, 287–292.

Mcdonald, J. A., and Austin, D. F. (1990). “Changes and Additions in *Ipomoea* Section Batatas (Convolvulaceae),” in Brittonia, 116–120.

Meyer, R. S., and Purugganan, M. D. (2013). “Evolution of crop species: Genetics of domestication and diversification,” in Nature Reviews Genetics, 840–852. doi:10.1038/nrg3605.

Montero-Pau, J., Blanca, J., Bombarely, A., Ziarsolo, P., Esteras, C., Martí-Gómez, C., et al. (2018). “De novo assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the *Cucurbita* genus,” in Plant Biotechnology Journal, 1161–1171. doi:10.1111/pbi.12860.

Muñoz-Rodríguez, P., Carruthers, T., Wood, J. R. I., Weitemier, K., Longway, L., Williams, B. R. M., et al. (2018). Reconciling Conflicting Phylogenies in the Origin of

Sweet Potato and Dispersal to Polynesia. Curr. Biol. 28, 1246-1256.e12. doi:10.1016/j.cub.2018.03.020.

Muñoz-Rodríguez, P., Carruthers, T., Wood, J. R. I., Williams, B. R. M., Weitemier, K., Kronmiller, B., et al. (2019). A taxonomic monograph of *Ipomoea* integrated across phylogenetic scales. Nat. Plants 5, 1136–1144. doi:10.1038/s41477-019-0535-4.

Muñoz-Rodríguez, P., Wells, T., Wood, J. R. I., Carruthers, T., Anglin, N. L., Jarret, R. L., et al. (2022). Discovery and characterization of sweetpotato's closest tetraploid relative. New Phytol. 234:4. doi:10.1111/nph.17991.

Muramatsu, M., and Shiotani, I. (1974). Closely related wild Ipomoea species of the sweet potato in Mexico and Guatemala. ed. M. Tanaka Kyoto: Faculty of Agriculture, Kyoto University Available at: https://m.kulib.kyoto-u.ac.jp/webopac/BB01184674.

Nishiyama, I. (1959). Collecting the sweet potato and its allied species in USA and Mexico, in Japanese Journal of Breeding, 73–78.

Nishiyama, I. (1961). The origin of the sweet potato. in Tenth Pacific Sci. Congr., Unviv. Hawaii (Honolulu), 119–128.

Nishiyama, I. (1971). Evolution and Domestication of the Sweet Potato. Bot. Maj. Tokyo 84, 377–387. Available at: https://www.jstage.jst.go.jp/article/jplantres1887/84/996/84_996_377/_pdf.

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. Japanese 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 (*Ipomoea trifida* (Kunth) G. Don), in Japanese Journal of Breeding, 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.

Nishiyama, I., and Teramura, T. (1962). "Mexican Wild Forms of Sweet Potato," in Economic Botany, 304–314.

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.,” in Theoretical and Applied Genetics, 617–624. doi:10.1007/BF00224220.

Ozias-Akins, P., and Jarret, R. L. (1994). Nuclear DNA content and ploidy levels in genus *Ipomoea*. J. Am. Soc. Hortic. Sci. 119, 110–115.

Pérez-Escobar, O. A., Bellot, S., Przelomska, N. A. S., Flowers, J. M., Nesbitt, M., Ryan, P., et al. (2021). Molecular Clocks and Archeogenomics of a Late Period Egyptian Date Palm Leaf Reveal Introgression from Wild Relatives and Add Timestamps on the Domestication. Mol. Biol. Evol. 38, 4475–4492. doi:10.1093/molbev/msab188.

Purugganan, M. D. (2019). Evolutionary Insights into the Nature of Plant Domestication. Curr. Biol. 29, R705–R714. doi:10.1016/j.cub.2019.05.053.

Quispe-Huamanquispe, D. G., Gheysen, G., Yang, J., Jarret, R., Rossel, G., and Kreuze, J. F. (2019). The horizontal gene transfer of *Agrobacterium* T-DNAs into the series Batatas (Genus *Ipomoea*) genome is not confined to hexaploid sweetpotato. Sci. Rep. 9, 1–13. doi:10.1038/s41598-019-48691-3.

Ramsey, J., and Schemske, D. W. (1998). Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. Annu. Rev. Ecol. Syst. 29, 467–501. doi:10.1146/annurev.ecolsys.29.1.467.

Razifard, H., Ramos, A., Della Valle, A. L., Bodary, C., Goetz, E., Manser, E. J., et al. (2020). Genomic evidence for complex domestication history of the cultivated tomato in Latin America. Mol. Biol. Evol. 37, 1118–1132. doi:10.1093/molbev/msz297.

Renner, S. S., Wu, S., Pérez-Escobar, O. A., Silber, M. V., Fei, Z., and Chomicki, G. (2021). A chromosome-level genome of a Kordofan melon illuminates the origin of domesticated watermelons. Proc. Natl. Acad. Sci. U. S. A. 118, 1–9. doi:10.1073/pnas.2101486118.

Roullier, C., Duputié, A., Wennekes, P., Benoit, L., Fernández Bringas, V. M., Rossel, G., et al. (2013). Disentangling the Origins of Cultivated Sweet Potato (*Ipomoea batatas* (L.) Lam.). PLoS One 8. doi:10.1371/journal.pone.0062707.

Shiotani, I., and Kawase, T. (1987). Synthetic Hexaploids Derived from Wild Species Related to Sweet Potato. Japanese J. Breed. 37, 367–376.

Shiotani, I., and Kawase, T. (1989). Genomic structure of sweet potato and hexaploids in *Ipomoea trifida* (H. B. K.) Don. Japanese J. Breed. 39, 57–66.

Srisuwan, S., Sihachakr, D., Martín, J., Vallès, J., Ressayre, A., Brown, S. C., et al. (2018). Change in nuclear DNA content and pollen size with polyploidisation in the sweet potato (*Ipomoea batatas*, Convolvulaceae) complex. Plant Biol., 0–2. doi:10.1111/plb.12945.

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. Plant Sci. 171, 424–433. doi:10.1016/j.plantsci.2006.05.007.

Stitzer, M. C., and Ross-Ibarra, J. (2018). Maize domestication and gene interaction. New Phytol. 220, 395–408. doi:10.1111/nph.15350.

Tayalé, A., and Parisod, C. (2013). Natural pathways to polyploidy in plants and consequences for genome reorganization. Cytogenet. Genome Res. 140, 79–96. doi:10.1159/000351318.

Teow, C. C., Truong, V. Den, McFeeters, R. F., Thompson, R. L., Pecota, K. V., and Yencho, G. C. (2007). Antioxidant activities, phenolic and β-carotene contents of sweet potato genotypes with varying flesh colours. Food Chem. 103, 829–838. doi:10.1016/j.foodchem.2006.09.033.

Teramura, T., and Nishiyama, I. (1962). Mexican wild forms of sweet potato. Econ. Bot. 16, 305–314. doi:10.1007/BF02860189.

Ting, Y. C., and Kehr, A. E. (1953). Meiotic studies in the sweet potato: (*Ipomoea batatas* Lam.). J. Hered. 44, 207–211. doi:10.1093/oxfordjournals.jhered.a106395.

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. Am. Nat. 91, 197–203.

van Bemmelen van der Plaat, A., van Treuren, R., and van Hintum, T. J. L. (2021). Reliable genomic strategies for species classification of plant genetic resources. BMC Bioinformatics 22, 1–18. doi:10.1186/s12859-021-04018-6.

van Ooststroom, S. J. (1953). “Convolvulaceae,” in Flora Malesiana I, ed. C. G. G. J. van Steenis (Batavia [Jakarta], Indonesia: Noordhoff-Kolff N.V.), 388–512.

Vaughan, D. A., Balázs, E., and Heslop-Harrison, J. S. (2007). From crop domestication to super-domestication, Annals of Botany 100(5), 893–901. doi:10.1093/aob/mcm224.

von Wettberg, E., Davis, T. M., and Smýkal, P. (2020). Editorial: Wild Plants as Source of New Crops. Front. Plant Sci. 11 , 1-3. doi:10.3389/fpls.2020.591554.

Weise, S., Lohwasser, U., and Oppermann, M. (2020). Document or lose it—on the importance of information management for genetic resources conservation in genebanks. Plants 9, 1–13. doi:10.3390/plants9081050.

Wood, J. R. I., Muñoz-Rodríguez, P., Williams, B. R. M., and Scotland, R. W. (2020). A foundation monograph of *Ipomoea* (Convolvulaceae) in the New World. PhytoKeys 143, 1–823. doi:10.3897/phytokeys.143.32821.

Yan, M., Quispe-huamanquispe, D. G., and Bock, R. (2021). Haplotype-based phylogenetic analysis uncovers the tetraploid progenitor of sweet potato. [Unpublished Pre-Print] ResearchSquare https://www.researchsquare.com/article/rs-750500/v1

Yang, J., Moeinzadeh, M.-H., Kuhl, H., Helmuth, J., Xiao, P., Haas, S., et al. (2017). Haplotype-resolved sweet potato genome traces back its hexaploidization history. Nat. Plants 3, 696–703. doi:10.1038/s41477-017-0002-z.

Zeder, M. A., Emshwiller, E., Smith, B. D., and Bradley, D. G. (2006). Documenting domestication: The intersection of genetics and archaeology. Trends Genet. 22, 139–155. doi:10.1016/j.tig.2006.01.007.

CHAPTER 3: DISTINGUISHING AUTO-TETRAPLOIDS FROM MODERN HYBRIDS IDENTIFIES THE PROGENITOR OF SWEET POTATO

(A revised version of this chapter was published in *New Phytologist*: Muñoz-Rodríguez P., Wells T., Wood J.R.I., Carruthers T., Anglin N.L., Jarret R.L., Scotland R.W. 2022. Discovery and characterization of sweetpotato's closest tetraploid relative. New Phytol. 234:4.)

CHAPTER ABSTRACT

Despite the status of sweet potato (*Ipomoea batatas* (L.) Lam.) as a globally significant food crop, its origin and evolution are still relatively poorly understood. Most importantly, how it evolved a hexaploid genome (2n=6x=90) remains unclear, impeding research and crop improvement breeding programmes. The transition to hexaploidy from a diploid ancestor (2n=2x=30) shared with its closest known relative *I. trifida* (Kunth) G.Don most likely involved a tetraploid intermediary (2n=4x=60). What is more, cultivated sweet potato is thought to be an allo-hexaploid, with two identical subgenomes and one that differs, in an AABBBB structure. This implies that it evolved from the union of a diploid (AA) (2n=2x=30) and tetraploid progenitor (BBBB), but while the AA subgenome is thought to have been inherited from a diploid ancestor shared with *I. trifida*, the tetraploid source of the BBBB subgenomes has never been definitively identified. Tetraploid plants collected growing wild in the Americas and resembling *I. batatas* exist in global germplasm collections, but no previous study has been able to demonstrate whether these are the descendants of sweet potato's tetraploid progenitor, or more recent hybrids between hexaploid *I. batatas* and diploid *I. trifida*. In this chapter, I integrate phylogenetic and genomic analyses with detailed morphometric study to distinguish recent hybrid tetraploids from autotetraploids that diverged prior to the origin of hexaploid *I. batatas*. The identification of wild, autotetraploid plants from Ecuador that are closely related to hexaploid *I. batatas* in phylogenetic analyses led to the description of this material

as a new species (*I. aequatoriensis* T.Wells & P. Munoz), which represents the closest crop wild relative (CWR) of sweet potato, and the probable descendant of its tetraploid progenitor. In contrast, tetraploids collected in Mexico and Central America are all most likely recent hybrids, akin to *I. tabascana* J.A. McDonald & D.F. Austin. The identification of the extant descendant of the tetraploid progenitor of hexaploid *I. batatas* represents a major milestone in sweet potato research, offering insights into the process of polyploidization, and potentially facilitating the incorporation of adaptive traits from it and other CWRs.

INTRODUCTION

A major barrier to understanding the origin and evolution of sweet potato (*Ipomoea batatas* (L.) Lam.) remains the difficulty of assembling its large, hexaploid genome (Isobe et al. 2019), which appears to be complicated by the likely AABBBB structure (sometimes denoted $B_1|B_1|B_2|B_2|B_2|B_2$) of its apparently divergent subgenomes (Ting and Kehr 1953; Ting et al. 1957; Jones 1965; Magoon et al. 1970; Nishiyama et al. 1975; Shiotani and Kawase 1987; Srisuwan et al. 2006; Gao et al. 2011; Yang et al. 2017). This structure suggests that *I. batatas* is an allohexaploid that arose from the union of a diploid and a tetraploid progenitor. The smaller subgenome (AA) has been identified as most likely to have been inherited from a diploid ancestor shared with closest known relative *I. trifida* (Roullier et al. 2013; Yang et al. 2017; Muñoz-Rodríguez et al. 2018), but the tetraploid progenitor of the two identical subgenomes (BBBB) has never been identified (Yang et al. 2017), complicating attempts to accurately assemble the sweet potato genome and confirm whether this allohexaploid hypothesis is correct. As such, definitively identifying the closest living relative of sweet potato's tetraploid progenitor is an important element to understanding the evolutionary origin and contemporary diversity of this globally important crop.

A complicating factor in efforts to identify the extant descendants of sweet potato's tetraploid progenitor is that since hexaploid plants will normally produce 3n gametes, hybridization between hexaploid *I. batatas* and its closest diploid relative *I. trifida* (which normally produces 1n gametes) will most often produce tetraploid offspring (3n + 1n = 4x) (Orjeda et al. 1991). As such, experimental evidence has shown that any recent hybrids between *I. trifida* and *I. batatas* are most likely to also be tetraploids (Orjeda et al. 1991). These tetraploids will possess a combination of genetic material from both their diploid and their hexaploid parent (Oracion et al. 1990; Srisuwan et al. 2006), and as a result will be recovered *intermediate* between

their two parents in phylogenetic analyses (Fig. 1), as is the case for the probable tetraploid hybrid *I. tabascana* (Rajapakse et al. 2004; Cao et al. 2014; Muñoz-Rodríguez et al. 2018).

Wild-collected tetraploid plants in germplasm collections that resemble *I. batatas* could therefore be either the direct descendants of sweet potato's tetraploid progenitor, or recent hybrids between hexaploid *I. batatas* and diploid *I. trifida*. Despite these crucial differences in their origin, they will be recovered in a similar position, closely related to hexaploid *I. batatas* in phylogenetic analyses of nuclear DNA sequence data (Fig.1). As a result, studies that rely purely on these analyses – without broader consideration of collection history, genomic structure, or chloroplast sequence data – are unlikely to be able to distinguish between the autotetraploid progenitor of hexaploid *I. batatas* and more recent hybrids between *I. batatas* and *I. trifida* (Fig.1).

Existing tetraploid collections previously associated with *I. batatas* and currently held in global germplasm repositories have never been comprehensively analysed together in a single integrative study based on a well-resolved phylogeny. Meanwhile descriptions of their morphology, collection history and other characteristics such as cytology have been partial, incomplete and sometimes contradictory or difficult to compare (see previous chapter). In order to be able to distinguish between autopolyploid and potential hybrid tetraploids among this material, I therefore designed an integrated approach. This approach combines phylogenetic, genomic and morphometric analyses of previously documented tetraploid specimens in the context of the well-resolved phylogenetic and taxonomic framework recently generated for the "Batatas group" of sweet potato's closest crop wild relatives (CWRs) (Muñoz-Rodríguez et al. 2018, 2019; Wood et al. 2020), and the systematic review undertaken in the previous chapter.

I designed the sampling strategy accordingly, by contacting germplasm repositories at CIP, USDA and NARO to request seeds, DNA, or silica-dried leaf material of all the available tetraploid germplasm collections previously discussed in relation to hexaploid *I. batatas*, as well as high-resolution images of accompanying herbarium specimens where these had been taken at the time of collection. I also visited, contacted or reviewed online digitized collections for herbaria in the UK (FHO,K,BM,E), Europe (P,ST,AAU), USA, (NY, US, MO, FAU, FTG), Mexico (MEXU, XAL), Colombia (COL), and Ecuador (QAP,QCA,QCNE,LOJA,GUAY) in order to gather details and morphological measurements from specimens, loans or high-resolution images of herbarium material described by other authors in connection with this tetraploid germplasm material (e.g., Dodson and Gentry 1978; Austin 1982; Dodson et al. 1985; Mcdonald and Austin 1990; Austin et al. 1993). This comprehensive sampling and integrative approach allowed me to better understand and interpret the relationships among different tetraploid collections, hexaploid *I. batatas*, and its closest diploid relative *I. trifida*.

MATERIALS & METHODS

Existing Phylogenetic Framework

The dataset used by Muñoz-Rodríguez et al. (2018, 2019) to reconstruct phylogenies for the “Batatas group” of sweet potato’s closest CWRs and the wider genus *Ipomoea* L. formed the backbone for this study. These samples had been sequenced for an initial 669 putatively single-copy nuclear loci and the whole chloroplast using the target capture method, Hyb-Seq (Muñoz-Rodríguez et al. 2018, 2019). Target capture approaches such as Hyb-Seq reduce costs and bioinformatic challenges, facilitating the analysis of a greater number of samples without sacrificing phylogenetic resolution (Andermann et al. 2020). The final 605

loci developed, sequenced, verified and analysed by Muñoz-Rodríguez et al. (2018) provided a well-resolved backbone phylogeny for *Ipomoea* and fully-resolved relationships for the 15 species in the Batatas group. This included recovering diploid *I. trifida* as the closest relative of cultivated hexaploid sweet potato, and demonstrating that no other currently known species was involved in its origin (Muñoz-Rodríguez et al. 2018).

Representative specimens of each of the 12 other Batatas group species (not including suspected hybrid *I. leucantha* Jacq.) were therefore treated as the outgroup in this study, along with *I. cryptica* J.R.I.Wood & Scotland, which was recovered sister to the Batatas group (Wood et al. 2020), with *I. trifida, I. batatas,* suspected tetraploid hybrid *I. tabascana,* and newly sequenced tetraploid germplasm material forming the ingroup. Since the aim of this study is to reconstruct likely relationships between sweet potato and its potential wild progenitor, I excluded previously sequenced accessions of hexaploid *I. batatas* that are likely to represent highly bred modern cultivars of sweet potato, as these often possess introgressed alleles resulting from modern breeding programmes that may confound interpretation (Razifard et al. 2020). Accessions of *I. batatas* from outside the Americas (sweet potato's presumed centre of origin), and those with names that included numbers were therefore not included in this study.

Sampling

The samples used in phylogenetic and genomic analyses in this study are listed in Table 1. These included 58 accessions of hexaploid *I. batatas*, 24 accessions of diploid *I. trifida*, 13 accessions of other Batatas group species, and 12 accessions of tetraploid germplasm collections from Mexico, Colombia and Ecuador. The germplasm material thus comprised representative samples from across the range of past tetraploid collections, including the accession PI518479, which was

previously described as the Batatas group species *I. tabascana* (McDonald and Austin 1990; Austin et al. 1991), and subsequently shown to be a likely hybrid between hexaploid *I. batatas* and diploid *I. trifida* (Rajapakse et al. 2004; Srisuwan et al. 2006; Cao et al. 2014; Muñoz-Rodríguez et al. 2018). Including this particular sample (*I. tabascana)* allowed me to use it as a point of reference for assessing how putative hybrids of *Ipomoea batatas* and *I. trifida* behave in phylogenetic and genomic analyses.

Additional DNA sequencing and analysis for this study were conducted to fit within the existing framework outlined above, but given the smaller number of samples and subsequent reduction in sequencing costs, WGS was used instead of Hyb-Seq for the 11 additional samples. This allowed me to assemble the same nuclear loci used by Muñoz-Rodríguez et al. (2018), as well as the whole chloroplast genome, while also facilitating genomic analyses with off-target reads.

Since morphological data were not available for the majority of the samples included in previous studies (Muñoz-Rodríguez et al. 2018), representative herbarium material for each taxon was gathered and analysed for a set of 13 morphological characters. This material is listed in Appendix 2 at the end of this thesis. It included 28 specimens of hexaploid *I. batatas*, 60 specimens of diploid *I. trifida*, and 109 specimens taken at the time of collection for tetraploid germplasm accessions, or previously cited in connection with them (e.g. Dodson & Gentry, 1978; Austin, 1982; Dodson et al., 1985; McDonald & Austin, 1990). Morphometric data was collected for all the sequenced tetraploid germplasm material with the exception of DLP2923 from Mexico, for which no herbarium voucher or seeds were available. One sample of diploid *I. trifida* (I_trifida_PW_108) previously sequenced by Muñoz-Rodríguez et al. (2018) was also included in both analyses, but this was not possible for any of the previously sequenced hexaploid *I. batatas* material.

Cultivation and Preparation of Germplasm Material

A total of eight accessions were received from several global germplasm repositories (CIP, USDA, & NARO) and grown in the greenhouses at the Department of Biology, Oxford. Seeds were scarified using a sharp scalpel, and soaked in water for 24 hours to stimulate germination. Two to four seeds of each accession were then sown in a 3:1 compost and vermiculite mix in 10cm pots, and placed in a closed propagator at ~20^{o}C. Successfully germinated seedlings were grown on and potted up in incrementally larger pots until they were large enough to be transferred to 25L Air-pots, which are designed to promote healthy root growth. All accessions were grown at an average temperature of 20^{o}C with a short-day lighting cycle (8 hours on, 16 hours off) to promote flowering (Teramura and Nishiyama 1962). Time to flowering was typically four to eight months. Flowering material was selected, pressed and dried in newspaper with a wooden plant press to facilitate comparison with herbarium material, and leaf samples were also taken in silica gel for DNA-extraction. Silica-dried leaf material or pre-extracted genomic DNA were also received from NARO and CIP for a further 2 (K500, K233, K123) and 1 (DLP2923) samples respectively.

DNA-Extraction & Sequencing

DNA-extraction of new samples was carried out by Dr Pablo Muñoz-Rodríguez or myself at Oxford using the Plant Tissue Mini protocol for the Qiagen DNEasy Plant MiniKit, or by Dr Masaru Tanaka at NARO, Japan using the same method. Genomic libraries were created using the NEBNext Ultra DNA LibraryPrep Kit for Illumina v.3.0 (New England BioLabs, Ipswich, MA, USA), before being sequenced at Novogene facilities in Cambridge, UK using Illumina NovoSeq6000. For each sample 150bp paired-end whole genome data was obtained, at an average of 11Gb per sample.

TABLE 1: SAMPLES USED IN PHYLOGENETIC ANALYSES

Sample	Source	Accession #	Collector	Number	Country	Ploidy	Sequencing
I_cryptica_DS_1487	USZ		Soto, D	1487	Bolivia	2	Hyb-Seq
I_australis_CIP460211	CIP	CIP 460211			Argentina	2	Hyb-Seq
I_cordatotriloba_CIP460604	CIP	CIP 460604			USA	2	Hyb-Seq
I_cynanchifolia_CIP460555	CIP	CIP 460555			Brazil	2	Hyb-Seq
I_grandifolia_CIP460311	CIP	CIP 460311			Paraguay	2	Hyb-Seq
I_lactifera_GCL_149	K		Caity Lero, G.	149	Bolivia	2	Hyb-Seq
I_lacunosa_JAS_76626	BM		Steyermark, J.A.	76626	USA	2	Hyb-Seq
I_littoralis_Anon_152892	SAN		Sugau, J.B.	SAN 152892	Malaysia	2	Hyb-Seq
I_ramosissima_CIP460033	CIP	CIP4 60033			Bolivia	2	Hyb-Seq
I_splendorsylvae_PW_472	BM		Wilkin, P.	472	Costa Rica	2	Hyb-Seq
I_tenuissima_AHC_495	BM		Curtiss, A.H.	495	Cuba	2	Hyb-Seq
I_tiliacea_CIP460524	CIP	CIP4 60524			Cuba	4	Hyb-Seq

Sample	Source	Accession #	Collector	Number	Country	Ploidy	Sequencing
I_triloba_CCHJ_10788	MA		Jongkind, C.C.H.	10788	Guinea	2	Hyb-Seq
I_trifida_CIP460021_VEN	CIP	CIP4 60021			Venezuela	2	Hyb-Seq
I_trifida_CIP460055_VEN	CIP	CIP4 60055			Venezuela	2	Hyb-Seq
I_trifida_CIP460061_VEN	CIP	CIP4 60061			Venezuela	2	Hyb-Seq
I_trifida_CIP460066_VEN	CIP	CIP4 60066			Venezuela	2	Hyb-Seq
I_trifida_CIP460096_VEN	CIP	CIP4 60096			Venezuela	2	Hyb-Seq
I_trifida_CIP460099_COL	CIP	CIP4 60099			Colombia	2	Hyb-Seq
I_trifida_CIP460120_COL	CIP	CIP4 60120			Colombia	2	Hyb-Seq
I_trifida_CIP460134_COL	CIP	CIP4 60134			Colombia	2	Hyb-Seq
I_trifida_CIP460144_COL	CIP	CIP4 60144			Colombia	2	Hyb-Seq
I_trifida_CIP460183_NIC	CIP	CIP4 60183			Nicaragua	2	Hyb-Seq
I_trifida_CIP460260_GTM	CIP	CIP4 60260			Guatemala	2	Hyb-Seq
I_trifida_CIP460274_GTM	CIP	CIP4 60274			Guatemala	2	Hyb-Seq
I_trifida_CIP460384_GTM	CIP	CIP4 60384			Guatemala	2	Hyb-Seq

Sample	Source	Accession #	Collector	Number	Country	Ploidy	Sequencing
I_trifida_CIP460396_NIC	CIP	CIP4 60396			Nicaragua	2	Hyb-Seq
I_trifida_CIP460403_NIC	CIP	CIP4 60403			Nicaragua	2	Hyb-Seq
I_trifida_CIP460407_GTM	CIP	CIP4 60407			Guatemala	2	Hyb-Seq
I_trifida_CIP460530_CUB	CIP	CIP4 60530			Cuba	2	Hyb-Seq
I_trifida_CIP460547_CUB	CIP	CIP4 60547			Cuba	2	Hyb-Seq
I_trifida_CIP460733_GTM	CIP	CIP4 60733			Guatemala	2	Hyb-Seq
I_trifida_CIP460739_GTM	CIP	CIP4 60739			Guatemala	2	Hyb-Seq
I_trifida_CIP460756_GTM	CIP	CIP4 60756			Guatemala	2	Hyb-Seq
I_trifida_PW_108	K		Wilkin, P	108	Cost Rica	2	Hyb-Seq
I_batatas_CIP400033_Abobora_BRA	CIP	CIP4 00033			Brazil	6	Hyb-Seq
I_batatas_CIP400155_Rojo_BOL	CIP	CIP4 00155			Bolivia	6	Hyb-Seq
I_batatas_CIP400157_Amarillo_BOL	CIP	CIP4 00157			Bolivia	6	Hyb-Seq
I_batatas_CIP400162_BlancoPapa_BOL	CIP	CIP4 00162			Bolivia	6	Hyb-Seq
I_batatas_CIP400205_Morado_BOL	CIP	CIP4 00205			Bolivia	6	Hyb-Seq

Sample	Source	Accession #	Collector	Number	Country	Ploidy	Sequencing
I_batatas_CIP400218_Tarijeno_BOL	CIP	CIP4 00218			Bolivia	6	Hyb-Seq
I_batatas_CIP400238_Amarelha_BRA	CIP	CIP4 00238			Brazil	6	Hyb-Seq
I_batatas_CIP400274_Branca_BRA	CIP	CIP4 00274			Brazil	6	Hyb-Seq
I_batatas_CIP400287_Branca_BRA	CIP	CIP4 00287			Brazil	6	Hyb-Seq
I_batatas_CIP400339_Iranduba_BRA	CIP	CIP4 00339			Brazil	6	Hyb-Seq
I_batatas_CIP400356_Amarillo_COL	CIP	CIP4 00356			Colombia	6	Hyb-Seq
I_batatas_CIP400361_Cordoba_COL	CIP	CIP4 00361			Colombia	6	Hyb-Seq
I_batatas_CIP400389_Huasca_COL	CIP	CIP4 00389			Colombia	6	Hyb-Seq
I_batatas_CIP400423_Blanca_COL	CIP	CIP4 00423			Colombia	6	Hyb-Seq
I_batatas_CIP400433_Batatilla_COL	CIP	CIP4 00433			Colombia	6	Hyb-Seq
I_batatas_CIP400435_NegraOMorena_COL	CIP	CIP4 00435			Colombia	6	Hyb-Seq
I_batatas_CIP400442_Narino_COL	CIP	CIP4 00442			Colombia	6	Hyb-Seq
I_batatas_CIP400453_Narino_COL	CIP	CIP4 00453			Colombia	6	Hyb-Seq
I_batatas_CIP400466_Blanca_COL	CIP	CIP4 00466			Colombia	6	Hyb-Seq

Sample	Source	Accession #	Collector	Number	Country	Ploidy	Sequencing
I_batatas_CIP400493_Batatas_COL	CIP	CIP4 00493			Colombia	6	Hyb-Seq
I_batatas_CIP400502_Amarilla_COL	CIP	CIP4 00502			Colombia	6	Hyb-Seq
I_batatas_CIP400507_Morada_COL	CIP	CIP4 00507			Colombia	6	Hyb-Seq
I_batatas_CIP400519_Pamplona_COL	CIP	CIP4 00519			Colombia	6	Hyb-Seq
I_batatas_CIP400551_Garita_CRI	CIP	CIP4 00551			Costa Rica	6	Hyb-Seq
I_batatas_CIP400572_Garita_CRI	CIP	CIP4 00572			Costa Rica	6	Hyb-Seq
I_batatas_CIP400784_LaLoquita_DOM	CIP	CIP4 00784			Dominican Rep.	6	Hyb-Seq
I_batatas_CIP400786_Copela_DOM	CIP	CIP4 00786			Dominican Rep.	6	Hyb-Seq
I_batatas_CIP400796_SolaNoEstan_DOM	CIP	CIP4 00796			Dominican Rep.	6	Hyb-Seq
I_batatas_CIP400803_ArroyoLoro_DOM	CIP	CIP4 00803			Dominican Rep.	6	Hyb-Seq
I_batatas_CIP400811_ArroyoLoro_DOM	CIP	CIP4 00811			Dominican Rep.	6	Hyb-Seq
I_batatas_CIP400815_Brujita_DOM	CIP	CIP4 00815			Dominican Rep.	6	Hyb-Seq
I_batatas_CIP400819_Tunita_DOM	CIP	CIP4 00819			Dominican Rep.	6	Hyb-Seq
I_batatas_CIP400822_Canabacoa_DOM	CIP	CIP4 00822			Dominican Rep.	6	Hyb-Seq

Sample	Source	Accession #	Collector	Number	Country	Ploidy	Sequencing
I_batatas_CIP400826_LaSelva_DOM	CIP	CIP4 00826			Dominican Rep.	6	Hyb-Seq
I_batatas_CIP400941_DeDulce_ECU	CIP	CIP4 00941			Ecuador	6	Hyb-Seq
I_batatas_CIP401056_Blanco_GTM	CIP	CIP4 01056			Guatemala	6	Hyb-Seq
I_batatas_CIP401062_Moral_GTM	CIP	CIP4 01062			Guatemala	6	Hyb-Seq
I_batatas_CIP401067_Morado_GTM	CIP	CIP4 01067			Guatemala	6	Hyb-Seq
I_batatas_CIP401068_Morado_GTM	CIP	CIP4 01068			Guatemala	6	Hyb-Seq
I_batatas_CIP401079_Blanco_GTM	CIP	CIP4 01079			Guatemala	6	Hyb-Seq
I_batatas_CIP401101_Rojo_GTM	CIP	CIP4 01101			Guatemala	6	Hyb-Seq
I_batatas_CIP401129_Blanco_GTM	CIP	CIP4 01129			Guatemala	6	Hyb-Seq
I_batatas_CIP401150_DosSemanas_HND	CIP	CIP4 01150			Honduras	6	Hyb-Seq
I_batatas_CIP401210_Cuitzeo_MEX	CIP	CIP4 01210			Mexico	6	Hyb-Seq
I_batatas_CIP401404_Merida_VEN	CIP	CIP4 01404			Venezuela	6	Hyb-Seq
I_batatas_CIP401443_Morada_VEN	CIP	CIP4 01443			Venezuela	6	Hyb-Seq
I_batatas_CIP401459_Blanco_VEN	CIP	CIP4 01459			Venezuela	6	Hyb-Seq

Sample	Source	Accession #	Collector	Number	Country	Ploidy	Sequencing
I_batatas_CIP401499_Merida_VEN	CIP	CIP4 01499			Venezuela	6	Hyb-Seq
I_batatas_CIP401523_Merida_VEN	CIP	CIP4 01523			Venezuela	6	Hyb-Seq
I_batatas_CIP420068_Sal_PER	CIP	CIP4 20068			Peru	6	Hyb-Seq
I_batatas_CIP420386_Acaro_PER	CIP	CIP4 20386			Peru	6	Hyb-Seq
I_batatas_CIP420602_Morado_PER	CIP	CIP4 20602			Peru	6	Hyb-Seq
I_batatas_CIP420882_Colombiano_PER	CIP	CIP4 20882			Peru	6	Hyb-Seq
I_batatas_CIP441494_Quiabo_BRA	CIP	CIP4 41494			Brazil	6	Hyb-Seq
I_batatas_CW_71	BM		Whitefoord, C.	71	Panama	6	Hyb-Seq
I_batatas_DMC_407	HUEFS		Carvalho, D.M.	407	Brazil	6	Hyb-Seq
I_batatas_JRIW_27946	USZ		Wood, J.R.W.I.	27946	Bolivia	6	Hyb-Seq
I_batatas_MN_37607	MO		Nee, M.	37607	Bolivia	6	Hyb-Seq
I_tabascana4X_DLP2946_MEX	FTG		De La Puente, F.	2946	Mexico	4	Hyb-Seq
I_tabascana4X_PI_518479_MEX	**USDA**	**PI 518479**	**De La Puente, F.**	**2946**	**Mexico**	**4**	**Illumina**
I_batatas_var_apiculata4X_DLP2908_MEX	**USDA**	**PI 518474**	**De La Puente, F.**	**2908**	**Mexico**	**4**	**Illumina**

Sample	Source	Accession #	Collector	Number	Country	Ploidy	Sequencing
I_batatas_var_apiculata4X_K233_MEX	NARO	K233	Muramatsu		Mexico	4	Illumina
I_sp4X_DLP2923_MEX	CIP	CIP 403958	De La Puente, F.	2923	Mexico	4	Illumina
I_sp4X_K500_COL	USDA	K500	Kobayashi		Colombia	4	Illumina
I_sp6X_K123_MEX	NARO	K123	Nishiyama		Mexico	6	Illumina
I_sp4X_DFA7817_ECU	USDA	PI 561258	Austin, D.F.	7817	Ecuador	4	Illumina
I_sp4X_DLP5282_ECU	USDA	PI 561246	De La Puente, F.	5282	Ecuador	4	Illumina
I_sp4X_DLP5284_ECU	USDA	PI 561248	De La Puente, F.	5284	Ecuador	4	Illumina
I_sp4X_DLP5296_ECU	USDA	PI 561255	De La Puente, F.	5296	Ecuador	4	Illumina
I_sp4X_R71_3_ECU	USDA	CH71.3	Rick, C.M.	SAM 2614	Ecuador	4	Illumina
I_sp4X_R81_2_ECU	USDA	CH81.2	Rick, C. M.	SAM 5316	Ecuador	4	Illumina

Table 1: Samples used in phylogenetic analyses, showing unique identifier based on taxonomic determination, followed by institutional accession number in the case of germplasm material, or collector and collection number in the case of herbarium specimens. Institution codes according to Index Hebariorum (https://sweetgum.nybg.org/science/ih/), with the exception of CIP, NARO, and USDA, which refer to the International Potato Centre (Peru), the National Agriculture and Food Research Organization (Japan), and the United States Department of Agriculture (USA) respectively.

Filtering and Quality Control

Raw sequencing reads were pre-processed with quality assessment, trimming and correction using fastp (Chen et al. 2018) with default settings and base correction (--correction). Adapters were identified and trimmed using overlap analysis, and reads shorter than 15bp, with a phred quality score less than 15, 40% unqualified bases, or more than 5 unidentified (N) bases were filtered and discarded. Base correction in fastp is based on detecting overlapping regions in paired end reads, where the minimum length of overlap is 30, the maximum number of mismatched bases in the overlapped region is 5, and the maximum percentage of mismatched bases is 20%. Mismatched bases are then corrected only if there is an imbalance in quality scores (e.g. one low quality (<Q15) and the other high (>Q30)).

Verification of Single-Copy Loci

Muñoz-Rodríguez et al. (2018) developed target capture probes for an initial 669 putatively low-copy nuclear loci covered by 20,020 custom DNA baits designed by MycroArray (Ann Arbor, MI). This was achieved by aligning partially assembled scaffolds of the *I. lacunosa* L. genome to *Solanum lycopersicum* L. coding sequences. A final set of 605 of these loci were found to map to unique positions on the 506 *I. lacunosa* scaffolds, and these were selected for use in their analyses (Muñoz-Rodríguez et al. 2018). Since the publication of that study however, high-quality chromosome-level whole genome assemblies have been completed for several other species of *Ipomoea* (Hoshino et al. 2016; Wu et al. 2018; Li et al. 2019). It is important to rule out duplication and paralogs in target-capture data (Zhou et al. 2021), so before conducting any new analyses, I reassessed the 605 loci in relation to the recent high-quality assembly for sweet potato's closest known relative, *I. trifida* (Wu et al. 2018). Dr Muñoz-Rodríguez provided me with both the original set of 669 *I. lacunosa* target loci, and a set of representative sequences

assembled for each of the final 605 *Ipomoea* loci, which he had compiled from multiple accessions of different *Ipomoea* species following sequencing and assembly. The original set of 669 *I. lacunosa* target loci were named for their location on the *I. lacunosa* scaffolds, with the first set of numbers designating the scaffold, and the last two (following the semi-colon) representing their start and end position on that scaffold (e.g.: 2468068_8659_420654_258057+,...,219245-;1195-6576). By contrast, the 605 assembled loci were named to correspond to the *S. lycopersicum* loci they and the *I. lacunosa* loci had originally been derived from, with an additional number (following the underscore) to designate assembled loci that were derived from the same initial *S. lycopersicum* locus (e.g.: Solyc00g006800.2.1_4). I mapped both sets of loci to the high-quality, chromosome-level *I. trifida* genome assembly (Wu et al. 2018) using the Burrows-Wheeler Alignment (bwa) mem function (Li and Durbin 2010), and visualized the results in geneious v.9.

Denovo Assembly with HybPiper

Once the raw sequencing reads had been quality-assessed, trimmed and corrected, they were mapped to the newly confirmed single-copy target loci using HybPiper, a set of python scripts specifically designed for the assembly of target capture sequence data (Johnson et al., 2016). HybPiper proceeds by first mapping reads to target loci using bwa mem (Li and Durbin 2010), and then distributing the mapped reads to separate directories for each locus. The successfully mapped reads for each locus are subsequently assembled denovo using SPAdes (Bankevich et al. 2012). Separate commands allow for the extraction of intronic regions either side of the targeted locus (intronerate), and for detecting putative paralogs (paralog_investigator). Since mapping reads with HybPiper is a memory-intensive and relatively slow process that generates large amounts of data, I developed a snakemake (Mölder et al. 2021) pipeline to parallelise runs such that multiple

samples could be run concurrently. I also edited the original HybPiper scripts to replace bwa mem with the improved version bwa-mem2, which runs significantly faster (Md et al. 2019), and to discard unaligned read ends (samclip) and unmapped reads (samtools view -F 4) (Danecek et al. 2021), thereby reducing the space required by output files and avoiding potential interference with subsequent genotyping pipelines.

Genotyping Pipeline

Resolving phylogenetic relationships in closely related or recently diverged taxa can be challenging because of conflicting signals resulting from incomplete lineage sorting (ILS) and high-levels of heterozygosity (Lischer et al. 2014). This is a particular issue for denovo assembly of sequence data, where high heterozygosity complicates assembly, and heterozygous alleles are often discarded from de Bruijn graph “bubbles” (Kajitani et al. 2014, 2019). For polyploid organisms that may possess divergent subgenomes, there is therefore a specific risk that denovo assembly will artificially obscure heterozygosity between different subgenomes in consensus sequences by discarding lower frequency variants, confounding the accurate reconstruction of phylogenetic relationships. To address this possibility, Dr Muñoz-Rodríguez developed a heterozygosity-aware pipeline based on genotype-calling, which I used with minor adjustments for this analysis.

For each sample:

1. Filtered sequencing reads were first mapped to the confirmed single-copy target loci using bwa-mem2 (Md et al. 2019).
2. Samtools (Danecek et al. 2021) was then used to extract all successfully mapped reads and to remove duplicate reads.
3. Picard Tools (http://broadinstitute.github.io/picard) was used to realign reads mapped around indels.

4. Bcftools (Danecek et al. 2021) was used to call variants, normalise indels and then filter variants such that heterozygous sites (0/1) where the dominant allele had more than twice the coverage depth of the alternate allele or vice versa were set to homozygous (0/0 or 1/1). This resulted in only heterozygous sites with >65% support in terms of read depth being retained.
5. The vcftools function vcf-sort (Danecek et al. 2011) was used to sort the resulting VCF file.
6. Consensus sequences were then generated for each accession using the bcftools consensus function, with the remaining heterozygous sites coded using IUPAC ambiguity codes, and sites assessed as having coverage <5 with bedtools genomecov masked.
7. For each accession, any loci shown to have a mean coverage <20X with bcftools query were excluded from subsequent analyses.

By explicitly recognizing and filtering potential heterozygosity, this method therefore allowed us to account for the likely presence of divergent subgenomes in the case of allohexaploid *I. batatas* and potential tetraploid hybrids, thereby helping to mitigate their impact on phylogenetic inference (but see discussion below for limitations of this approach).

Phylogenetic Analysis of Nuclear Sequence Data

For each set of loci (denovo-assembled and genotyped), the following pipeline was implemented in snakemake (Mölder et al. 2021) to align sequences from all samples at each recovered locus, trim poorly aligned regions, generate gene tree phylogenies for each locus, and then species trees based on coalescent analysis of all gene trees:

1. For each locus, sequences shorter than 500bp or with more than 10% unknown bases (N) were excluded.

2. Sequences for each locus were then aligned using mafft (Katoh and Standley 2013), and poorly aligned regions trimmed using trimal (Capella-Gutiérrez et al. 2009). Summary statistics were calculated for each aligned locus using AMAS (Borowiec 2016).
3. Individual gene trees were generated using IQTree (Minh et al. 2020). The default ModelFinder (Kalyaanamoorthy et al. 2017) option was used to choose the appropriate substitution model, and ultrafast bootstrap approximation (UFBoot) was run using 1000 replicates.
4. A concatenated alignment of all loci was generated using AMAS, and analysed using the same IQTree parameters.
5. Gene tree branches with less than 10% support were collapsed (Zhang et al. 2018) using Newick Utilities (Junier and Zdobnov 2010; Czech et al. 2017).
6. A species tree was generated using ASTRAL-III (Zhang et al. 2018), with both local posterior probabilities and alternative quartet scores (Sayyari and Mirarab 2016) calculated for the resulting topology.

For each assembly method (denovo-assembled and genotyped), this pipeline was implemented for the full set of 107 samples, and a second reduced set with an equal number of ingroup samples (12) for each ploidy group (diploid, tetraploid, hexaploid), to ensure that the relatively high proportion of hexaploid *I. batatas* samples in the full sampling set (58) was not biasing results (e.g. Hillis 1998; Plazzi et al. 2010).

Challenges of Network-based Analysis

Since phylogenetic analyses based on consensus sequence data are unlikely to be able to accurately resolve reticulate evolutionary relationships resulting from hybridization and allopolyploidy (Fig. 1), network-based or population genetic approaches were considered for their potential to provide better clarity as to

the relationships among ingroup samples. For example, the python package HyDe (Blischak et al. 2018) can be used to detect hybridization from phylogenetic "invariants" under the coalescent model. It does "D-statistic", whereby the using ABBA or BABA patterns should occur with equal frequency in the absence of admixture. HyDe has been used to better understand hybridization and phylogenetic relationships within the genus *Salix* L. (Salicaceae) for example (Wagner et al. 2020). Interestingly, this included species of differing ploidies, despite HyDe not being designed to handle non-diploid datasets.

Unfortunately, this and the majority of other network-based methods are not ploidy-aware and therefore require the artificial diploidization of data. This is also the case when creating PCAs based on genomic variants, for example using software such as PLINK (Purcell et al. 2007). By limiting possible genotypes to two in polyploid sequence data, these tools therefore risk discarding important genetic variation. In this sense, the difficulty of handling mixed-ploidy datasets is not restricted to phylogenetic analysis, but rather presents itself at the sequence assembly stage, thus also impacting network-based approaches. It was therefore necessary to use alternative methods to establish whether tetraploid material is autopolyploid or of likely hybrid origin. By using the results of the phylogenetic analyses carried out here in conjunction with these other methods, I endeavoured to ensure that the limitations of phylogenetic approaches were mitigated and accounted for during interpretation.

In future analyses, I intend to test two further approaches, in the hope of overcoming these issues. The first is to use the admixture analysis program STRUCTURE (Pritchard et al. 2000), which is capable of handling mixed-ploidy datasets, though this is not immediately clear from the manual. STRUCTURE is a deceptively simple, yet extremely powerful method for detecting evidence of hybridization and has been shown to be more robust that alternative methods in

mixed-ploidy datasets such as this one (Stift et al. 2019). The second promising approach is the use of highly repetitive transposons as naturally-occurring species-specific markers in sequence data, which can be used to tease apart sequence data potentially inherited through allopolyploidy (Gordon et al. 2019). This approach has been successful in a number of other polyploid crops such as Tobacco and Wheat (Gordon et al. 2019), with new tools and approaches recently developed (Session and Rokhsar 2023).

K-Mer Analyses of Heterozygosity in Nuclear Sequence Data

The first approach I chose to integrate with the results of the phylogenetic analyses is the use of kmer frequencies to assess whether samples are likely autopolyploid or else of hybrid origin. The program GenomeScope2.0 (Ranallo-Benavidez et al. 2020) calculates heterozygosity from k-mer frequencies of unaligned sequencing reads, so that relative frequency patterns can be used to infer whether a tetraploid sample is autopolyploid or allopolyploid. Since tetraploid hybrids between hexaploid *I. batatas* and diploid *I. trifida* are likely to have a sub-genomic structure similar to allotetraploids (Fig. 1: A+ABB=AABB), I applied this method to all tetraploid accessions with sufficient sequencing depth. Initial k-mer counting and histogram construction were carried out on the filtered but unaligned sequencing reads using Jellyfish (Marçais and Kingsford 2011). I then ran both Jellyfish and GenomeScope2.0 with a maximum coverage of 100,000 and the default k-mer value of 21.

Assembly & Analysis of Chloroplast Sequence Data

Previous phylogenetic analysis of chloroplast sequence data has shown that sweet potato possesses two chloroplast lineages (Roullier et al. 2011, 2013; Muñoz-Rodríguez et al. 2018), one of which – Lineage 2 *sensu* Muñoz-Rodríguez et al. (2018) – is more closely related to *I. trifida* than it is to the other *I. batatas* lineage.

No evidence of these two lineages was found in the nuclear genome by Muñoz-Rodríguez et al. (2018), who suggested that the most likely explanation was therefore chloroplast capture, whereby hybridization and repeated backcrossing results in chloroplast-only introgression (Soltis 1991; Folk et al. 2017), giving rise to Lineage 2. Some tetraploid germplasm accessions have previously been shown to associate with one or other of these lineages in phylogenetic analysis of the chloroplast barcode marker *rpl32-TrnL*(UAG), perhaps depending on geographic origin (Roullier et al. 2013a; Muñoz-Rodríguez pers. comm.) (see Chapter 2), but this result has not been confirmed for a comprehensive sampling of tetraploid germplasm material, or re-examined in light of the chloroplast capture hypothesis.

On the basis that Lineage 1 represents the "ancestral" chloroplast lineage in hexaploid *I. batatas*, the analysis of chloroplast sequence data undertaken here was proposed as a means of further testing whether these and other tetraploid germplasm samples represent the progenitor of hexaploid *I. batatas*, or are recent hybrids between *I. batatas* and *I. trifida*. Dr Pablo Muñoz-Rodríguez assembled a list of 56 chloroplast genes and variable regions previously shown to be phylogenetically informative. I then used HybPiper to assemble these target loci from the filtered sequencing reads, with intronerate and paralog detector activated.

Phylogenetic Analysis of Chloroplast Sequence Data

The assembled chloroplast regions were then:

1. Filtered for sequences shorter than 500bp or with more than 10% unknown bases (N).
2. Aligned with Mafft v.7.310 (Katoh and Standley 2013).
3. Trimmed of poorly aligned regions using TrimAl (Capella-Gutiérrez et al. 2009).
4. Concatenated using AMAS (Borowiec 2016).

5. Analysed with IQTree (Minh et al. 2020) to generate a Maximum Likelihood based phylogeny, with default ModelFinder option and 1000 ultrafast bootstrap replicates.

Morphometrics

In order to quantify morphological differences among the major clades identified in phylogenetic analysis of nuclear sequence data, I compiled and compared a set of 12 morphological measurements based on those commonly used in *Ipomoea* taxonomy (Austin 1978; Bohac et al. 1993; Wood et al. 2020) and by sweet potato breeders (Huaman 1991). These characters were then reassessed in light of a review of relevant herbarium and germplasm material to assess their likely utility and phylogenetic informativeness. Measurements were taken from dried and pressed herbarium specimens using digital callipers and a digital protractor, or in the case of digitized herbarium specimens, using the biological-image analysis software Fiji (Schindelin et al. 2012).

The analysis of herbarium material presents particular challenges due to the drying and pressing of specimens, and this meant that some of the characters assessed by sweet potato breeders from living material (e.g. Huaman 1991) are not appropriate, or had to be modified. An additional challenge is the variation caused by the developmental stage and/or part of the plant that is collected. For example, collecting material from the very tips and young growth of the plant can influence characters including stem thickness, leaf size and even inflorescence form, while the number, developmental stage and condition of flowers on specimens can be highly variable. To account for this, specimens with a total length shorter than 15cm were not included in the analysis, and measurements of vegetative characters (e.g. leaf and internode dimensions) were taken from the largest visible organ on each

herbarium sheet, while floral characters were only measured from fully open flowers to ensure comparable levels of maturity.

Stem internode length was measured to the nearest 0.5cm on the longest internode. Stem internode diameter was measured in millimetres to 2dp, also on the longest internode. Leaf dimensions of the largest leaf were measured to the nearest 0.5cm. Length of leaf was measured from the tip to furthest point on the base of the blade, and width was measured between the widest two points of the lobes, where present. Lobes can vary greatly in their depth and prominence, so were counted based on the number of teeth, one for each lobe.

The maximum number of flowers per inflorescence was assessed by counting pedicels, or pedicel scars if necessary. The pedicel length was measured only from open flowers. Sepal dimensions (length and width) were measured in millimetres to 2decimal places for the shortest (outermost) and longest (innermost) sepals on an open flower. The difference between these two measurements was then calculated as a percentage and treated as a separate character. The angle of the apex of the largest (innermost) sepal was also measured to the nearest 10^{o}. Corolla length and width at base of corolla tube were both measured in millimetres to 2dp. The length was taken from the base of the calyx to the top of the corolla tube on a fully opened flower, and the width of the corolla tube was measured immediately above the longest (innermost) sepal, following Bohac et al. (1993).

Data visualisation and statistical analyses were performed in R, initial data visualisation was carried out using the ggplot2 package v.3.3.5 (Wickham 2016), and a correlation matrix for coefficients of all pairs of variables was calculated using the corrplot function. Correlation tests of significance for each correlation coefficient were then calculated using the correlation function from the package easystats. Scatter and violin plots were used to visualise pairs of related variables and their

distributions, with ellipses depicting 95% confidence level added to scatter plots using the stat_ellipse function, and a linear regression model fitted using the lm function and plotted using geom_abline.

I then ran a principal component analysis (PCA) to investigate phenotypic clustering among the four major clades recovered in the phylogenetic analysis of genotyped loci (diploid *I. trifida*, Ecaudorian/Colombian tetraploids, Mexican tetraploids, and hexaploid *I. batatas*) using the factominer package v.2.4 (Le et al. 2008). The data were scaled by using the scale function to subtract each variable's mean and divide by its standard deviation prior to analysis in order to homogenise dimensionality (Borcard et al. 2011).

Where pairs of variables showed evidence of strong correlation (r>0.75 or ,<-0.75) one of the variables was excluded to avoid biasing the results (Björklund 2019). I also used the package mass v.7.3.54 to assess how well individual specimens could be classified into their respective clades using a linear discriminant analysis (LDA), which was trained on 70% of the data. The results of the PCA and LDA were both plotted using the ggplot2 package v.3.3.5 (Wickham 2016) in R, with ellipses depicting 95% confidence added using the stat ellipse function. To assess levels of difference in each character among the separate clades, an analysis of variance (ANOVA) was performed for each variable using the aov function in R (R Core Team 2018). Where the results of the ANOVA indicated that there were significant differences among clades, a Tukey test was conducted to compare each pair of clades using the TukeyHSD function, as well as a bonferoni correction using the pairwise.t.test function.

RESULTS

Redundancy of Loci

Aligning both sets of loci to the *I. trifida* genome with bwa mem revealed evidence of some redundancy. In some cases, multiple assembled loci mapped to identical positions more or less contiguous with the same *I. lacunosa* target locus on the *I. trifida* genome (Fig. 2A-B). In other cases, two assembled loci might align to neighbouring areas of the genome, but were still more or less covered by a single *I. lacunosa* target locus (Fig. 2C). These duplications were likely caused by the use of the only partially assembled *I. lacunosa* scaffolds in the initial design of the target capture probes. In all cases the assembled loci shared the same prefix based on their corresponding *S. lycopersicum* target locus, but possessed different suffixes, suggesting some recognition that they were derived from the same original target locus (Fig. 2A-C). Fortunately, this redundancy did not have an impact on the topological results of earlier phylogenetic analyses (Muñoz-Rodríguez *et al* 2018; pers. comm. Dr Pablo Muñoz-Rodríguez). In order to proceed with new analyses using the best possible set of loci, I selected the 249 target loci with unique matches, and the 139 which had received multiple matches, renaming them to match their *S. lycopersicum* counterparts and discarding the suffixes. This required discarding 217 of the 605 assembled loci, leaving 388 loci that are likely single-copy.

HybPiper denovo assembly results

Sequences of all 388 newly verified loci were successfully assembled by HybPiper for all 107 samples (Fig. 3). For the majority of these loci, the assembled sequences covered over 75% of their target length. The mean number of reads mapped for each sample was 2.5million, with a maximum of nearly 10million in one sample of hexaploid *I. batatas* (CIP400218) and a minimum of 297,832 for outgroup species *I. lactifera* J.R.I.Wood & Scotland.

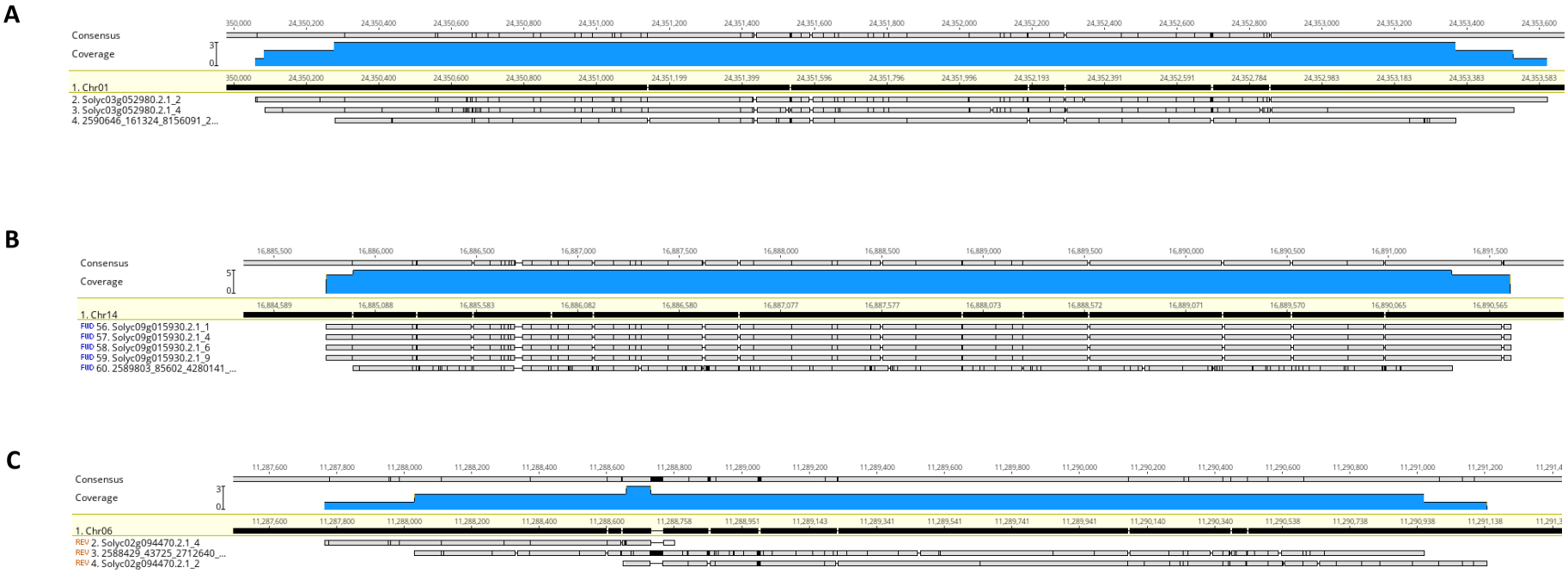


Figure 2: Showing putatively single copy loci aligned to the *I. trifida* genome assembly. **A** and **B** show multiple assembled loci aligned to the same area as a single target locus, and **C** shows two assembled loci aligned to contiguous regions of the *I. trifida* genome, covered by a single target locus.

Number of reads mapped was at or below the lower quartile for the majority of newly-sequenced samples with the exception of *I. tabascana* (PI518479) and the Ecuadorian sample DLP5284, which had been sequenced at higher coverage for planned genome assembly (Fig. 3).

A heatmap of assembled locus lengths for each sample (Fig. 4) also shows that the majority of loci were recovered at or above their target length and that assembled locus length was relatively uniform across samples, regardless of number of reads mapped or sequencing method (HTS vs WGS). Vertical turquoise lines indicate loci with low sequence length compared to the target locus, reflecting poor assembly across samples. Three loci (Solyc01g100500.2.1, Solyc02g087240.2.1 & Solyc12g010600.1.1) received potential paralog warnings from HybPiper's pararlog investigator and were excluded from all subsequent analyses.

BCFTOOLS Genotyping Statistics

The results of the genotyping pipeline showed considerable differences in mean coverage depth across loci, as well as the number of loci with >20X coverage (Fig. 5). Four samples had fewer than 50% of loci with >20X coverage. Aside from outgroup species *I. lactifera*, these included three tetraploid samples from Ecuador (DLP5296, DLP5282 and DFA7817), suggesting a difference in locus coverage between sequencing methods. The three other tetraploid samples from Ecuador, and K500 from Colombia all had >90% of loci with >20% coverage however, as did the newly sequenced Mexican tetraploids and hexaploid K123.

The heatmap of relative coverage also reflects this variability across samples and loci (Fig. 6), showing samples with both low and high coverage across ingroup taxa and sequencing methods.

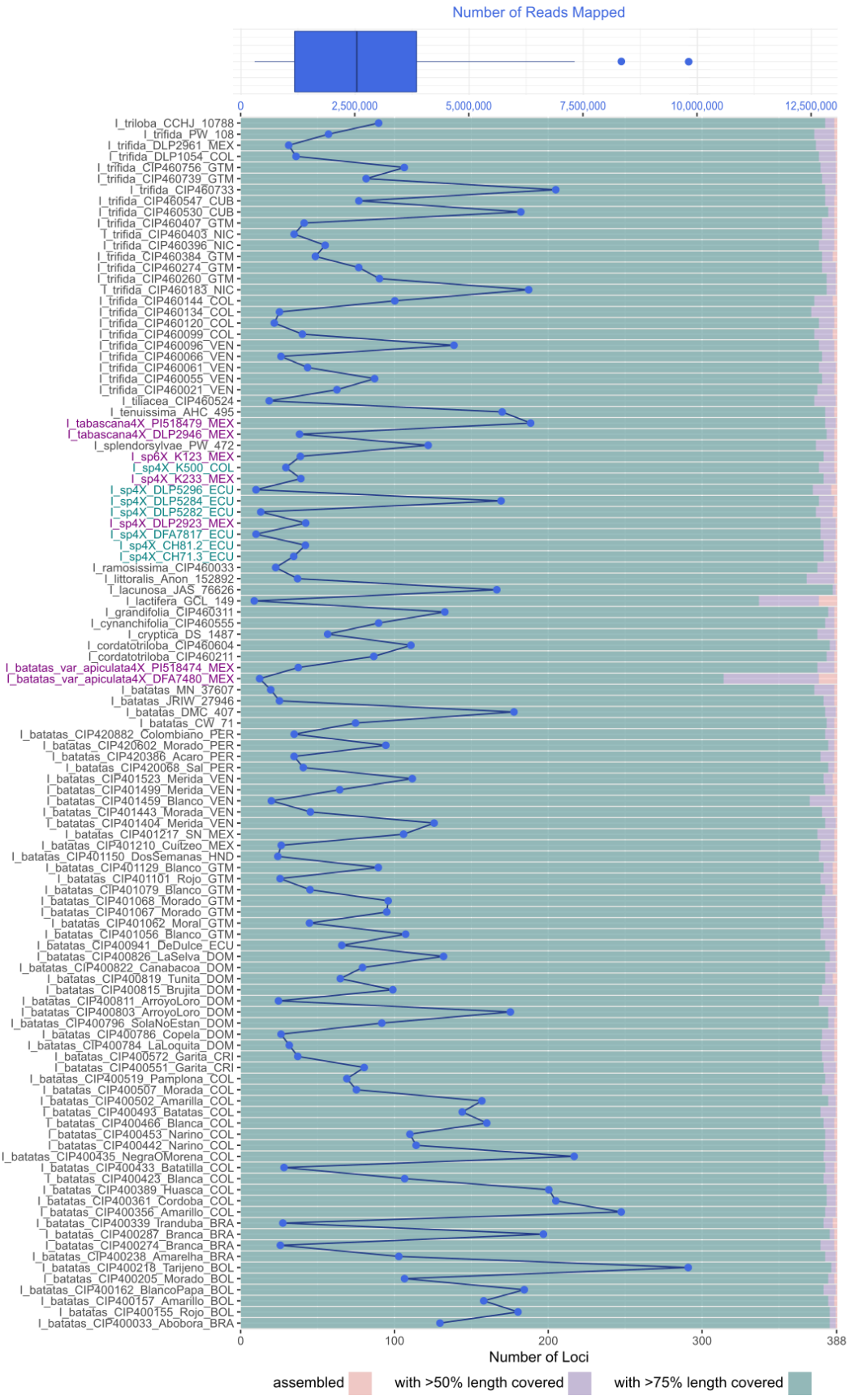


Figure 3: Number of reads mapped and loci assembled for the HybPiper denovo assembly pipeline. Blue boxplot and line graph show total number of reads mapped to all 388 loci for each sample. Stacked barchat shows total number of loci assembled (red), number of loci with assembled sequences covering over 50% (purple) and 75% (green) of their length.

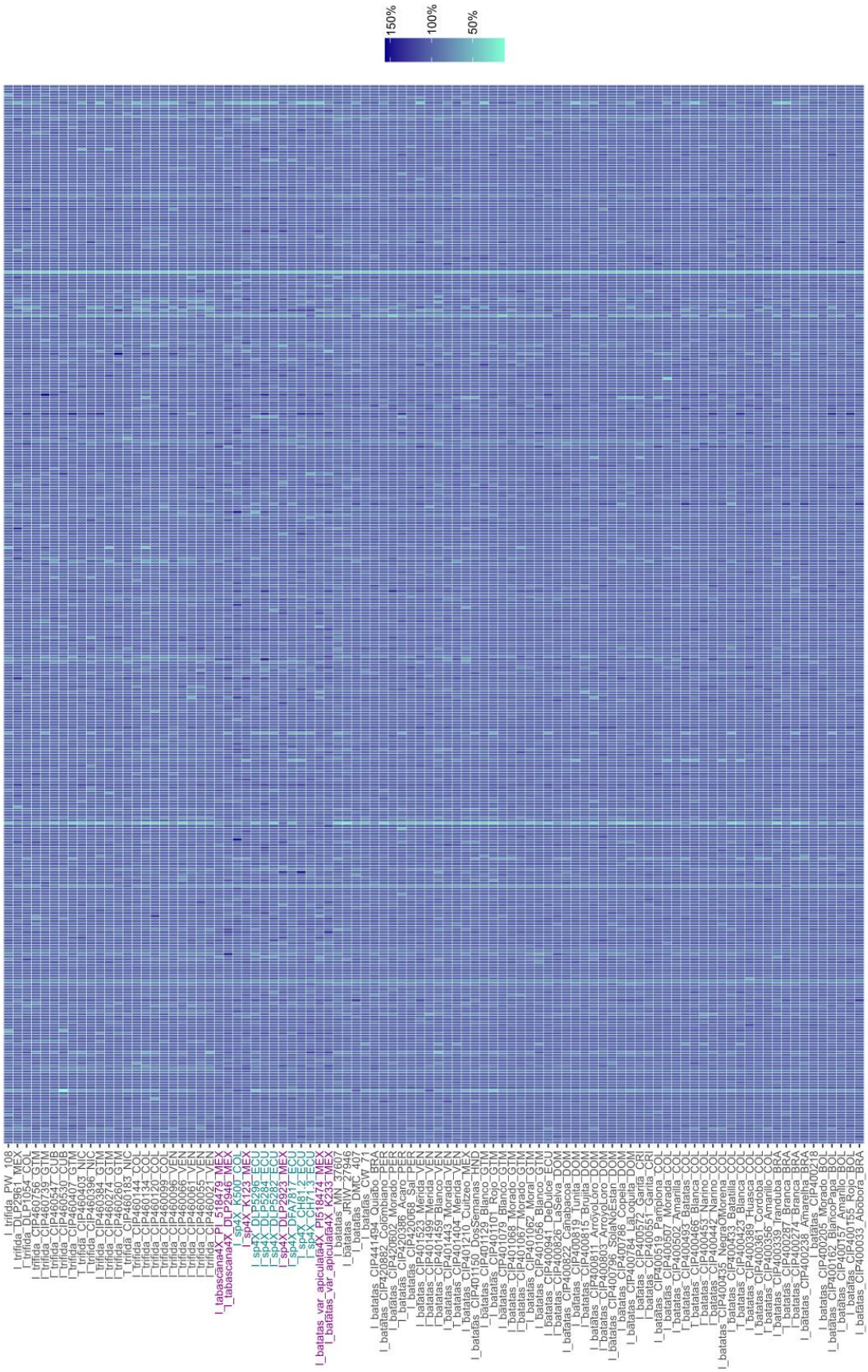


Figure 4: Heatmap showing relative length of sequence assembled by HybPiperfor each of the 388 loci for each ingroup sample as a percentage of the length of the original target locus. Higher percentages are represented by darker blues and lower proportions by turquoise.

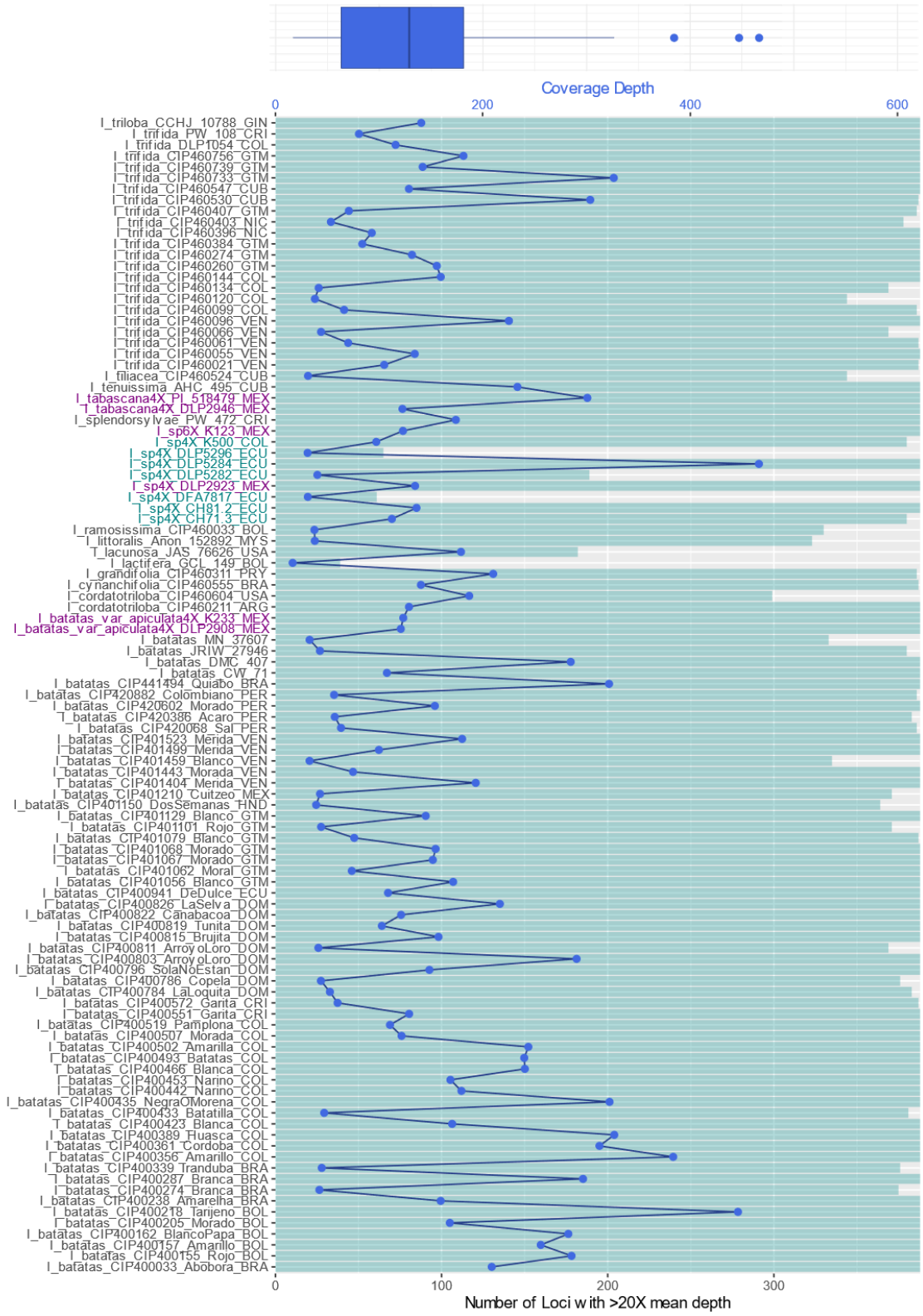


Figure 5: Barchart showing number of genotyped loci with mean coverage >20X, and boxplot and line graph showing mean coverage across all loci.



Figure 6: Heatmap showing relative depth of coverage for each sample at each genotyped locus. Darker blues represent high values, and paler tourquoises lower ones.

Coalescent-based species trees in ASTRAL-III

SUPPORT VALUES: { Quartet Scores (q1/q2/q3) | 0.38/0.34/0.28
Posterior Probability | 0.81 [Posterior Probability=1 ●]

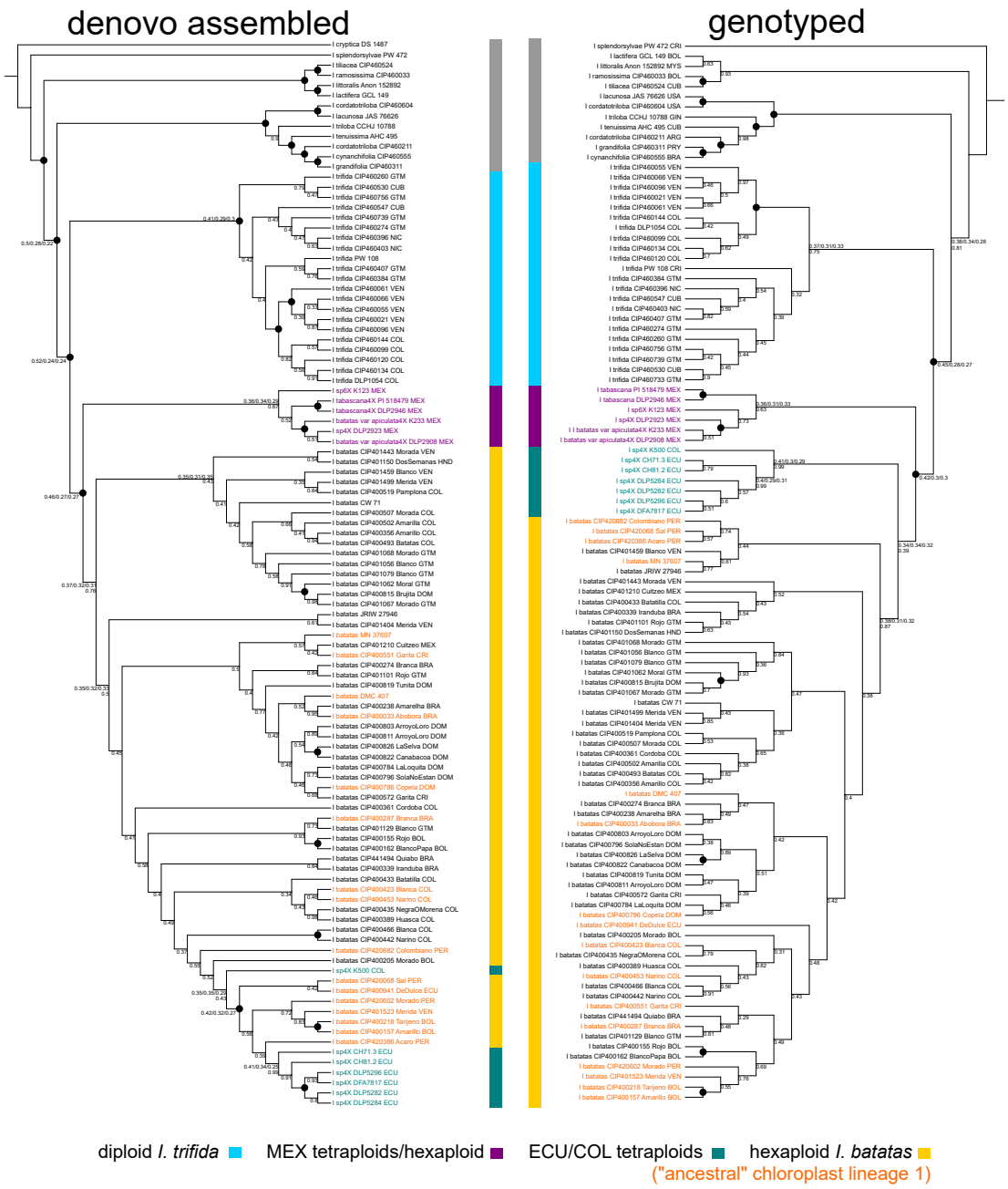


Figure 7: Coalescent-based species trees for both assembly methods (LHS:denovo and RHS:genotyping) with full sampling. Posterior probabilities provided at all nodes, plus quartet scores for major clades. Black circles indicate posterior probability of 1. Both trees recover a well-supported ingroup clade (PP:1/1), containing a well-supported monophyletic diploid *I. trifida* (PP:1/0.75), sister to a moderately- to poorly-supported clade containing hexaploid *I. batatas* and all tetraploid accessions (PP:0.76/0.39). Mexican tetraploids and hexaploid K123 are recovered monophyletic in both trees (PP:0.67/0.63) sister to hexaploid *I. batatas* and tetraploids from Ecuador and Colombia (PP:1/1). Ecuadorian tetraploids are also monophyletic in both trees (PP:0.99/0.99), but are nested within hexaploid *I. batatas* in the denovo assembled analysis (PP:0.76), or sister to it in the genotyped analysis (PP:0.39). Seven hexaploid *I. batatas* accessions with a Lineage 1 chloroplast (in orange) are recovered in a strongly supported subclade with the Ecuadorian material in the denovo analysis (PP:1), but are scattered throughout hexaploid *I. batatas* in the genotyped analysis.

Phylogenetic Analysis of Denovo-Assembled Loci

Coalescent-based phylogenetic analysis of the denovo-assembled loci recovered a well-supported ingroup clade containing diploid *I. trifida*, hexaploid *I. batatas* and all tetraploid accessions (PP=1; QS=0.52/0.25/0.24) (Fig.7). Diploid *I. trifida* was recovered as monophyletic (PP=1; QS=0.41/0.29/0.3), sister to a well-supported clade containing all accessions of hexaploid *I. batatas* and all analysed tetraploid germplasm accessions (PP=1; QS=0.46/0.27/0.27). Within this clade, the tetraploid material was recovered in several separate places in relation to hexaploid *I. batatas*. Wild-collected hexaploid accession K123 and the tetraploid germplasm collections from Mexico, including the suspected hybrid *I. tabascana* (PI518479) were recovered monophyletic (PP=0.67; QS=0.36/0.34/0.29), sister to a clade containing all hexaploid *I. batatas* accessions and all tetraploid accessions from Ecuador and Colombia (PP=0.76; QS=0.37/0.32/0.31).

Within the clade of Mexican accessions, hexaploid K123 was recovered sister to a clade containing *I. tabascana*, *I. batatas* var. *apiculata*, K233, and DLP2923 with low support (PP=0.52). The latter three samples were recovered monophyletic with strong support (PP=1). Meanwhile the Ecuadorian and Colombian tetraploids were recovered nested within hexaploid *I. batatas*. The Ecuadorian tetraploids themselves formed a well-supported clade (PP=0.99; QS=0.41/0.34/0.25), nested within another well-supported clade with seven hexaploid *I. batatas* accessions from Ecuador, Colombia, Venezuela, Peru and Bolivia (PP=1; QS=0.42/0.32/0.27), all of which possessed a Lineage 1 chloroplast haplotype (see below). The tetraploid Colombian specimen was recovered sister to this clade containing the Ecuadorian tetraploids and the seven hexaploid *I. batatas* accessions from northwest South America, but with low support (PP=0.43; QS=0.35/0.35/0.29).

Coalescent-based species trees in ASTRAL-III

SUPPORT VALUES: { Quartet Scores (q1/q2/q3) | 0.38/0.34/0.28
Posterior Probability | 0.81 [Posterior Probability=1 ●]

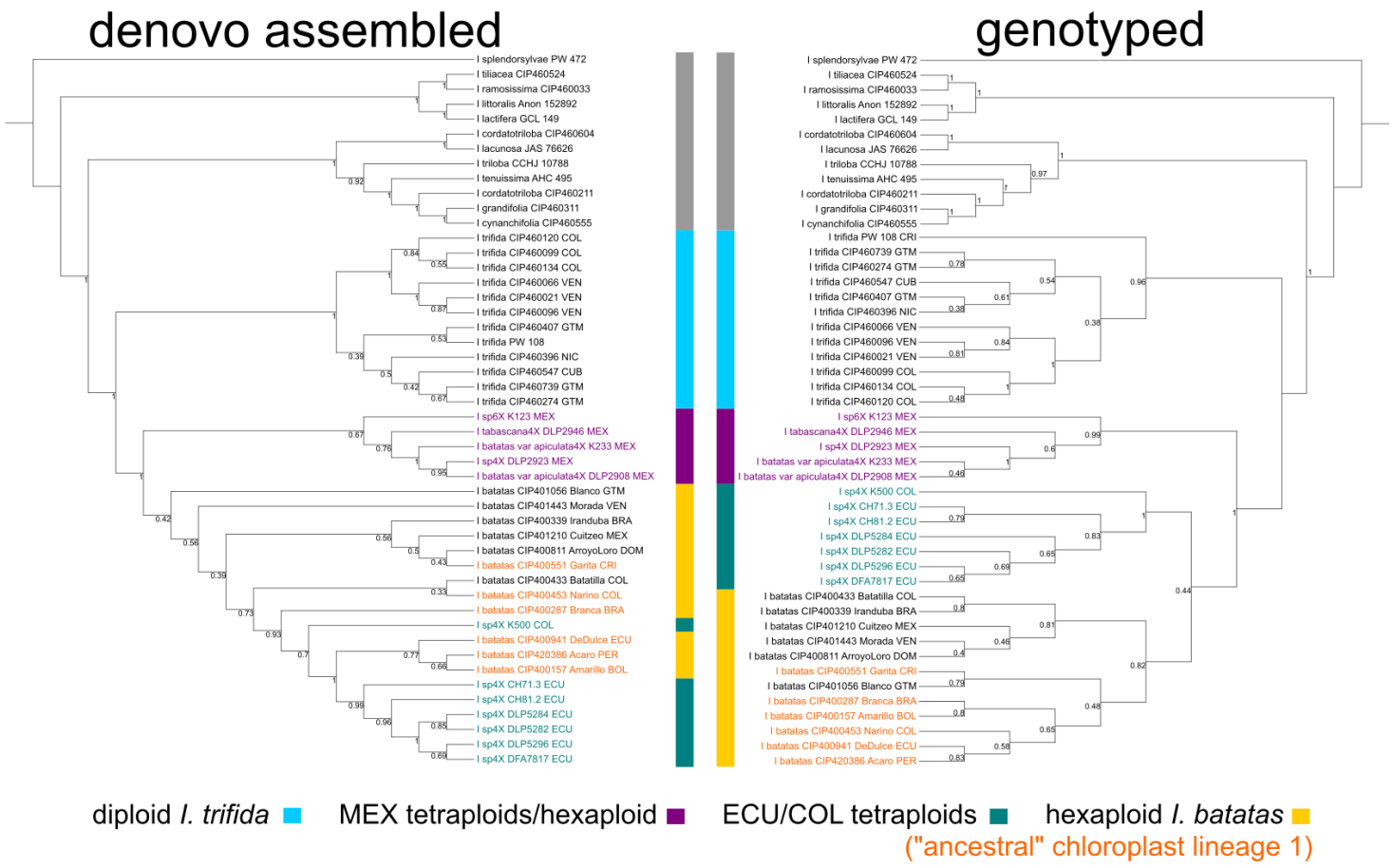


Figure 8: Coalescent-based species trees for both assembly methods (LHS:denovo and RHS:genotyping) with reduced sampling. Showing more or less identical topology of relationships to equivalent analyses with full-sampling in Fig. 7, but with increased support at key nodes in the genotyped analysis.

The topology of relationships within the rest of hexaploid *I. batatas* was otherwise generally unresolved and poorly supported (PP<0.6; QS1<0.35). A similar topology was recovered for the same analysis of the reduced sampling denovo-assembled dataset (Fig. 8).

Support for the ingroup clade, for a monophyletic diploid *I. trifida*, and for the clade containing all tetraploid and hexaploid samples remained high (PP=1), as did support for the clade of Ecuadorian tetraploids, and their position within a clade of hexaploid *I. batatas* accessions from northwest South America (PP=0.99-1), while support for the wider clade containing all hexaploid *I. batatas* and the tetraploids from Ecuador and Colombia remained relatively low (PP=0.42), as did support for the clade containing hexaploid K123 and tetraploid accessions from Mexico (PP=0.67).

Phylogenetic Analysis of Genotyped Loci

Coalescent-based phylogenetic analysis of the genotyped loci full-sampling dataset also recovered a well-supported ingroup clade of diploid *I. trifida*, hexaploid *I. batatas*, and all sampled tetraploid material (Fig. 7; PP=1; QS=0.45/0.28/0.27). Diploid *I. trifida* was again recovered monophyletic (PP=0.75; QS=0.37/0.31/0.33) and sister to a clade containing all tetraploid and hexaploid accessions (PP=1; QS=0.42/0.3/0.3), within which hexaploid K123 and tetraploid germplasm accessions from Mexico were once more recovered monophyletic (PP=0.63; QS=0.36/0.31/0.33), sister to a clade containing all accessions of hexaploid *I. batatas* and all tetraploid accessions from Ecuador and Colombia (PP=0.39; QS=0.34/0.34/0.32).

Within this final clade however, the topology differed from that of the analysis of denovo-assembled loci, with the tetraploids from Ecuador and Colombia recovered together in a separate clade (PP=0.99; QS=0.41/0.3/0.29) sister to a monophyletic hexaploid *I. batatas* (PP=0.87; QS=0.38/0.31/0.32), instead of nested within it as in the results for the denovo-assembled dataset. This topology was again replicated in the reduced sampling analysis of genotyped loci (Fig. 8), but with increased support for the clade of Mexican tetraploids and hexaploid K123 (P=0.99), and for the sister relationship between hexaploid *I. batatas* and the tetraploids from Ecuador and Colombia (P=0.44).

Genomic Differentiation

The Genomescope2 results comparing nucleotide heterozygosity in k-mer spectra from filtered, unassembled sequencing reads recovered a likely genome length of between 344 and 624Mb for the seven successfully modelled samples (Fig. 9). This broad range likely indicates relatively poor accuracy in the lower coverage samples.

Homozygosity was estimated between 88.1% and 96.4%, and the read error rate was between 0.07% and 0.42% (Appendix 3). For all Mexican samples analysed, the proportion of heterozygous sites with an aabb structure averaged 9-9.72%, compared to 0.01% for aaab (Fig. 9). The model failed to accurately converge with the data for all but one of the Ecuadorian samples analysed (DLP5284), which had been sequenced at significantly higher coverage for genome assembly. The proportion of aabb for DLP5284 was 1.18%, compared to 2.29% for aaab (Fig. 9).

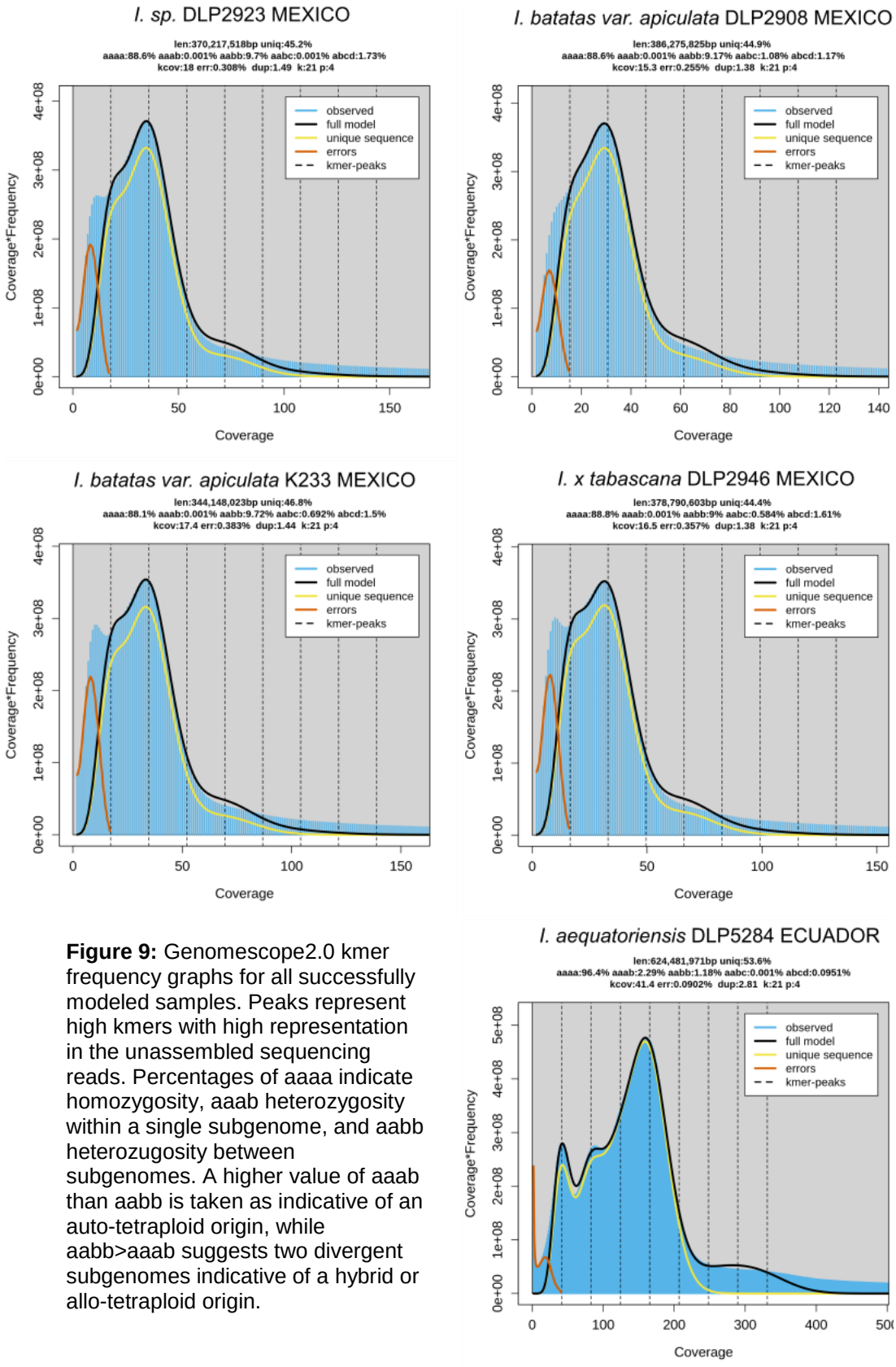


Figure 9: Genomescope2.0 kmer frequency graphs for all successfully modeled samples. Peaks represent high kmers with high representation in the unassembled sequencing reads. Percentages of aaaa indicate homozygosity, aaab heterozygosity within a single subgenome, and aabb heterozugosity between subgenomes. A higher value of aaab than aabb is taken as indicative of an auto-tetraploid origin, while aabb>aaab suggests two divergent subgenomes indicative of a hybrid or allo-tetraploid origin.

Phylogenetic Analysis of Chloroplast Sequence Data

HybPiper successfully assembled sequences for the majority of samples for each of the 56 selected chloroplast loci. Three of these loci (VR24, ndhG and rpl2) received paralog warnings for multiple samples. This suggests an issue with assembly given that paralogs are extremely unlikely in the chloroplast. These regions were therefore excluded from subsequent analyses, while psaA was not successfully assembled for a large number of missing samples, so was also excluded.

Phylogenetic analysis of concatenated sequences for the remaining 52 chloroplast loci recovered three main ingroup clades with high support (BS:99-100%) (Fig. 10), similar to those recovered in previous studies (Roullier et al. 2013; Muñoz-Rodríguez et al. 2018). These clades represent diploid *I. trifida*, and two separate lineages of hexaploid *I. batatas*, labelled “Lineage 1” and “Lineage 2” (Muñoz-Rodríguez et al. 2018), or “Southern” and “Northern” (Roullier et al. 2013).

As in previous studies (Muñoz-Rodríguez et al. 2018), *I. batatas* Lineage 2 was recovered sister to *I. trifida*, with high support (BS=94%). The tetraploid samples from Ecuador were recovered in a grade at the base of *I. batatas* Lineage 1 (BS=99%), or nested within it (BS=90%) (Fig. 10).

Meanwhile, the Mexican tetraploids including the putative hybrid *I. tabascana* as well as the Colombian tetraploid (K500) were recovered in a grade at the base of *I. batatas* Lineage 2 (BS=100%). Hexaploid accession K123 was recovered nested within Lineage 2 (BS=99%) (Fig. 10).

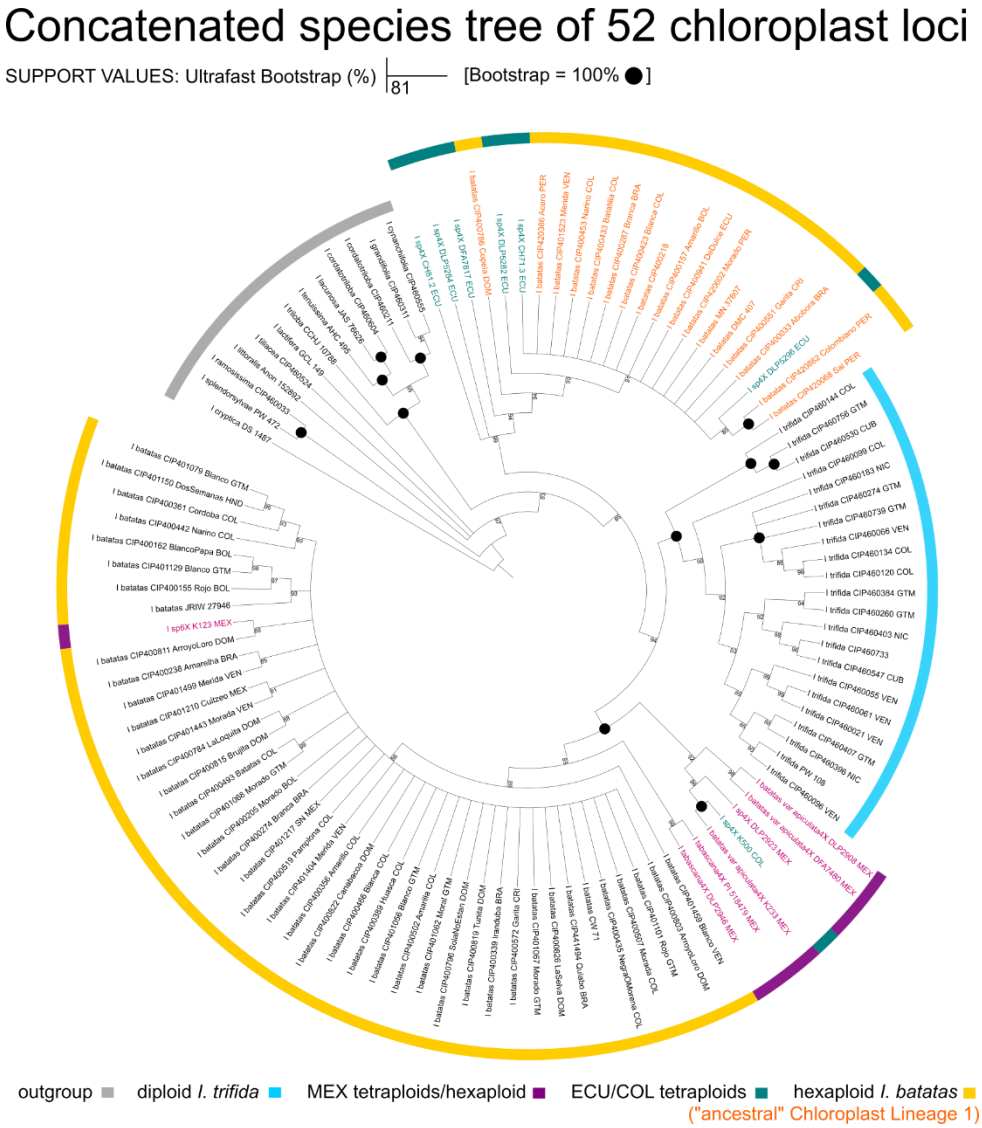


Figure 10: Species tree generated from concatenated sequences of 52 chloroplast loci, showing the two chloroplast lineages (yellow band), one of which is recovered sister to diploid *I. trifida* (pale blue band) with strong support (94%). Ecuadorian tetraploids (green band) are recovered with *I. batatas* Lineage 1 (orange text), the proposed "ancestral" lineage of *I. batatas*, while those from Mexico (purple band) and Colombia (green band) are recovered with *I. batatas* Lineage 2, which is believed to have originated through asymmetric intorgession and chloroplast capture with *I. trifida*.

Morphometrics

All but one of the 13 morphological characters analysed showed evidence of highly significant ($p<0.001$) variation between means in the one-way ANOVA (Table 2A). The only character to show no evidence of significant variation between means was pedicel length. Two variables (lamina width and outer sepal length) were found to be strongly correlated with other variables ($t>0.75$ or $t<-0.75$, $p<0.001$) (Figs. 11&12), and were therefore excluded from the later clustering analyses.

A	MEAN MEASUREMENTS					
Character	**I. trifida**	**4X (ECU/COL)**	**4X (MEX/GTM)**	**I. batatas**	**F**	**p**
Internode diameter (mm)	1.362931	2.0582	1.32	3.0652	**71.519**	**< 2.2E-16 *****
Internode length (cm)	7.224138	9.5395	5.5	6.24	**13.687**	**9.069E-08 *****
Lamina length (cm)	5.017241	8.9625	3.5	9.42	**44.536**	**< 2.2E-16 *****
Lamina width (cm)	4.224138	7.625	3.4	9.08	**42.738**	**< 2.2E-16 *****
No. lamina lobes	2.827586	1.65	5.4	3.16	**15.338**	**1.509E-08 *****
Pedicel length (mm)	8.025862	7.231825	7.312	8.534	0.9012	0.4427
No. flowers/inflorescence	4.566552	14.425	2.4	8.52	**47.764**	**< 2.2E-16 *****
Shortest sepal length (mm)	6.505	5.87165	7.502	9.3684	**42.437**	**< 2.2E-16 *****
Longest sepal length (mm)	9.290345	7.85535	10.178	11.2264	**27.763**	**8.371E-14 *****
Diff. sepal length (%)	29.584948	25.269965	25.11083	16.49034	**12.998**	**1.944E-07 *****
Sepal apex angle (o)	44.827586	102.75	118	74.8	**51.933**	**< 2.2E-16 *****
Corolla length (mm)	30.536379	38.095075	35.804	39.8004	**51.933**	**< 2.2E-16 *****
Corolla tube width (mm)	6.434828	5.049275	9.78	9.3168	**45.49**	**< 2.2E-16 *****

B	DIFFERENCE IN MEANS (TUKEY/BONFERRONI)					
Character	**2Xtrifida-4X(ECU/COL)**	**2Xtrifida-4X(MEX/GTM)**	**2Xtrifida-6Xbatatas**	**4x(ECU/COL)-4X(MEX/GTM)**	**4x(ECU/COL)-6Xbatatas**	**4x(MEX/GTM)-6Xbatatas**
Internode diameter (mm)	**0.69526897***/***	-0.04293103	**1.70226897***/***	-0.7382*/*	**1.007***/***	**1.7452***/***
Internode length (cm)	**2.3153621***/***	-1.7241379	-0.9841379	-4.0395**/**	**-3.2995***/***	0.74
Lamina length (cm)	**3.945259***/****	-1.517241	**4.402759***/***	**-5.4625***/***	0.4575	**5.92***/***
Lamina width (cm)	**3.4008621***/***	-0.8241379	**4.8558621***/***	**-4.225***/***	1.455*/*	**5.68***/***
No. lamina lobes	**-1.1775862***/***	**2.5724138***/***	0.3324138	**3.75***/***	**1.51***/***	-2.24**/**
Pedicel length (mm)						
No. flowers/inflorescence	**9.858448***/***	-2.166552	**3.953448***/***	**-12.025***/***	**-5.905***/***	6.12*/*
Shortest sepal length (mm)	-0.63335	0.997	**2.8634***/***	1.63035*/*	**3.4967***/***	1.8664*/*
Longest sepal length (mm)	**-1.4349948***/***	0.8876552	**1.9360552***/***	2.3226500**/**	**3.3710500***/***	1.0484000
Diff. sepal length (%)	-4.41498324	-4.47412170	**-13.09460916***/***	-0.05913846	**-8.67962592***/**	-8.62048746
Sepal apex angle (o)	**57.92241***/***	**73.17241***/***	**29.97241***/***	15.25	**-27.95000***/***	-43.20000**/**
Corolla length (mm)	**7.558696***/***	5.267621	**9.264021***/***	-2.291075	1.705325	3.996400
Corolla tube width (mm)	**-1.385553***/***	**3.345172***/***	**2.881972***/***	**4.730725***/***	**4.267525***/***	-0.463200

Table 2: Showing the results of ANOVAs (**A**) comparing each character between the four clades (*I. trifida*, 4X(ECU/COL), 4X(MEX/GTM) & *I. batatas*), with mean measurements for each clade, F and p values, and degree of significance (*p=<0.05, **p=<0.01, ***p=<0.001); and (B) the results of Tukey Tests comparing each character between each pair of clades, with the difference in means for each pair, and degree of significance (*p=<0.05, **p=<0.01, ***p=<0.001)

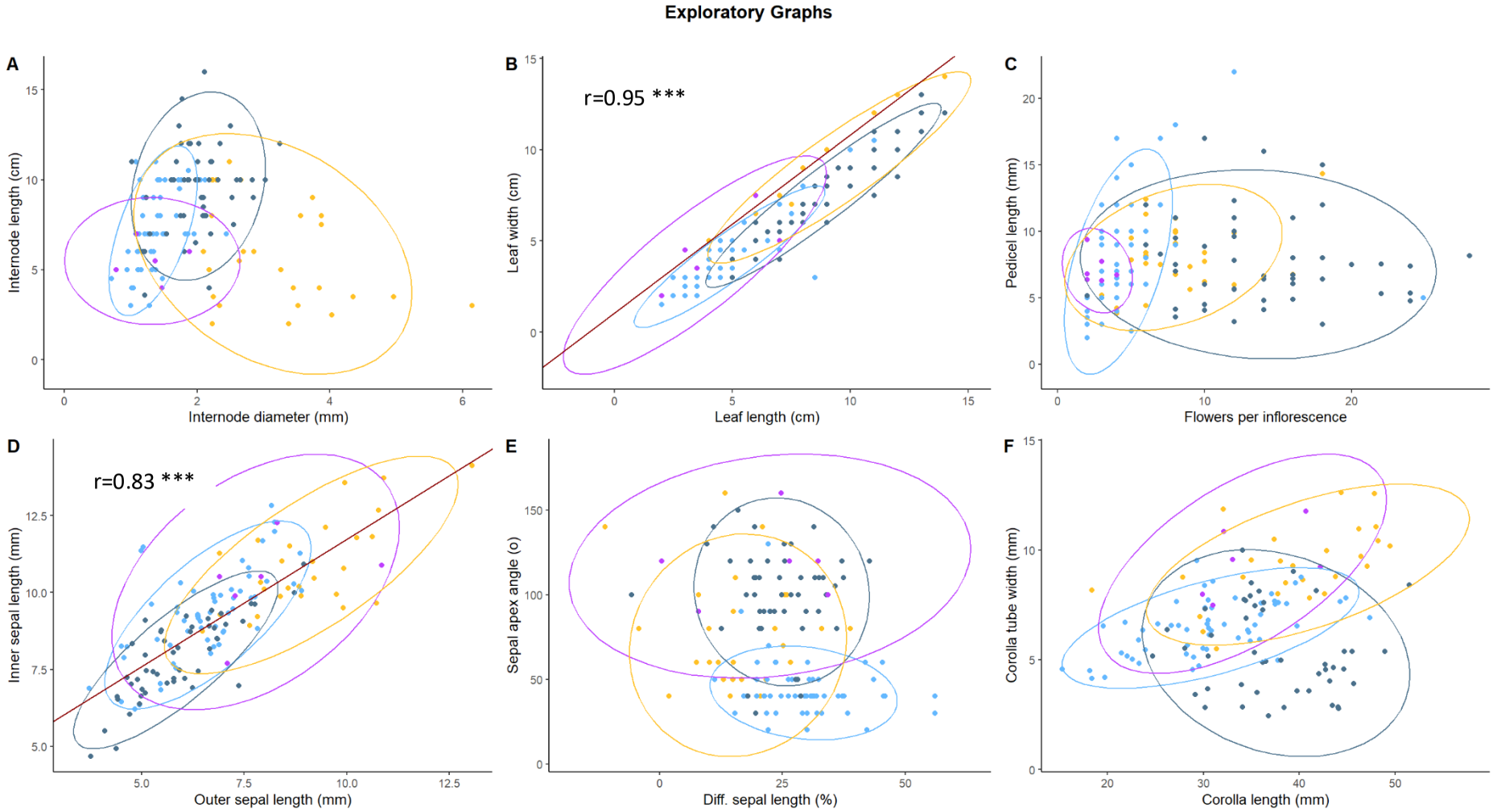


Figure 11: Showing exploratory scatter plots for each pair of potentially correlated morphological variables for *I. batatas* (orange) *I. trifida* (pale blue), Ecuadorian and Colombian tetraploids (dark blue), and Mexican tetraploids (pink). Correlations shown with red line for pairs of variables with strong (r>0.75), statistically significant (p<0.05) (**B** & **D**).

In the Tukey tests and Bonferroni corrections for the twelve characters showing significant variation between means, *I. trifida* showed highly significant variation in 10 characters with both *I. batatas* and the Ecuadorian/Colombian tetraploids, but in only three with the Mexican tetraploids (Table 2). The two groups of tetraploids (Ecuadorian and Colombian vs Mexican) showed highly significant variation with each other in five characters, moderately significant ($p<0.01$) variation in two, and weakly significant variation in two ($p<0.05$) (Table 2). The Ecuadorian and Colombian tetraploids showed highly significant variation with hexaploid *I. batatas* in nine characters, and weakly significant variation in one (Table 2). Whereas the Mexican tetraploids showed highly significant variation with *I. batatas* in only three characters, plus two each with moderately or weakly significant variation (Table 2). There was therefore evidence of strongly significant variation in the majority of characters among *I. trifida*, *I. batatas*, and Ecuadorian tetraploids, and strongly or moderately significant variation between the two groups of tetraploids, but not between Mexican tetraploids and *I. trifida* or *I. batatas*.

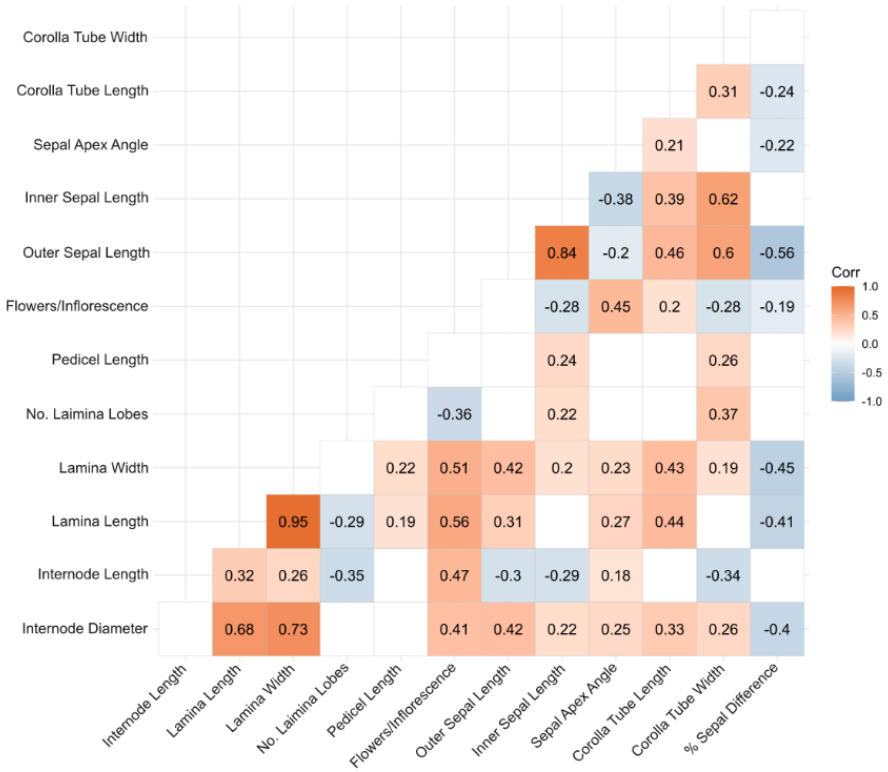


Figure 12: Showing correlation coefficients for all pairs of variables (with darker colours for stronger correlations), and statistically insignificant correlations in white.

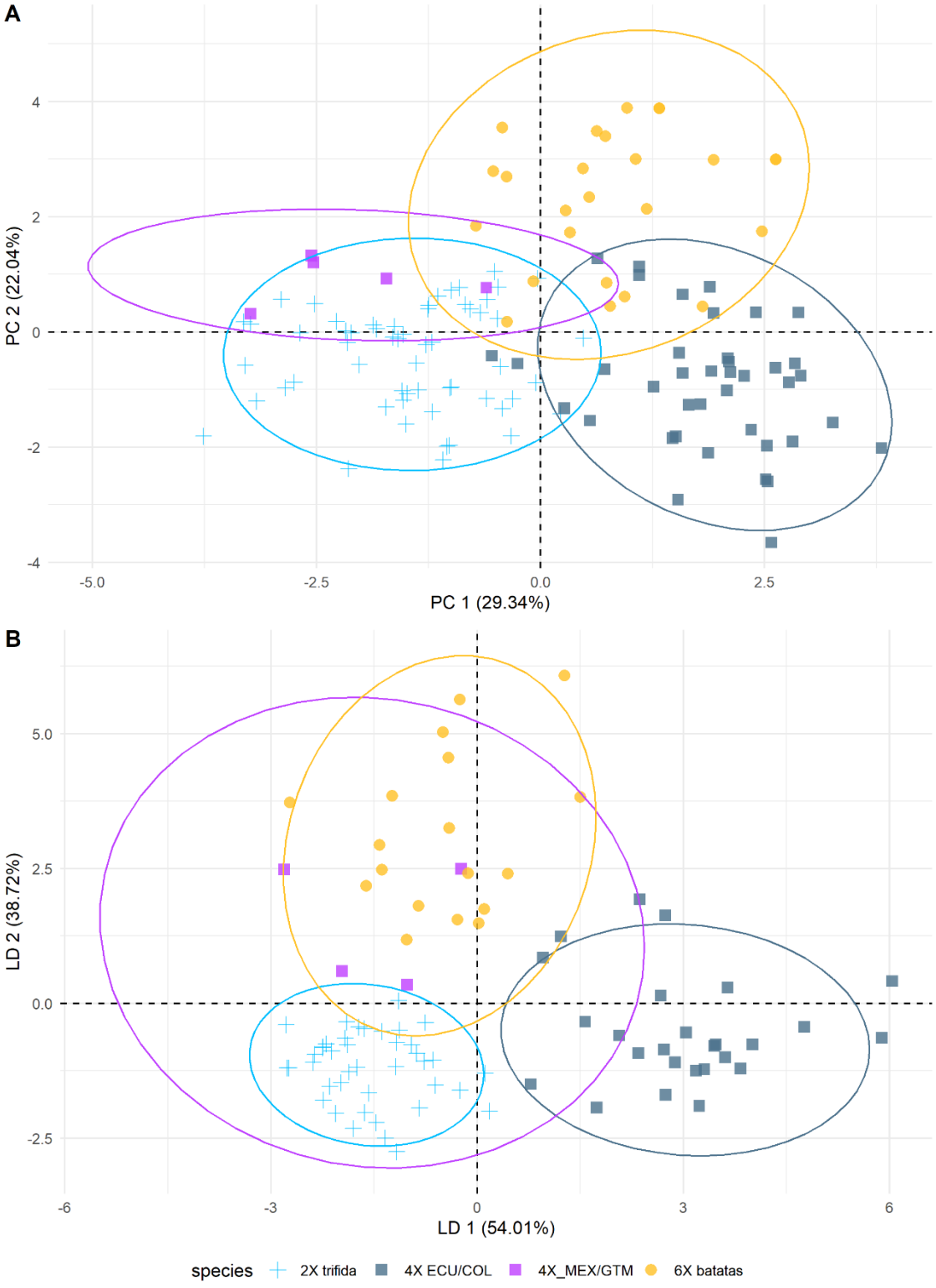


Figure 13: Showing the results of the clustering analyses performed using the 11 uncorrelated variables. In both the principal component analsysis (**A**) and the the linear discriminant analysis (**B**), *I. trifida, I. batatas*, and the tetraploid material from Ecuador and Colombia form relatively distinct clusters, with partial overlaps between 95% concidence interval ellipses. In contrast, the tetraploid samples from Mexico are recovered within the either *I. trifida* (**A**) or *I. batatas* (**B**).

This pattern was also reflected in both the clustering analyses. In the PCA (Fig. 13A) performed using the 11 uncorrelated variables for example, components one and two explained 27.6% and 23.4% of the variation respectively, and recovered three clusters corresponding to diploid *I. trifida*, hexaploid *I. batatas* and tetraploid material from Ecuador and Colombia, with some overlap between 95% confidence intervals of each. Meanwhile, the majority of tetraploid samples from Mexico, including *I. batatas* var. *apiculata* (PI 518474) and *I. tabascana* (PI518479), fell within the clusters of *I. trifida* and *I. batatas* (Fig.13A). The same three clusters were recovered in the LDA, but with less overlap (Fig. 13B), and there was a 97.4% success rate in accurately identifying the test data. This time, the tetraploid plants from Mexico and Guatemala were all recovered within the 95% confidence interval for hexaploid *I. batatas*.

DISCUSSION

Taken together, the results of the different analyses reveal two distinct groups of tetraploids closely related to cultivated hexaploid *I. batatas*, each of which are likely to have originated by different means. The results consistently suggest that those from Ecuador are a distinct group that most likely represents the extant descendants of the tetraploid progenitor of hexaploid *I. batatas*. The exact identity of those from Mexico remains somewhat uncertain, but the combined evidence suggests they are most likely the result of more recent hybridization between hexaploid *I. batatas* and diploid *I. trifida*.

Tetraploids from Colombia appear to be closely related to those from Ecuador, but may also have arisen through hybridization – between the Ecuadorian tetraploids and hexaploid *I. batatas* with a Lineage 2 chloroplast.

Congruence & Incongruence In Nuclear Phylogenies

Phylogenetic analysis based on both denovo-assembled and genotyped nuclear sequence data produced similar topologies in terms of three major ingroup clades. The first of these clades contains all accessions of diploid *I. trifida*, and is itself recovered sister to the two other ingroup clades, one of which contains all the Mexican tetraploids and hexaploid K123, and the other of which contains all accessions of cultivated hexaploid *I. batatas* and all tetraploid accessions from Ecuador and Colombia. The results for the two datasets differed however, with respect to the position of the Ecuadorian and Colombian tetraploids relative to hexaploid *I. batatas*. In the denovo-assembled analysis, the Ecuadorian and Colombian tetraploids were paraphyletic, nested within hexaploid *I. batatas*, within a subclade of predominantly northwest South American hexaploid *I. batatas* accessions with a Lineage 1 chloroplast (Fig. 7). Whereas in the genotyped analysis they were recovered together in a separate clade, sister to a monophyletic hexaploid *I. batatas*.

The position of the Ecuadorian and Colombian tetraploids in relation to hexaploid *I. batatas* thus differed between results from the two assembly methods. By contrast, in the analyses of both the denovo-assembled and genotyped datasets, the hexaploid (K123) and tetraploid accessions collected in Mexico, including the putative tetraploid hybrid *I. tabascana*, were consistently recovered in a clade sister to the clade containing all hexaploid *I. batatas* accessions and the tetraploids from Ecuador and Colombia. Both topologies therefore provide strong evidence of two independent tetraploid clades with distinct origins: one from Mexico, and the other from Ecuador, both of which are monophyletic and distinct both from each other and from diploid *I. trifida*. Colombian tetraploid accession K500 was recovered either in a well-supported clade with the Ecuadorian tetraploids, or nested within hexaploid *I. batatas*, but paraphyletic with respect to the Ecuadorian tetraploids and several

northwest South American accessions of hexaploid *I. batatas*. This suggests a close association with the Ecuadorian tetraploids, but its exact nature will require further research.

The fact that in the analyses of both datasets, the clade of Mexican tetraploids and hexaploid K123 is repeatedly recovered sister to the clade containing all cultivated hexaploid *I. batatas* accessions and the tetraploids from Ecuador and Colombia suggests that the Mexican tetraploids share a similar genomic structure to that of diploid *I. trifida* than either hexaploid *I. batatas* or the Ecuadorian and Colombian tetraploids do. By contrast, the fact that the Ecuadorian and Colombian tetraploids were either recovered nested within hexaploid *I. batatas*, or in a separate clade sister to it, suggests these accessions are more closely related to hexaploid *I. batatas* than the Mexican tetraploid accessions are.

At the same time, the results also underline the close relationships and potentially complex evolutionary history within this group of taxa. These two factors make resolving and accurately interpreting the exact phylogenetic relationships among them extremely challenging, thanks to the difficulty of reconstructing past hybridization, geneflow and polyploidy in dichotomously branching trees (Fig.1). For example, conventional cladistic interpretation of the position of the clade containing the Mexican tetraploids and hexaploid K123 sister to that of hexaploid *I. batatas* and the Ecuadorian and Colombian tetraploids would suggest that these two clades share a recent common ancestor. This is unlikely to be the case for *I. tabascana* however, since it is thought to have originated through recent hybridization between hexaploid *I. batatas* and diploid *I. trifida* (Srisuwan et al. 2006). Its position (and by extension, that of the other Mexican tetraploids recovered with it) is therefore more likely to be attributable to its possession of genetic material inherited from both diploid *I. trifida* and hexaploid *I. batatas*, leading to a position intermediate between the two (Fig.1) (Rajapakse et al. 2004; Cao et al. 2014; Muñoz-Rodríguez et al.

2018). Wild-collected Mexican hexaploid accession K123, which was recovered within the clade of Mexican tetraploids, could also be a hybrid between *I. trifida* and *I. batatas*, since experimental data shows these crosses can produce hexaploids as well as tetraploids (Orjeda et al. 1991).

The position of the Ecuadorian tetraploids nested within hexaploid *I. batatas* in the results of the denovo-assembled analyses is also difficult to interpret, since it appears to imply that these tetraploids are derived form a hexaploid ancestor. While aneuploidy can lead to the loss of one or more chromosomes, the only mechanism leading to a wholesale reduction in ploidy from hexaploid to tetraploid is through hybridization with a diploid, as described for *I. tabascana* and the other Mexican tetraploids above. This is unlikely to be the case in this instance however, since the tetraploids are nested within the hexaploids, whereas hybridization with a diploid relative would most likely lead to a sister relationship, as seen here and in previous studies for *I. tabascana* and other artificially generated hybrids (Rajapakse et al. 2004; Cao et al. 2014; Muñoz-Rodríguez et al. 2018). As a result, the only plausible explanation for this topology is that it is a result of an artefact in the denovo assembly method. In addition to the challenges of representing reticulate evolution in dichotomously branching phylogenetic trees, the results therefore also illustrate the difficulty of reconstructing phylogenetic relationships among mixed ploidy taxa using consensus sequences.

The genotyped analysis explicitly takes into account heterozygosity, which has been shown to offer better phylogenetic resolution in closely related taxa (Lischer et al. 2014), and is likely to be especially important in this case, given that *I. batatas* is believed to be an allohexaploid (AABBBB) that arose from a diploid (AA) and a tetraploid (BBBB) progenitor (Yang et al. 2017). In the denovo-assembly pipeline, alleles from hexaploid *I. batatas*' smaller subgenome (AA) are likely to be discarded in favour of the more numerous alleles from its pair of identical

subgenomes (BBBB). This would lead to the assembly of loci resembling the tetraploid progenitor of *I. batatas*, at the expense of alleles inherited from its diploid progenitor.

If the Ecuadorian tetraploids are the extant descendants of the tetraploid progenitor of the two identical subgenomes possessed by *I. batatas* (BBBB), then this could therefore explain their recovery within the hexaploid *I. batatas* clade in the results of the denovo-assembly-based analysis. By contrast, if the tetraploids from Mexico arose by hybridization between allohexaploid *I. batatas* and the extant descendant of its diploid progenitor (*I. trifida*), they would possess more or less equal amounts of alleles from both the tetraploid progenitor of *I. batatas* and the diploid progenitor or its descendant (*I. trifida*), potentially explaining why they are recovered intermediate between *I. trifida* and *I. batatas*, regardless of the assembly method used. That is, a hybrid between allohexaploid *I. batatas* (AABBBB) and the descendant of its diploid progenitor *I. trifida* (AA), would most likely produce a tetraploid with equal volumes of A and B genome alleles (ABB+A=AABB). Meanwhile, if K123 is a hexaploid hybrid between *I. batatas* and *I. trifida*, then it must have arisen through the production of unreduced gametes by one or both parents (Orjeda et al. 1991), making its resulting genomic composition more complicated (A+AABBB/AA+ABBB?), but still resulting in a genome with a more or less equal combination of A and B genome alleles.

Nonetheless, while the recognition of ambiguous sites in genotyped loci can account for potential heterozygosity between different subgenomes, it is still based on unphased and artificially diploidized data, with only two potential calls (0/1) at each position. It therefore provides no information on relative contribution or dosage of the different subgenomes for polyploid accessions. In addition, during the filtering stage of the genotyping pipeline, potentially heterozygous sites where one allele has fewer than half as many supporting reads as the alternative are set to homozygous.

While this step is arguably necessary to protect against potential sequencing and alignment errors that might introduce artificial heterozygosity, the specific ratio of 1:2 is potentially problematic for hexaploid *I. batatas*, since that is the suspected ratio between its A and B subgenomes. The genotyping pipeline may therefore still be excluding relevant heterozygosity between them.

Ultimately, building a phylogenetic tree that accurately reflects the evolutionary history of hexaploid *I. batatas* would ideally require the phasing of each subgenome for all polyploid taxa, such that they could each be included as separate accessions in the alignment and subsequent phylogenetic analyses. Given that the genome of the tetraploid progenitor of sweet potato has not yet been assembled, and that no subgenome-level assembly exists for hexaploid *I. batatas*, this is currently impossible. It is therefore necessary to use the results of the other analyses to test the hypothesis of phylogenetic relationships outlined above based on the topology observed here.

While the results of the phylogenetic analyses presented here cannot be used to draw definitive conclusions about the relationship of different tetraploid accessions to cultivated hexaploid *I. batatas* and their role in its origin, they do support the following conclusions: (i) that tetraploids from Mexico, Ecuador and Colombia are more closely related to *I. batatas* than closest known relative diploid *I. trifida* or any other species; (ii) that they form separate clades in all analyses, with moderate to strong support; (iii) that the tetraploids from Mexico share greater genomic similarity with *I. trifida* than either hexaploid *I. batatas* or the tetraploids from Ecuador and Colombia are; (iv) that the tetraploids from Ecuador and Colombia are more closely related to hexaploid *I. batatas* than the Mexican tetraploids are; and (v) that the Mexican tetraploids share a highly similar genomic composition to suspected hybrid *I. tabascana*, suggesting they originated by a similar means.

Further support for these conclusions is explored in comparison of the results of the subsequent analyses below for the different clades of tetraploids recovered here.

Genomic structure

The Genomescope2 results comparing nucleotide heterozygosity in k-mer spectra from filtered, unassembled sequencing reads suggest that the representative Ecuadorian tetraploid analysed (PI403553) has a genomic structure strongly consistent with an autopolyploid rather than allopolyploid origin, with proportions of aaab consistently higher than aabb (Appendix 3). This pattern is indicative of two identical or highly similar sub-genomes originating from WGD (Ranallo-Benavidez et al. 2020). The same analysis for suspected hybrid *I. tabascana* (PI 518479) and the three other tetraploid specimens from Mexico (PI518474, DLP2923 and K233) shows instead a higher proportion of aabb than aaab (Appendix 3), suggesting that two distinct sub-genomes have been derived from a recent hybridization event in these collections.

If the Mexican samples arose by hybridization between hexaploid *I. batatas* and diploid *I. trifida* then they are not strictly speaking allo-tetraploids. Nonetheless, the fact that *I. trifida* is believed to be directly descended from the diploid progenitor of hexaploid *I. batatas* means that genetic material inherited from *I. trifida* will likely closely resemble that inherited from the *I. batatas* AA subgenome by any hybrids between *I. trifida* and *I. batatas*. That is, a cross between AABBBB and AA will yield AABB (Fig. 1), which would most likely resemble a true allo-tetraploid in these analyses.

The difference in sequencing coverage between the Ecuadorian and Mexican samples, and the failure of the model to converge in the other Ecuadorian samples with similar coverage to the Mexican samples mean that the results for these samples should still be interpreted cautiously. Closer inspection of the

minimum and maximum values in Table S2, show higher variability for the Mexican samples (which had a lower sequencing coverage) than for the Ecuadorian sample, as well as a lower model fit, higher predicted sequencing error rate and a smaller haploid genome length. This pattern is also reflected in the graphs for Mexican samples (Fig. 8), which show considerably less pronounced peaks than the graph for the Ecuadorian sample.

Nonetheless, the results for the more comprehensively sequenced Ecuadorian sample can be confidently said to indicate an autopolyploid origin. It is not clear why the other Ecuadorian samples with lower sequencing coverage failed to produce useable results, but autopolyploids may present more of a challenge for the program due to lower overall heterozygosity and high similarity of subgenomes.

Chloroplast Phylogeny

The strongly supported position of all Ecuadorian tetraploids in the Lineage 1 clade in phylogenetic analysis of chloroplast sequence data supports the hypothesis that these accessions are closely related to what is considered to be the ancestral lineage of hexaploid *I. batatas.* Likewise, the position of all Mexican tetraploid accessions and K123 with Lineage 2 suggests that they are associated with this supposedly derived lineage.

Naturally, this interpretation is dependent on acceptance of the chloroplast capture hypothesis, further evidence for or against which is not provided by these results. In fact, interpreting the exact nature of the relationship between the tetraploids and their respective *I. batatas* Lineages within the framework of the chloroplast capture hypothesis is somewhat complicated by the exact topologies within each clade.

For example, the Mexican tetraploids are recovered in two well-supported clades, which themselves form a grade sister to all Lineage 2 hexaploid *I. batatas*

samples and hexaploid K123, rather than being nested within Lineage 2, as we might expect if they are indeed recent hybrids between hexaploid *I. batatas* and diploid *I. trifida*. This would seem to imply that these accessions diverged from a common ancestor with *I. batatas* Lineage 2, rather than having arisen through repeated independent hybridization events between *I. batatas* and *I. trifida.* Unlike in the nuclear phylogenies, this sister relationship is unlikely to be explained by a possession of genetic material from each ancestor, given the uniparental inheritance and lack of recombination in the plastid genome. Biparental chloroplast inheritance is not unheard of in plant hybrids (Barnard-Kubow et al. 2017), but were it the case here, we might expect to have found evidence for it in the form of numerous paralog alerts from the HybPiper assembly pipeline.

On the other hand, while some Ecuadorian tetraploids are recovered sister to the hexaploid *I. batatas* samples in Lineage 1, they form a paraphyletic grade, and several other samples are recovered nested within the hexaploid samples in Lineage 1. There are two possible explanations for this topology that can still be reconciled with the Ecuadorian tetraploids being the progenitor of hexaploid *I. batatas*. The first is that there has been insufficient time for divergence between the ancestral hexaploid *I. batatas* lineage and its tetraploid progenitor for it to be visible in the more conserved plastid genome, and the second is that there has been some recent geneflow between the two through, but these explanations could just as easily be applied to the Mexican tetraploids. What is more, the inclusion of the Colombian tetraploid (K500) within the Mexican tetraploid clade sister to Lineage 2 contradicts its position close to or within the Ecuadorian material in the results of the nuclear analyses. It is again possible this could be indicative of a further example of chloroplast capture resulting from a cross between a relative of the Ecuadorian tetraploids and a Lineage 2 hexaploid *I. batatas*, which might also explain its position paraphyletic with respect to hexaploid *I. batatas* and the Ecuadorian

tetraploids in the results for the denovo-assembled nuclear loci (Fig. 7). It underlines the fact that these results are not easily explained within the context of the chloroplast capture hypothesis however, and reinforces the complex nature of the evolutionary history for this group of taxa.

Over all, the results of the phylogenetic analysis of chloroplast sequence data presented here cannot therefore be said to conclusively confirm the hypothesis of relationships put forward based on the results of the nuclear analyses, but nor do they entirely preclude it either. Were it not for the lack of a corresponding topology in the nuclear phylogeny, the topology of the chloroplast phylogeny would in fact seem to be more parsimoniously explained by separate origins for hexaploid *I. batatas* from two different tetraploid progenitor populations, one in Ecuador and the other somewhere between Mexico and Colombia, as was suggested by Roullier et al. (2013) when they designated the two chloroplast lineages “Northern” and “Southern”.

As it is, the aforementioned issues in the assembly and phylogenetic analysis of nuclear loci in mixed-ploidy datasets continue to make this hypothesis difficult to entirely rule out either. There is some evidence for an association between the Ecuadorian tetraploids and a group of Lineage 1 hexaploid *I. batatas* accessions from South America found in the results of the denovo-assembled nuclear analysis, for example (Fig. 7). Seven of the 16 included *I. batatas* Lineage 1 accessions – all of which are from South America – are recovered in a well-supported clade (PP=1; QS=0.41/0.34/0.27) with the Ecuadorian tetraploids. The remaining Lineage 1 accessions are scattered paraphyletic throughout the rest of a poorly resolved *I. batatas* clade however, as all Lineage 1 accessions are in the full-sampling genotyped analysis. In the reduced sampling genotyped analysis this pattern is again reversed, with all but one of the Lineage 1 accessions recovered in a moderately supported clade (PP=0.48) sister to a separate and well-supported

clade of exclusively Lineage 2 samples (PP=0.81). The topology within hexaploid *I. batatas* is therefore poorly resolved and variable dependent on assembly-method and sampling.

What is more, an additional challenge to confirming or discarding a dual-origin hypothesis is the impact of human-mediated translocation and breeding on the contemporary genetic make-up of cultivated crops like sweet potato. This challenge has been recognized in attempts to re-construct the evolutionary history of tomato for example (Razifard et al. 2020), where despite attempts to account for the impact of 20th Century breeding programs by sampling only putative landraces, introgression of alleles believed to be the result of human-mediated selection and breeding still complicated attempts to accurately reconstruct relationships between CWRs and the modern crop.

Sweet potato has been a global crop since at least the 16th century, and before that is likely to have been extensively cultivated and traded within the Americas (Lebot 2008; Loebenstein and Thottappilly 2009). Efforts were made in the sample selection for this study to avoid modern cultivars, but this was done purely on the basis of recorded cultivar name and country of collection. None of the samples used here have detailed accession notes on historic translocation or ancestry, which is all but impossible to ascertain in the present day. It is possible therefore, that if sweet potato did have two origins from separate tetraploid populations, subsequent human crossing between Lineage 1 and 2 cultivars could have led to the erasure of much of the evidence for it in the nuclear genome, potentially explaining its absence in the results presented here and by Muñoz-Rodríguez et al. (2018).

As such, the chloroplast capture hypothesis continues to require definitive confirmation. What the results of the chloroplast analysis do confirm however, is that

the Mexican and Ecuadorian tetraploids do not share a common origin, but are both closely related to hexaploid *I. batatas*, and that the Mexican tetraploids share more genomic similarity to diploid *I. trifida* than the Ecuadorian tetraploids are. Further phylogenetic and population genetic analyses from phased nuclear sequence data will be required to entirely resolve the evolution of hexaploid *I. batatas* from its tetraploid ancestor. This will now be greatly facilitated by the discovery of autotetraploid plants from Ecuador, closely related to hexaploid *I. batatas* in both nuclear and chloroplast analyses, and therefore its most likely tetraploid progenitor.

Morphological & Geographical Coherence

The results of the morphometric analyses suggest that while tetraploid plants from Ecuador and Colombia are morphologically coherent and distinct from both *I. batatas* and *I. trifida*, tetraploid plants from Mexico display a heterogeneous set of morphological characteristics intermediate between the two (Table 2 & Fig.13). This is clear from both the ANOVA and Tukey tests (Table 2) – where the Mexican tetraploids share significantly less variation with any of the three other clades – and the clustering analyses (Fig 13), where they are predominantly recovered within the 95% confidence intervals for either *I. batatas*, or *I. trifida*. What is known about the Mexican collections also suggests they are morphologically and ecologically heterogeneous, again supporting repeated independent hybrid origins, rather than a single origin as might be expected given their monophyly and sister relationship to *I. batatas* in phylogenetic analyses.

One potential limitation of this result is the relatively small number of Mexican samples available for study here. Unlike the Ecuadorian tetraploids, for which numerous citations of related herbarium specimens allowed me to expand the sampling beyond germplasm collections, relatively few samples are recorded for *I. tabascana* and *I. batatas* var. *apiculata* (McDonald and Austin 1990), and much of

the previously cited germplasm material was unavailable, did not germinate or flower, and lacked herbarium vouchers taken at the time of collection. If the Mexican tetraploids do indeed represent recent hybrids, then this may explain their relatively limited distribution and thus collection history compared to their Ecuadorian counterparts. Hybrids are often ephemeral, can suffer from fertility issues, and struggle to establish themselves beyond disturbed habitats (Abbott et al. 2013). Nonetheless, it is also possible that similar material is concealed within herbarium collections of unknown ploidy. Wood et al. (2020) and Contreras et al. (1994) for example, reported numerous wild-growing populations of unknown ploidy, resembling "weedy" sweet potatoes collected in similar areas to known tetraploid germplasm collections (see Chapter 1). Establishing the ploidy of these samples and including them in future analyses would allow us to explore the frequency of hybridization between *I. batatas* and its closest diploid relative, and verify the results presented here.

A New Tetraploid Species Sister to Sweet Potato

Taken together, the results of the phylogenetic, genomic and morphometric analyses show that the tetraploids from Ecuador are autotetraploid, consistently monophyletic and more closely related to hexaploid *I. batatas* than any other taxon in phylogenetic analyses of a comprehensive set of low-copy nuclear loci; and that they are morphologically distinct from their closest relatives, and distributed in a contiguous region in western Ecuador. We therefore described this material as a new species, *Ipomoea aequatoriensis* T.Wells & P. Muñoz (Muñoz-Rodríguez et al. 2022), following the newly developed *heuristic species concept*, which requires evidence of phenotypic, genotypic and ecotypic cohesion (Wells et al. 2021, 2022). A detailed discussion of this decision, along with a diagnosis, detailed description, illustration of key morphological characters, specimen citation and notes on distribution and ecology are provided in the following chapter.

The Most Likely Tetraploid Progenitor of Sweet Potato

Any candidate for the extant descendant of the tetraploid progenitor of sweet potato should meet the following four criteria: (i) it should be monophyletic and more closely related to hexaploid *I. batatas* than any other taxon in phylogenetic analyses based on multiple vouchered specimens and a comprehensive set of genomic loci; (ii) it should be autotetraploid, with no evidence of past hybridization in its genome. This is important because of the apparent AABBBB structure of the hexaploid *I. batatas genome*, which requires that the tetraploid progenitor possess two identical sub-genomes, corresponding to the identical pair (BBBB) inherited by hexaploid *I. batatas* (Fig. 1); (iii) it should possess a chloroplast haplotype similar to the hypothesised ancestral lineage of sweet potato (Muñoz-Rodríguez et al. 2018); and (iv) it should be a clearly wild entity with a well-documented collection history and provenance (i.e., not feral, from cultivation, or derived from modern breeding programs). The newly described *I. aequatoriensis* fits these four criteria. Most importantly, it is autotetraploid and monophyletic, and a clearly distinct species with a coherent set of morphological characters and distribution growing wild across Western Ecuador. It can therefore be confidently described as the most likely extant descendant of sweet potato's tetraploid progenitor.

In contrast, the wild tetraploid and hexaploid collections from Mexico (PI 518474, K233, DLP2923, K123) previously referred to as *I. littoralis*, *I. batatas* var. *apiculata*, "*I. trifida* 4X/6X" or "*I. batatas* 4X" (see previous chapter) are recovered with suspected hybrid *I. tabascana*, between *I. trifida* and the well-supported clade containing the Ecuadorian tetraploids and hexaploid *I. batatas* in all nuclear phylogenetic results. They are most closely related to the putatively derived sweet potato chloroplast lineage 2 in the chloroplast analyses; and genomic analysis shows that they possess two distinct sub-genomes, consistent with a history of hybridization. They also share a high degree of morphological similarity with both *I.*

batatas and *I. trifida*. Taken together, these results strongly indicate that unlike the Ecuadorian material, the Mexican tetraploids are most likely the result of modern hybridization between hexaploid *I. batatas* and diploid *I. trifida* (Fig.1).

CONCLUSION:

While the newly described *I. aequatoriensis* meets the criteria outlined above for the most likely extant descendant of the tetraploid progenitor of *I. batatas*, some doubt remains as to the exact route to hexaploidy in sweet potato. In particular, the results of the phylogenetic analysis of chloroplast sequence data fail to provide conclusive evidence confirming chloroplast capture as the means by which the two *I. batatas* lineages arose. This hypothesis is potentially undermined by the position of the putative hybrid tetraploids sister to Lineage 2, rather than nested within it as would be expected if they are recent hybrids and is the case for hexaploid K123. In addition, the fact that none of these putative tetraploid hybrids possesses a chloroplast similar to diploid *I. trifida* suggests that the chloroplast is predominantly maternally inherited in crosses between *I. batatas* and *I. trifida*, again complicating the chloroplast capture hypothesis. The incongruence in the topologies of nuclear phylogenies based on different assembly methods (Figs. 7&8), and the difficulty of accurately resolving potentially reticulate phylogenetic relationships in a dichotomously branching tree based on consensus sequences (Fig.1) also negate precise statements on the exact evolutionary history of this group of taxa, which will require a population genomics approach, and ultimately, the phased assembly of the relevant genomes and subgenomes.

Nonetheless, what is clear is that the discovery of a new autotetraploid species of *Ipomoea* that is the closest wild relative of cultivated sweet potato, and most likely the direct descendant of its tetraploid progenitor represents a major

discovery. These results definitively resolve a question that has occupied researchers ever since the discovery of sweet potato's hexaploid genome in the mid-20th century (Ting and Kehr 1953; Ting et al. 1957; Nishiyama 1959, 1961, 1971; Nishiyama et al. 1961, 1975; Teramura and Nishiyama 1962; Jones 1964, 1967, 1970; Magoon et al. 1970; Martin et al. 1974; Austin 1977; Kobayashi 1984; Shiotani and Kawase 1987, 1989; Oracion et al. 1990; Orjeda et al. 1990, 1991; Jarret et al. 1992; Bohac et al. 1993; Ozias-Akins and Jarret 1994; Jarret and Austin 1994; Huang and Sun 2000; Becerra Lopez-Lavalle and Orjeda 2002; Rajapakse et al. 2004; Srisuwan et al. 2006; Roullier et al. 2013), setting the stage for an exciting new phase in its domestication.

Perhaps most importantly, once assembled, the autotetraploid genome of *I. aequatoriensis* will provide a reference for more accurately assembling the divergent subgenomes in *I. batatas*. This will facilitate further research into the process of polyploidization that led to hexaploid sweet potato and its economically important storage root. It could also allow the incorporation of adaptive traits from the wild in crop improvement breeding programmes, which is crucial to crop improvement in the face of predicted climate change and population growth (Vaughan et al. 2007; Guarino and Lobell 2011; Dempewolf et al. 2017; von Wettberg et al. 2020). The identification of the extant descendant of the tetraploid progenitor of *I. batatas* is thus an essential step towards improving and accelerating research into sweet potato's origin and contemporary diversity.

REFERENCES:

Abbott R., Albach D., Ansell S., Arntzen J.W., Baird S.J.E., Bierne N., Boughman J., Brelsford A., Buerkle C.A., Buggs R., Butlin R.K., Dieckmann U., Eroukhmanoff F., Grill A., Cahan S.H., Hermansen J.S., Hewitt G., Hudson A.G., Jiggins C., Jones J., Keller B., Marczewski T., Mallet J., Martinez-Rodriguez P., Möst M., Mullen S., Nichols R., Nolte A.W., Parisod C., Pfennig K., Rice A.M., Ritchie M.G., Seifert B., Smadja C.M., Stelkens R., Szymura J.M., Väinölä R., Wolf J.B.W., Zinner D. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Andermann T., Torres Jiménez M.F., Matos-Maraví P., Batista R., Blanco-Pastor J.L., Gustafsson A.L.S., Kistler L., Liberal I.M., Oxelman B., Bacon C.D., Antonelli A. 2020. A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project. Front. Genet. 10:1–20.

Austin D.F. 1977. Hybrid polyploids in *Ipomoea* section Batatas. J. Hered. 62:259–260.

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

Austin D.F. 1982. Convolulaceae. In: Harling G., Sparre B., editors. Flora of Ecuador. Stockholm: Swedish Research Councils. p. 30–69.

Austin D.F., Jarret R.L., Tapia C., de la Puente F. 1993. Collecting tetraploid *I. batatas* (L.) Lam. in Ecuador. FAO/IBPGR Plant Genet. Resour. Newsl. 91/92:33–35.

Austin D.F., de la Puente F., Contreras J. 1991. *Ipomoea tabascana*, an Endangered Tropical Species. Econ. Bot. 45:435.

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., Pyshkin A. V, Sirotkin A. V, Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology. p. 455–77.

Barnard-Kubow K.B., McCoy M.A., Galloway L.F. 2017. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. New Phytol. 213:1466–1476.

Becerra Lopez-Lavalle L.A., Orjeda G. 2002. Occurrence and Cytological Mechanism of 2n Pollen Formation in a Tetraploid Accession of *Ipomoea batatas* (Sweet Potato). J. Hered. 93:185–192.

Björklund M. 2019. Be careful with your principal components. Evolution (N. Y). 73:2151–2158.

Blischak P.D., Chifman J., Wolfe A.D., Kubatko L.S. 2018. HyDe: A python package for genome-scale hybridization detection. Syst. Biol. 67:821–829.

Bohac J.R., Austin D.F., Jones A. 1993. Discovery of Wild Tetraploid Sweetpotatoes. Econ. Bot. 47:193–201.

Borcard D., Gillet F., Legendre P. 2011. Numerical Ecology with R. New York: Springer.

Borowiec M.L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016.

Cao Q., Tang J., Li A., Gruneberg W., Huamani K., Ma D. 2014. Ploidy Level and Molecular Phylogenic Relationship among Novel *Ipomoea* Interspecific Hybrids. p. 32–38.

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.

Chen S., Zhou Y., Chen Y., Gu J. 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34:i884–i890.

Contreras J., de la Puente F., Austin D.F., Diaz J. 1994. Biodiversity of sweetpotato (*Ipomoea batatas*, Convolvulaceae) in Southern Mexico. Econ. Bot. 49:286–296.

Czech L., Huerta-Cepas J., Stamatakis A. 2017. A critical review on the use of support values in tree viewers and bioinformatics toolkits. Mol. Biol. Evol. 34:1535–1542.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158.

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M. 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10:1–4.

Dempewolf H., Baute G., Anderson J., Kilian B., Smith C., Guarino L. 2017. Past and future use of wild relatives in crop breeding. Crop Sci. 57:1070–1082.

Dodson C.H., Gentry A.H. 1978. FLORA OF THE RIO PALENQUE SCIENCE CENTER: Los Rios Province ECUADOR. Selbyana. 4:301–302.

Dodson C.H., Gentry A.H., Valverde F.M. 1985. La Flora de Jauneche, Los Rios, Ecuador. Quito: Banco Central del Ecuador.

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

Gao M., Ashu G.M., Stewart L., Akwe W.A., Njiti V., Barnes S. 2011. Wx intron variations support an allohexaploid origin of the sweetpotato [*Ipomoea batatas* (L.) Lam]. Euphytica. 177:111–133.

Gordon S.P., Levy J.J., Vogel J.P. 2019. PolyCRACKER, a robust method for the unsupervised partitioning of polyploid subgenomes by signatures of repetitive DNA evolution. BMC Genomics. 20:1–14.

Guarino L., Lobell D.B. 2011. A walk on the wild side. Nat. Clim. Chang. 1:374–375.

Hillis D.M. 1998. Taxonomic Sampling, Phylogenetic Accuracy, and Investigator Bias. Syst. Biol. 47:3–8.

Hoshino A., Jayakumar V., Nitasaka E., Toyoda A., Noguchi H., Itoh T., Shin T., Minakuchi Y., Koda Y., Nagano A.J., Yasugi M., Honjo M.N., Kudoh H., Seki M., Kamiya A., Shiraki T., Carninci P., Asamizu E., Nishide H., Tanaka S., Park K. Il, Morita Y., Yokoyama K., Uchiyama I., Tanaka Y., Tabata S., Shinozaki K., Hayashizaki Y., Kohara Y., Suzuki Y., Sugano S., Fujiyama A., Iida S., Sakakibara Y. 2016. Genome sequence and analysis of the Japanese morning glory *Ipomoea nil*. Nat. Commun. 7:1–10.

Huaman Z. 1991. Descriptors for sweet potato. CIP. Lima, Peru

Huang J.C., 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.

Isobe S., Shirasawa K., Hirakawa H. 2019. Current status in whole genome sequencing and analysis of Ipomoea. Plant Cell Rep.

Jarret R.L., Austin D.F. 1994. Genetic diversity and systematic relationships in sweetpotato (*Ipomoea batatas* (L.) Lam.) and related species as revealed by RAPD analysis. Genet. Resour. Crop Evol. 41:165–173.

Jarret R.L., Gawel N., Whittemore A.T. 1992. Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. J. Am. Soc. Hortic. Sci. 117:633–637.

Jones A. 1964. Chromosome numbers in the genus *Ipomoea*. J. Hered.55(5):216-219

Jones A. 1965. Cytological observations and fertility measurements of sweet potato (*Ipomoea batatas* (L.) Lam.). Proc. Am. Soc. Hortic. Sci. 86:527–537.

Jones A. 1967. Should Nishiyama's K123 (*Ipomoea trifida*) Be Designated *I. batatas*? Economic Botany. p. 163–166.

Jones A. 1970. Asynapsis in *Ipomoea gracilis*. J. Hered. 61:151–152.

Junier T., Zdobnov E.M. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics. 26:1669–1670.

Kajitani R., Toshimoto K., Noguchi H., Toyoda A., Ogura Y., Okuno M., Yabana M., Harada M., Nagayasu E., Maruyama H., Kohara Y., Fujiyama A., Hayashi T., Itoh T. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 24:1384–1395.

Kajitani R., Yoshimura D., Okuno M., Minakuchi Y., Kagoshima H., Fujiyama A., Kubokawa K., Kohara Y., Toyoda A., Itoh T. 2019. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. Nat. Commun. 10:1–15.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., Von Haeseler A., Jermiin L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software

version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–80.

Kobayashi M. 1984. The *Ipomoea trifida* Complex Closely Related to Sweet Potato. Proceedings, Sixth Symposium of the International Society for Tropical Root Crops, hosted by CIP in Lima, Peru. p. 561–568.

Le S., Josse J., Husson F. 2008. FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software. 25:1–18.

Lebot V. 2008. Sweet Potato. In: Atherton J., Rees A., editors. Tropical root and tuber crops: cassava, sweet potato, yams and aroids. Wallingford, UK: CABI Publishing Group. p. 99–182.

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

Li M., Yang S., Xu W., Pu Z., Feng J., Wang Z., Zhang C., Peng M., Du C., Lin F., Wei C., Qiao S., Zou H., Zhang L., Li Y., Yang H., Liao A., Song W., Zhang Z., Li J., Wang K., Zhang Y., Lin H., Zhang J., Tan W. 2019. The wild sweetpotato (*Ipomoea trifida*) genome provides insights into storage root development. BMC Plant Biol. 19:1–17.

Lischer H.E.L., Excoffier L., Heckel G. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: Implications for the evolutionary history of *Microtus* voles. Mol. Biol. Evol. 31:817–831.

Loebenstein G., Thottappilly G. 2009. The Sweetpotato. Springer Netherlands.

Magoon M.L., Krishnan R., Vijaya Bai K. 1970. Cytological evidence on the origin of sweet potato. Theor. Appl. Genet. 40:360–366.

Marçais G., Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27:764–770.

Martin F.W., Jones A., Ruberte R.M. 1974. A Wild *Ipomoea* Species Closely Related to the Sweet Potato. Econ. Bot. 28:287–292.

McDonald J.A., Austin D.F. 1990. Changes and Additions in *Ipomoea* Section Batatas (Convolvulaceae). Brittonia. p. 116–120.

Md V., Misra S., Li H., Aluru S. 2019. Efficient architecture-aware acceleration of

BWA-MEM for multicore systems. Proc. - 2019 IEEE 33rd Int. Parallel Distrib. Process. Symp. IPDPS 2019.:314–324.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37:1530–1534.

Mölder F., Jablonski K.P., Letcher B., Hall M.B., Tomkins-tinch C.H., Sochat V., Forster J., Lee S., Twardziok S.O., Kanitz A., Wilm A., Holtgrewe M., Rahmann S., Nahnsen S., Köster J. 2021. Sustainable data analysis with Snakemake [version 1 ; peer review : 1 approved , 1 approved with reservations]. F1000Research.:1–25.

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., Rausher M.D., Kelly S., Liston A., Scotland R.W., Williams B.R.M., Kronmiller B., Harris S.A., Anglin N.L., Ellis D., Rausher M.D., Kelly S., Liston A., Scotland R.W. 2018. Reconciling Conflicting Phylogenies in the Origin of Sweet Potato and Dispersal to Polynesia. Curr. Biol. 28:1246-1256.e12.

Muñoz-Rodríguez P., Carruthers T., Wood J.R.I., Williams B.R.M., Weitemier K., Kronmiller B., Goodwin Z., Sumadijaya A., Anglin N.L., Filer D., Harris D., Rausher M.D., Kelly S., Liston A., Scotland R.W. 2019. A taxonomic monograph of *Ipomoea* integrated across phylogenetic scales. Nat. Plants. 5:1136–1144.

Muñoz-Rodríguez P., Wells T., Wood J.R.I., Carruthers T., Anglin N.L., Jarret R.L., Scotland R.W. 2022. Discovery and characterization of sweetpotato’s closest tetraploid relative. New Phytol. 234:4.

Nishiyama I. 1959. Collecting the sweet potato and its allied species in USA and Mexico. Japanese Journal of Breeding. 9:73–78.

Nishiyama I. 1961. The origin of the sweet potato. Tenth Pacific Sci. Congr., Unviv. Hawaii.:119–128.

Nishiyama I. 1971. Evolution and Domestication of the Sweet Potato. Bot. Maj. Tokyo. 84:377–387.

Nishiyama I., Fujise K., Teramura T., 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. Japanese J. Breed. 11:37–43.

Nishiyama I., Miyazaki T., Sakamoto S. 1975. Evolutionary autoploidy in the sweet potato (*Ipomoea Batatas* (L.) Lam.) and its progenitors. Euphytica. 24:197–208.

Oracion M.Z., Niwa K., Shiotani I. 1990. Cytological analysis of tetraploid hybrids between sweet potato and diploid *Ipomoea trifida* (H. B. K.) Don. Theoretical and Applied Genetics. p. 617–624.

Orjeda G., Freyre R., Iwanaga M. 1990. Production of 2n pollen in diploid *Ipomoea trifida*, a putative wild ancestor of sweet potato. J. Hered. 81:462–467.

Orjeda G., Freyre R., 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., Jarret R.L. 1994. Nuclear DNA content and ploidy levels in genus *Ipomoea.* J. Am. Soc. Hortic. Sci. 119:110–115.

Plazzi F., Ferrucci R.R., Passamonti M. 2010. Phylogenetic representativeness: A new method for evaluating taxon sampling in evolutionary studies. BMC Bioinformatics. 11.

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

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., De Bakker P.I.W., Daly M.J., Sham P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

R Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rajapakse S., Nilmalgoda S.D., Molnar M., Ballard R.E., Austin D.F., Bohac J.R. 2004. Phylogenetic relationships of the sweetpotato in *Ipomoea* series Batatas (Convolvulaceae) based on nuclear β-amylase gene sequences. Mol. Phylogenet. Evol. 30:623–632.

Ranallo-Benavidez T.R., Jaron K.S., Schatz M.C. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. 11.

Razifard H., Ramos A., Della Valle A.L., Bodary C., Goetz E., Manser E.J., Li X., Zhang L., Visa S., Tieman D., Van Der Knaap E., Caicedo A.L. 2020. Genomic evidence for complex domestication history of the cultivated tomato in Latin America. Mol. Biol. Evol. 37:1118–1132.

Roullier C., Duputié A., Wennekes P., Benoit L., Fernández Bringas V.M., Rossel G., Tay D., McKey D., Lebot V. 2013. Disentangling the Origins of Cultivated Sweet Potato (*Ipomoea batatas* (L.) Lam.). PLoS One. 8.

Roullier C., Rossel G., Tay D., McKey D., Lebot V. 2011. Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. Mol. Ecol. 20:3963–3977.

Sayyari E., Mirarab S. 2016. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. Mol. Biol. Evol. 33:1654–1668.

Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., Preibisch S., Rueden C., Saalfeld S., Schmid B., Tinevez J.Y., White D.J., Hartenstein V., Eliceiri K., Tomancak P., Cardona A. 2012. Fiji: An open-source platform for biological-image analysis. Nat. Methods. 9:676–682.

Session A.M., Rokhsar D.S. 2023. Transposon signatures of allopolyploid genome evolution. Nat. Commun. 14:1–14.

Shiotani I., Kawase T. 1987. Synthetic Hexaploids Derived from Wild Species Related to Sweet Potato. Japanese J. Breed. 37:367–376.

Shiotani I., Kawase T. 1989. Genomic structure of sweet potato and hexaploids in Ipomoea trifida (H. B. K.) Don. Japanese J. Breed. 39:57–66.

Soltis D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants. 5:65–84.

Srisuwan S., Sihachakr D., Siljak-Yakovlev S. 2006. The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. Plant Sci. 171:424–433.

Stift M., Kolář F., Meirmans P.G. 2019. Structure is more robust than other

clustering methods in simulated mixed-ploidy populations. Heredity (Edinb). 123:429–441.

Teramura T., Nishiyama I. 1962. Mexican wild forms of sweet potato. Econ. Bot. 16:305–314.

Ting Y.C., Kehr A.E. 1953. Meiotic studies in the sweet potato: (*Ipomoea batatas* Lam.). J. Hered. 44:207–211.

Ting Y.C., Kehr A.E., Miller J.C. 1957. A Cytological Study of the Sweet Potato Plant *Ipomoea Batatas* (L.) Lam. and its Related Species. Am. Nat. 91:197–203.

Vaughan D.A., Balázs E., Heslop-Harrison J.S. 2007. From crop domestication to super-domestication. Annals of Botany. 100(5):893–901.

Wagner N.D., He L., Hörandl E. 2020. Phylogenomic Relationships and Evolution of Polyploid Salix Species Revealed by RAD Sequencing Data. Front. Plant Sci. 11:1–15.

Wells T., Carruthers T., Muñoz-Rodríguez P., Sumadijaya A., Wood J.R.I., Scotland R.W. 2021. Species as a Heuristic: Reconciling Theory and Practice. Syst. Biol. 71(5):1233–1243.

Wells T., Carruthers T., Scotland R.W. 2022. Heuristics, species, and the analysis of systematic data. Trends Plant Sci. 27:524–527.

von Wettberg E., Davis T.M., Smýkal P. 2020. Editorial: Wild Plants as Source of New Crops. Front. Plant Sci. 11:591554.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer Verlag.

Wood J.R.I., Muñoz-Rodríguez P., Williams B.R.M., Scotland R.W. 2020. A foundation monograph of *Ipomoea* (Convolvulaceae) in the New World. PhytoKeys. 143:1–823.

Wu S., Lau K.H., Cao Q., Hamilton J.P., Sun H., Zhou C., Eserman L., Gemenet D.C., Olukolu B.A., Wang H., Crisovan E., Godden G.T., Jiao C., Wang X., Kitavi M., Manrique-carpintero N., Vaillancourt B., Wiegert-rininger K., Yang X., Bao K., Schaff J., Kreuze J., Gruneberg W., Khan A., Ghislain M., Ma D., Jiang J., Mwanga R.O.M. 2018. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. Nature

Communications. p. 1–12.

Yang J., Moeinzadeh M.-H.H., Kuhl H., Helmuth J., Xiao P., Haas S., Liu G., Zheng J., Sun Z., Fan W., Deng G., Wang H., Hu F., Zhao S., Fernie A.R., Boerno S., Timmermann B., Zhang P., Vingron M. 2017. Haplotype-resolved sweet potato genome traces back its hexaploidization history. Nat. Plants. 3:696–703.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. BMC Bioinformatics. p. 15–30.

Zhou W., Soghigian J., Xiang Q.-Y. (Jenny). 2021. A New Pipeline for Removing Paralogs in Target Enrichment Data. Syst. Biol. 0:1–16.

CHAPTER 4: *Ipomoea aequatoriensis* T.Wells & P.Muñoz: a new autotetraploid species of *Ipomoea* from Ecuador & Colombia

CHAPTER ABSTRACT

The discovery of wild, autotetraploid plants from Ecuador that are the closest crop wild relative (CWR) of cultivated hexaploid sweet potato (*Ipomoea batatas* (L.) Lam.) and the probable descendants of its tetraploid progenitor, led to their description as a new species, *I. aequatoriensis* T.Wells & P.Munoz. In this chapter, I give a detailed review of the decision to recognise *I. aequatoriensis* as a separate species, in the context of the newly developed *heuristic species concept*. This method of species delimitation arose from the Foundation Monograph approach to overhauling the taxonomy of large, complex groups of tropical plants. It requires combined evidence of genotypic, phenotypic, and ecotypic cohesion for the recognition of a separate species. The chapter includes background to the heuristic species concept, a summary of the taxonomic history of *I. aequatoriensis*, citations of existing and newly collected specimens, an updated description of its key morphological characteristics and diversity, a review of its phylogenetic position within the Batatas clade of sweet potato's closest CWRs, and detailed notes on its present distribution and ecology. Integrative analysis of this combined dataset supports the description of *I. aequatoriensis* as a separate species, and the inclusion of Colombian specimens within that delimitation. This illustrates how explicit recognition of the inherently heuristic nature of species delimitation and their use in biological research helps alleviate the theoretical and practical difficulties presented by the "species problem" – one of biology's most intractable challenges. It is also hoped that the description of this new species and its distinction from other tetraploid material collected growing wild in Latin America will bring much needed clarity to sweet potato research and crop improvement breeding programmes.

INTRODUCTION:

The "Species Problem"

The notion of distinct species that can be counted, measured and compared is fundamental to most branches of biological research. This is because species form a basic unit for researchers working in systematics, ecology, conservation, and evolutionary biology, where they are central to important scientific insights such as the spatiotemporal distribution (von Humboldt 1805; de Candolle 1855; Wallace 1876; Brundin 1972; Myers et al. 2000; Hillebrand 2003), evolutionary drivers (Darwin 1859; Hughes et al. 2013; Gao and Rieseberg 2020), and future resilience of biodiversity (Maclaurin and Sterelny 2008; Zografou et al. 2021; Bauman et al. 2022). Identifying and describing species is thus a key step in understanding the patterns and processes that underpin Earth's ecosystems.

Unfortunately, there continues to be widespread disagreement among biologists and philosophers of biology about what species are and how they should be delimited from one another (e.g. Mallet 2020; Wang et al. 2020). What is more, advances in our understanding of evolution and genetics have led some to question whether species represent a real unit of biology at all (e.g. Gilmour 1958; Vrana and Wheeler 1992; Bachmann 1998; Ereshefsky 1998; Pleijel 1999; Hendry et al. 2000; Pleijel and Rouse 2000).

The major issue at the heart of this "species problem", is the necessity of reconciling the variability inherent to evolving organisms with the need for a more or less stable and comparable unitary category (Hull 1997). This is essential if species are to be identified and delimited accurately and consistently, which is undermined by the over 30 different definitions – or "species concepts" – currently in existence (Zachos 2016).

Species as Historical Lineages

Attempts to create a single, unified species definition have tended to focus on the notion of species as *historical individuals* or lineages (e.g. Hull 1976, 1978; Ghiselin 1987; de Queiroz 1998, 1999, 2007; Padial and De la Riva 2020). Under this conception, species can be identified and delimited based on "contingent properties", or criteria indicative of their status as independent lineages with a unique evolutionary history and trajectory (de Queiroz 2007). Part of the popularity of this approach is arguably derived from the rise of molecular phylogenetics and contemporary "phylogenetic thinking", which gives researchers a rapid and empirical way of testing lineage divergence (Freudenstein et al. 2017; Padial and De la Riva 2020).

This has led some to claim that an increase in the power, sampling and accuracy of these techniques should lead to more accurate and consistent species recognition (Padial and De la Riva 2020). But others have pointed out that deciding exactly how much lineage divergence is necessary and sufficient for species delimitation remains an inherently subjective issue (Freudenstein et al. 2017). And that an over reliance on the increasing power of molecular phylogenetics to resolve genetic structure risks recognizing temporarily isolated populations – or "evolutionary ephemera" – as species (Freudenstein et al. 2017; Sukumaran and Knowles 2017; Hillis 2020). Thus, in turn potentially reducing consistency and comparability among species.

Integrative Taxonomy

A lineage-based species concept is not inherently tied to molecular phylogenetics however, and de Queiroz (2007) explicitly called for the combined analysis of different contingent properties in species delimitation. This combination of different types of systematic data is often referred to as "integrative taxonomy"

(e.g. Dayrat 2005; Will et al. 2005), despite the lack of a formal definition for that term (Padial and De La Riva 2010; Yeates et al. 2011). Nonetheless, the assumption that simply integrating more types of data will lead to better species delimitation faces challenges on two fronts.

The first is the lack of an agreed mechanism for dealing with incongruence between the evidence from different data sources for and against species delimitation (Padial and De La Riva 2010). Incongruence is an inevitable consequence of the fact that symptoms of lineage divergence do not occur simultaneously or in a particular order (de Queiroz 1998, 2007). As such, under a lineage-based species concept, incongruence between different data sources should be expected and should not automatically rule out species status (Padial and De La Riva 2010). But deciding which characters should take precedence in these instances risks resurrecting the inconsistency and sense of arbitrariness that lineage-based approaches sought to escape (Freudenstein et al. 2017).

Secondly, and perhaps more importantly, decisions about whether a given group of organisms represents a distinct species are invariably made from limited or incomplete data. In this sense, taxonomists are constrained by both the partiality of their data and their own limits as biological organisms bounded by space and time. Even for well-sampled and studied taxa, our ability to perceive both the evolutionary history and future trajectory of a group of organisms, and the complexity of their contemporary diversity and relationships present fundamental limits that inherently impact decision-making.

Species as a Heuristic

In light of these fundamental restrictions on our ability to recognise and delimit species, species delimitation is perhaps therefore best recognised as a *heuristic* endeavour: an attempt to resolve complex decision problems in the face of

insurmountable uncertainty. Heuristic methods are those that rely on the strategic analysis of a subset of all the potentially available data to make decisions that would otherwise be irresolvable (Kulkarni and Simon 1988; Gigerenzer and Brighton 2009; Marewski and Gigerenzer 2012). To do this they often involve repeated trial and error, providing a provisional answer that is as accurate as possible at the time, but open to revision should new evidence come to light.

Heuristics describe well the process of taxonomy and species delimitation, where species hypotheses are generated and tested based on strategic analysis of a set of selected characters for a sampling of representative specimens or individuals. New samples, data or methods may lead to a re-appraisal of that initial decision, but in the interim a heuristic approach to species provides “a practical method not guaranteed to be optimal or perfect, but sufficient for the immediate goals” (Wood et al. 2020).

The use of species themselves in wider biological research can also be described as heuristic, since they allow us to investigate and understand complex phenomena, such as evolution, or the processes underlying the distribution of biodiversity. What is more, approaching species in a heuristic manner aligns with our theoretical understanding of species as variable and evolving groups of organisms. Subject to continuous change over time, species are in a resultant state of flux or “homeostasis” depending on the interaction of different ecological and evolutionary forces. This means that in recognising a species, we must also find a way to recognise and accommodate evidence indicative of these competing forces.

Species Cohesion

In this sense, the notion of *species cohesion* is central to the heuristic approach to species delimitation (Wells et al. 2021, 2022). This term was coined by Templeton (1989) in his development of the *Cohesion Species Concept*. In it, he

identified “cohesion mechanisms” as the forces responsible for generating and maintaining independent species. He therefore described the basic task in species delimitation as one of identifying these mechanisms, which help to maintain an evolutionary lineage in a state of extended equilibrium, or “homeostasis”. In so doing, he combined elements of Mayr's (1963) *Biological Species Concept* (which he followed Paterson (1985) in terming the Isolation Species Concept) with those from the *Recognition* (Paterson 1985) and *Evolutionary* (Simpson 1951) *Species Concepts*, thereby advocating an early form of integrative taxonomy. The two key terms he used for this approach were *genetic* and *demographic exchangeability*, the relative importance of which will vary based on the specific biology of the organisms in question.

Genetic exchangeability expresses the possibility of geneflow (predominantly by sexual reproduction) between organisms. This may incorporate mate recognition and other pre- and post-zygotic isolation barriers. Genetically exchangeable individuals share a common evolutionary history and thus genetic similarity resulting from common descent. Genetic exchangeability alone is not necessary or sufficient to define a species, however. Instead it has to be considered in conjunction with demographic exchangeability. This describes individuals that share a fundamental ecological niche, facilitated by a shared phenotype, which is itself derived from the similar genotypes of individuals. These individuals share similar phenotypic characters and ecological tolerances derived from a similar genetic make-up. As ecological tolerances differ within a group because of differences in that genetic make-up, and differing selective pressures are able to act on these differences, so demographic exchangeability falls. When this happens, not all individuals are able to respond in the same way to the same stimuli – they are no longer demographically exchangeable – and cohesion begins to break down. It is this interplay between

genetic and demographic exchangeability that underpins the notion of species cohesion (Templeton 1989).

In a similar vein, Maclaurin and Sterelny (2008, p.38) describe a counterbalance between “ecological mosaics” of individuals and the “metapopulation dynamics” of gene flow between those individuals. Together these act as balancing forces that hold species in a form of homeostasis or “punctuated equilibrium”. Only when the forces of selection acting on the ecological mosaic of different individuals are strong enough to exceed balancing metapopulation dynamics will that equilibrium (cohesion) breakdown, and speciation occur. In species delimitation therefore, the task of a taxonomist is to search for clusters of properties that indicate genetic and demographic exchangeability, or counterbalancing ecological and evolutionary forces. These cohesion mechanisms and the properties indicative of them can then be traced to identify the “most inclusive group with the potential for cohesion” – a species (Templeton 1989).

Pigliucci (2003) described the Cohesion Species Concept as a close match for the idea of species as cluster concepts, citing Wittgenstein's (1953) notion of family resemblance. A separate attempt to apply cluster concepts to the species problem can be found in the notion of species as *homeostatically sustained clusters of properties and relations* (Boyd 1991, 1999; Wilson et al. 2007).

Homeostatic Property Clusters

This attempted reformulation of the species category arose in opposition to the idea that species should be thought of as historical individuals (or lineages) rather than biological kinds. Boyd (1999) and later Wilson et al. (2007) asserted that closer analysis of the species problem revealed instead the requirement for a new definition of biological kinds like species. One which recognised them as neither traditionally essentialist, nor as definite individuals, but instead as what they termed

Homeostatic Property Clusters (HPC). This HPC definition of biological kinds such as species advocated the necessity of using flexibly delimited clusters of properties to identify them. No single property – nor specific set of properties – is strictly necessary or sufficient for membership of a species, but rather overlapping subsets of properties cohesively unite them. To this end, the HPC method sets out a specific means of assessing cohesion: one based on the identification of *causally basic properties*, and the diagnostic properties indicative of them. Properties are considered causally basic when they cannot themselves be easily explained by reference to other properties. Examples of this include genealogical and reproductive relationships, which are responsible for the possession of similar genotypes and sets of traits (Wilson et al. 2007).

The beak morphology of finches for example, is a result of geneflow distributing the genotype underpinning that phenotypic trait between their ancestors. This geneflow in turn occurs under the influence of the selective pressures that made that particular beak shape adaptive. Since the underlying mechanisms are causally dependent (geneflow is dependent on compatible genotype, and compatible genotypes are maintained by geneflow) they will tend to form clusters, not unrelated sets of properties. The shared phenotypic trait (the beak) results from genes inherited as a result of shared genealogical relationships, making it a potentially diagnostic property, though not a causally basic one. Other phenotypic traits and ecological tolerances that result from these genealogical relationships will also be potentially diagnostic, though not all of them will necessarily be manifested in all individuals, or have obviously adaptive functions.

Different diagnostic properties will therefore also tend to cluster together because they share causally basic roots. And although the clustering of these diagnostic properties may not be perfect, it will display a homeostatic integrity drawn from the shared underlying mechanisms. This in turn means that clusters of

diagnostic properties are indicative of species cohesion because they provide evidence of shared, causally basic properties.

Heuristic Species Delimitation

By approaching speciation as an evolutionary genetic process, Templeton (1989) believed he had achieved a definition of species "… that illuminates, rather than obscures or misleads, the mechanisms of speciation and their genetic consequences". Like de Queiroz (2007), Templeton (1989) therefore thought that a unified definition of species was possible, while still accommodating the variability inherent to the process of evolution. Rather than leaving the choice of properties indicative of lineage divergence to individual researchers however, cohesion-based approaches to the species problem provide some guidelines about what properties to use and how to integrate them.

A heuristic approach reinforces these guidelines and places them at the centre of the species delimitation process. It does this by showing that the same forces leading to species cohesion also give rise to the heuristic utility of species in biological research. The reason species have played such a central role in understanding Earth's biodiversity is their location at the intersection of the evolutionary and ecological forces holding them in a state of cohesion. Describing, delimiting and comparing species thus allows us to use them as a heuristic tool for understanding those forces that govern the generation and maintenance of Earth's biodiversity. In this sense, a finch's beak can illustrate evolution by natural selection (Darwin 1859), and the distributions of particular plant and animal species can tell us about geological history (Wallace 1876). If species are clusters of closely related individuals responding in a similar manner to comparable sets of evolutionary and ecological forces (Wells et al. 2021), then not only can they be identified by

contingent properties indicative of those forces, but doing so simultaneously provides valuable insights into the forces themselves.

As many have pointed out (e.g. Lee 2000; Landrum 2001; Wheeler 2020), this is why the most important aspect of systematics is not resolving phylogenetic relationships per se, but rather the "description and mapping of the world biota" (Wilson 2000). The collective metadata attached to species and compiled largely by taxonomists are what makes species such an invaluable resource to biological research. In reviewing the delimitation and description of *I. aequatoriensis* T.Wells & P.Muñoz below, I hope to provide just such a resource for the closest wild relative of a major global crop.

Taxonomic History of I. aequatoriensis

Distinctive, wild-growing plants from western Ecuador resembling wild sweet potatoes (*Ipomoea batatas* (L.) Lam.) were first highlighted by Charles Rick when he collected the germplasm specimen CH71.3 near the city of Quininde, in the northwest province of Esmeraldas (Martin et al. 1974) (see Chapter 2). Martin et al. (1974) thought that these plants and other similar ones collected by Rick in southern Colombia (CH73.1) – which were all found to be tetraploid – represented promising leads in the hunt for the wild progenitor of sweet potato. However, Austin (1977, 1978) initially dismissed them as most likely to be recent hybrids between *I. batatas* and a close diploid relative.

Dodson and Gentry (1978) listed similar material from the Ecuadorian province of Los Rios as *I. trifida* (Kunth.) G.Don, describing it as a vine, with deeply cordate leaves and cymose inflorescences of "salverform" flowers (Fig. 1). As part of a Scandinavian-led project to document the Flora of Ecuador based on extensive new collecting, Austin (1982) later pointed out that *I. trifida* was not found in Ecuador. He instead placed the Ecuadorian material described by Dodson and

Gentry (1978) in *I. batatas*, noting its short outermost sepal, with an obtuse-caudate tip and 3-5 raised veins on the back, which he gave as a defining character for the whole of *I. batatas*.

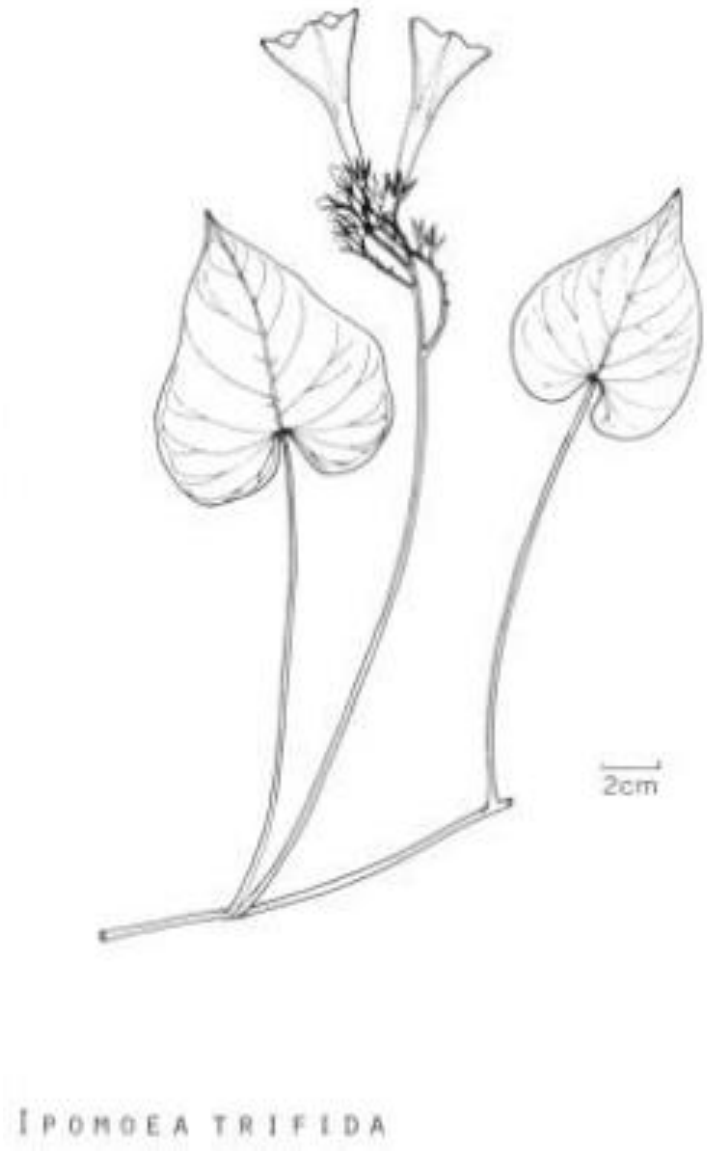


Figure 1: Reproduced from Dodson and Gentry (1978). Illustration of plants collected in Rio Palenque reserve, Los Rios province, Ecuador, and originally identified as *I. trifida*. Showing their cordate leaves, twining stems and dense cymose inflorescences of "salverform" corollas, with a slender tube and widening limb.

Dodson et al. (1985) highlighted the shorter sepals in the Ecuadorian samples compared to those seen in cultivated sweet potatoes, and Austin et al. (1993) also mentioned this character (Fig. 2D) – along with a "weedy" habit, and a tendency to twine – when they collected new germplasm and herbarium material

from across western Ecuador. Bohac et al. (1993) agreed, and added a constricted corolla tube just above the sepals as an additional distinction from cultivated sweet potatoes (Fig. 2B), though they also included some now lost non-Ecuadorian tetraploid samples in their delimitation (see Chapter 2).



Figure 2: Comparison of flowers in *I. trifida*, *I. batatas* and *I. aequatoriensis*. A-C reproduced from Bohac et al. (1993), D-E from Austin et al. (1993). A: *I. trifida*; B&D: (CH80.3 & DFA 7816) *I. aequatoriensis*; C&E: *I. batatas*. Showing short, blunt sepals and constricted corolla tube in *I. aequatoriensis*. Not to scale.

Roullier et al. (2013) showed that the Ecuadorian material collected by Austin et al. (1993) shared a chloroplast haplotype with their “Southern” lineage of *I. batatas*, but they could not provide sufficient evidence to distinguish it from potential feral or hybrid escapes of cultivated sweet potato (see Chapter 2). In the Foundation Monograph of the genus *Ipomoea* in the Americas, Wood et al. (2020) described a distinctive wild-growing form of *I. batatas* from the Ecuadorian province of Esmeraldas as a sparsely pubescent, twining herb, with chartaceous, predominantly obovate-obrhomboid, unequal sepals, which have a single vein on the back and a mucronate tip (Fig. 3). They also highlighted other material from the province of Pichincha with a larger corolla (Fig. 4), and from Manabí and Guayaquil with trilobed leaves and long, acuminate sepals (Fig. 5).

Comprehensive phylogenetic, genomic and morphological analysis has since conclusively shown that the material from Ecuador is autotetraploid and distinct from *I. batatas* and *I. trifida*, as well as other tetraploid plants collected elsewhere in Latin America, which are most likely hybrids between *I. batatas* and *I. trifida* (Muñoz-Rodríguez et al. 2022) (see previous chapter). The Colombian material originally cited by Martin et al. (1974), and later added to by Kobayashi (1984), has not been included in any subsequent taxonomic studies with the exception of Bohac et al. (1993), but was shown to share a close, if somewhat unresolved relationship to Ecuadorian material in phylogenetic analysis of 388 nuclear loci (Muñoz-Rodríguez et al. 2022).

The decision whether to describe the tetraploid material from Ecuador and Colombia as a new species separate from *I. batatas* was made on the basis of these results, as well as a review of other material held in global herbaria. Subsequent new collections from across western Ecuador were made over a three-week period in 2022, and they are also included here for further discussion about the diversity, distribution and ecology of *I. aequatoriensis*.



Figure 3: Digital image of specimen collected in Esmeraldas province, Ecuador (Sparre 15308, 1967 [S]) highlighted as a distinct form of *I. batatas* by Wood et al. (2020), based on its short, blunt calyx. Also showing very subtle lobed margin with small, blunt teeth on uppermost leaves on page. Determination: *I. aequatoriensis.*



Figure 4: Digital image of specimen collected in Pichincha province, Ecuador (Sparre 14810, 1967 [S]) highlighted as a distinct form of *I. batatas* with a large corolla by Wood et al. (2020). Note also the longer, more acuminate and foliaceous sepals. Determination: *I. aequatoriensis.*



Figure 5: Digital image of specimen collected in Manabí province, Ecuador (Asplund 15966, 1955 [S]) highlighted as a distinct form of *I. batatas* with trilobed leaves and long, acuminate sepals by Wood et al. (2020). Indetermined.

MATERIALS & METHODS:

Pre-existing Collections

Alongside the systematic review of tetraploid germplasm material described in Chapter 2, a search for similar material in herbarium collections was conducted through a thorough literature review, herbarium visits, loan requests, and online access to digital images of specimens. This led to the discovery of 15 specimens from Colombia, and 62 from Ecuador resembling tetraploid germplasm collections from the same areas (Fig. 6 & Appendix 4).

The majority of these specimens were collected between 1967 and 1992, but several date back to the mid-19th Century, and the most recent was collected in 2018. In addition to *I. batatas*, these specimens had previously been determined with names including *I. trifida*, *I. tiliacea* (Willd.) Choisy, *I. fastigiata* Choisy, *I. variabilis* Choisy, *I. glaucifolia* Ruiz & Pavon (*nom. inval*., non L.), *I. philomega* (Vell.) House, *I. sp.*, *Convolvulaceae sp*., *Convolvulus sp*., *Jacquemontia sp*. Several vernacular names are also listed, such as “Batatilla”, “Camotillo” and “Camotillo blanco”, which indicate an association with sweet potato (Batata, and Camote are both names used for sweet potato in South America).

Specimens lacking geolocation data were georeferenced using the Map Point Editor tool in BRAHMS 8, which allows text descriptions of locality to be searched against Google Maps, and the most likely result manually selected based on visual observation.

New Collections (2022)

A trip to Ecuador to collect new material of *I. aequatoriensis* was initially planned for June 2020, but was postponed until May 2022 due to the SARS-CoV-2 pandemic. The route and collecting sites were chosen based on the previously

collected specimens (Fig. 6), and discussion with collaborators in Ecuador including Dr Katya Romoleroux and David Espinel-Ortiz at QCA, Carlos Cerón at QAP, and Xavier Cornejo at GUAY. In total,119 new collections were made between Quininde, Esmeraldas in the north and the city of Loja in the south (Fig. 6), with accompanying photographs and notes on morphological variation, distribution, and ecology.

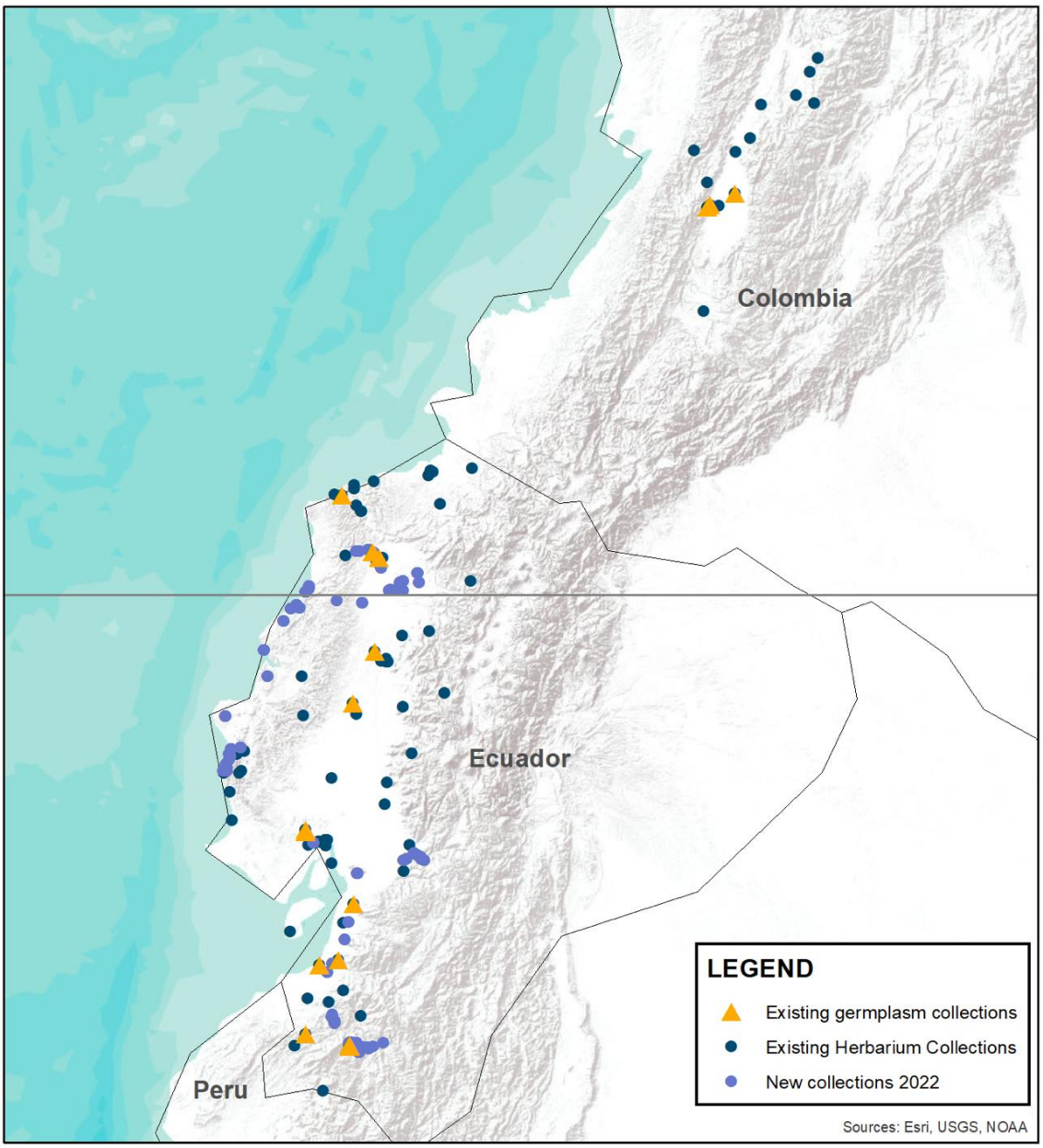


Figure 6: Map showing existing herbarium (navy dots) and germplasm collections (orange triangles) of *I. aequatoriensis* from Ecuador and Colombia, as well as new collections made in June 2022 (purple dots). Showing their distribution west of the Andes in Ecuador and in the inter-Andean Cauca Valley in Southern Colombia

Heuristics:

Since species are lineages or clusters of closely related individuals, with a shared ecological role derived from shared morphological traits (see above and Wells et al. 2021, 2022), they should ideally be delimited and identified using properties indicative of combined genotypic, phenotypic, and ecotypic cohesion (Fig. 7). While this often necessitates the integration of multiple forms of data, some properties can be simultaneously indicative of more than one type of cohesion at a time (Table 1). For example, as described above and in Wilson et al. (2007), the shared possession of a particular beak morphology by individuals of a given species of finch constitutes phenotypic cohesion. That beak is also potential evidence of the shared ancestry that led to its inheritance – genotypic cohesion. And it is also an adaptive trait, allowing the individuals in that species to fill a particular ecological niche, thereby resulting in ecotypic cohesion.

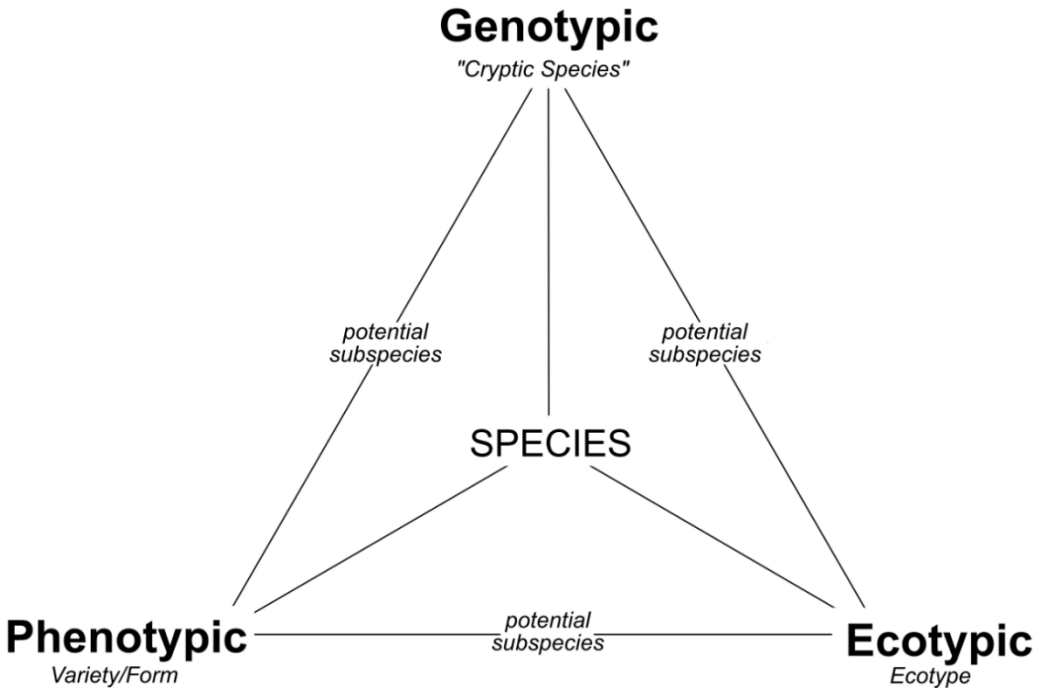


Figure 7: The properties used in species recognition should ideally provide evidence of genotypic, phenotypic, and ecotypic cohesion. Properties indicative of one form of cohesion may indicate a variety, ecotype, or 'cryptic' lineage, two together could perhaps lead to recognition as a subspecies, but all three are ideally required for a species. Some properties can be simultaneously indicative of more than one type of cohesion (see Table 1 above).

Table I. Integrating data for species recognition[a]

Contingent Property	Example	Type of cohesion		
		Genotypic	Phenotypic	Ecotypic
Monophyly	DNA barcode	•	–	–
	Fixed morphological trait	•	•	○
Population structure	Multispecies coalescent	•	–	–
Shared habitat	Distribution data	•	–	•
	Edaphic data	–	○	•
	Climatic data	–	○	•
Mutualism (e.g., pollinator)	Floral structure/reward	•	•	•
	Field observations	•	–	•
Adaptive trait	Character with clear function	•	•	•
Not apparently adaptive trait	No clear function	•	•	–

[a]Key: • indicate evidence of cohesion, ○ indicate potential evidence of cohesion, – indicate a lack of evidence.

Table 1: Examples of data that might be used to infer contingent properties indicative of species cohesion. Some individual properties can provide evidence of all three types of cohesion, since for example, shared possession of putatively adaptive morphological traits results from genotypic cohesion, and enables accommodation of the same ecological pressures.

In the case of *I. aequatoriensis*, phenotypic cohesion was assessed based on the results of the morphometric analyses in the previous chapter and comparative review of the existing herbarium, germplasm and new field collections listed here. The possession of a cohesive set of morphological characters also provided preliminary hypotheses of potential genotypic cohesion (see Table 1), confirmation of which was sought in the results of the phylogenetic analyses in the previous chapter. Evidence of ecotypic cohesion was assessed through comparison of distribution data and GIS with satellite images and regional biome maps, and ecological notes from specimen collections. The possession of thickened storage roots is thought to be an adaptation to fire and drought (Muñoz-Rodríguez et al. 2019), and on this basis it was also treated as an adaptive morphological trait potentially indicative of ecotypic cohesion (Table 1).

RESULTS & DISCUSSION:

Phenotypic Cohesion:

While wild-collected Ecuadorian plants resembling *I. batatas* have been repeatedly noted as distinctive since at least the 1970s (Martin et al. 1974), there has also been some regional variability described among them (Wood et al., 2020, see above and figs. 3-5). This variability was also noted during the review of new and existing collections carried out here.

Notes from herbarium collections and observations from germplasm and field collections consistently confirm a slender twining habit in all material analysed here. This habit is common to most of the batatas group of sweet potato CWRs (Wood et al. 2020). In the case of the material observed here, plants often reached heights of 4-5m, climbing in small trees, or on fences and exposed banks (Fig. 8D). Some specimens were also found to trail several metres along the ground (Fig. 8C). Young stems are terete, generally around 2mm in diameter and green or tinged slightly purple in colour, often pubescent to pilose towards the tip (Fig. 8C & 10A). Older stems turn grey-green and become somewhat angular with age. They are also often somewhat verrucose and woody towards the base of the plant, where they can reach diameters of 5-8mm (Figs. 9B, 10B & 11B). Trailing stems will root at the nodes where they come into contact with the ground, as is also the case in *I. batatas*.

Leaves are generally simple, entire and cordate (often deeply so) and up to 15cm long and wide. Subtly lobed leaves are not uncommon, especially on trailing stems. In general, this takes the form of small, blunt teeth (Fig. 3), but truly lobed leaves are also seen (Figs. 9B & 11C).

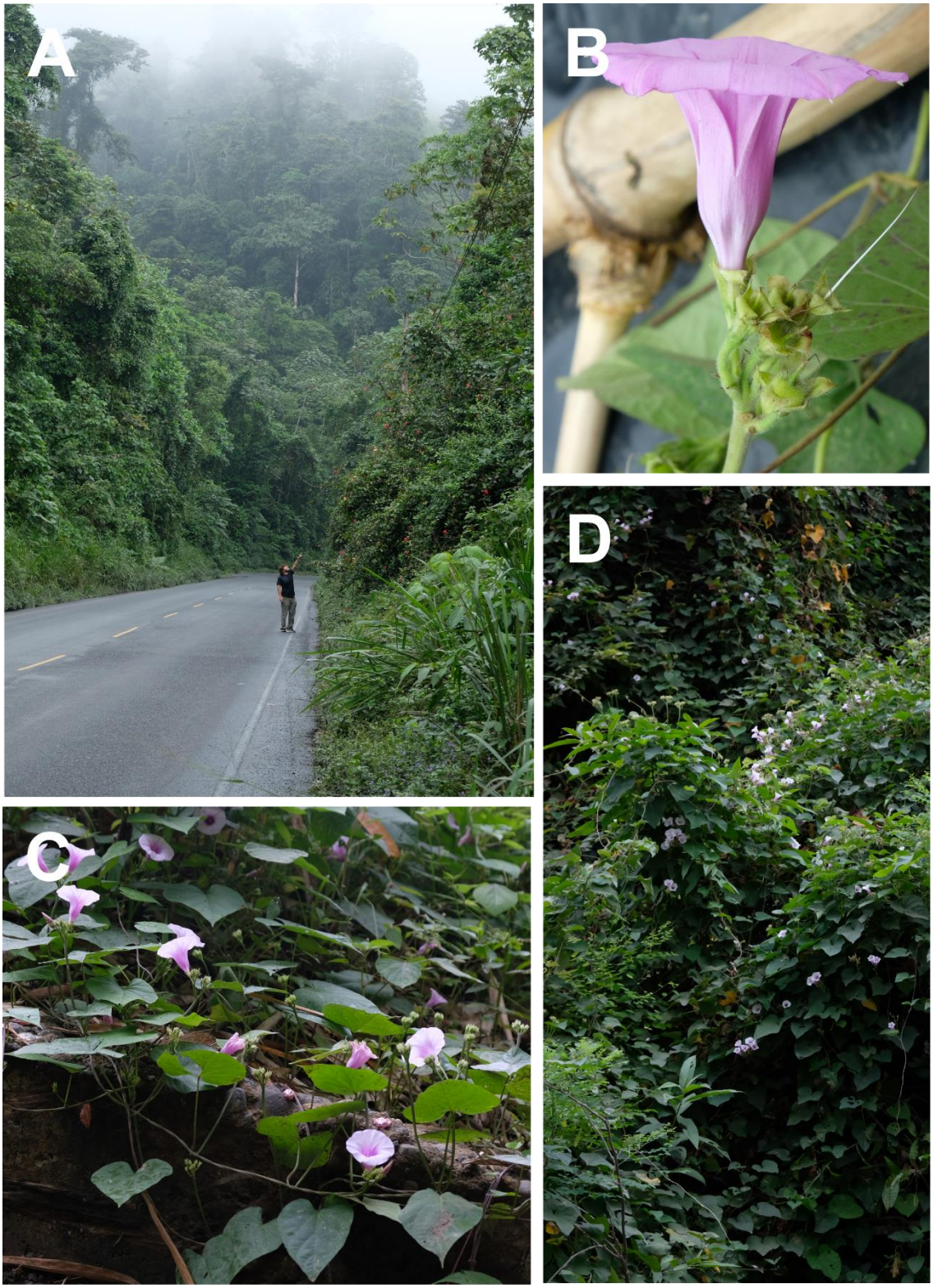


Figure 8: Phtoographs of specimens collected in June 2022 in western Ecuador. Showing **A:** Collection site of Munoz 49 (*I. aequatoriensis*), by roadside in moist evergreen forest. **B:** Inflorescence of Munoz 49, showing more foliaceous and acuminate sepals, often reflexed and drying brown towards the tips. **C:** Specimen of *I. aequatoriensis* collected by the roadside near Quininde in Esmerladas province. With slender stems, cordate leaves, and chartacesous, pale green to off-white sepals with mucronate tip. The corolla tube pale pink to white, with constriction just above calyx. **D:** Specimen of *I. aequatoriensis* climbing in low trees and shrubs on exposed slope in Cerro Blanco dry forest reserve, Guayas Province.

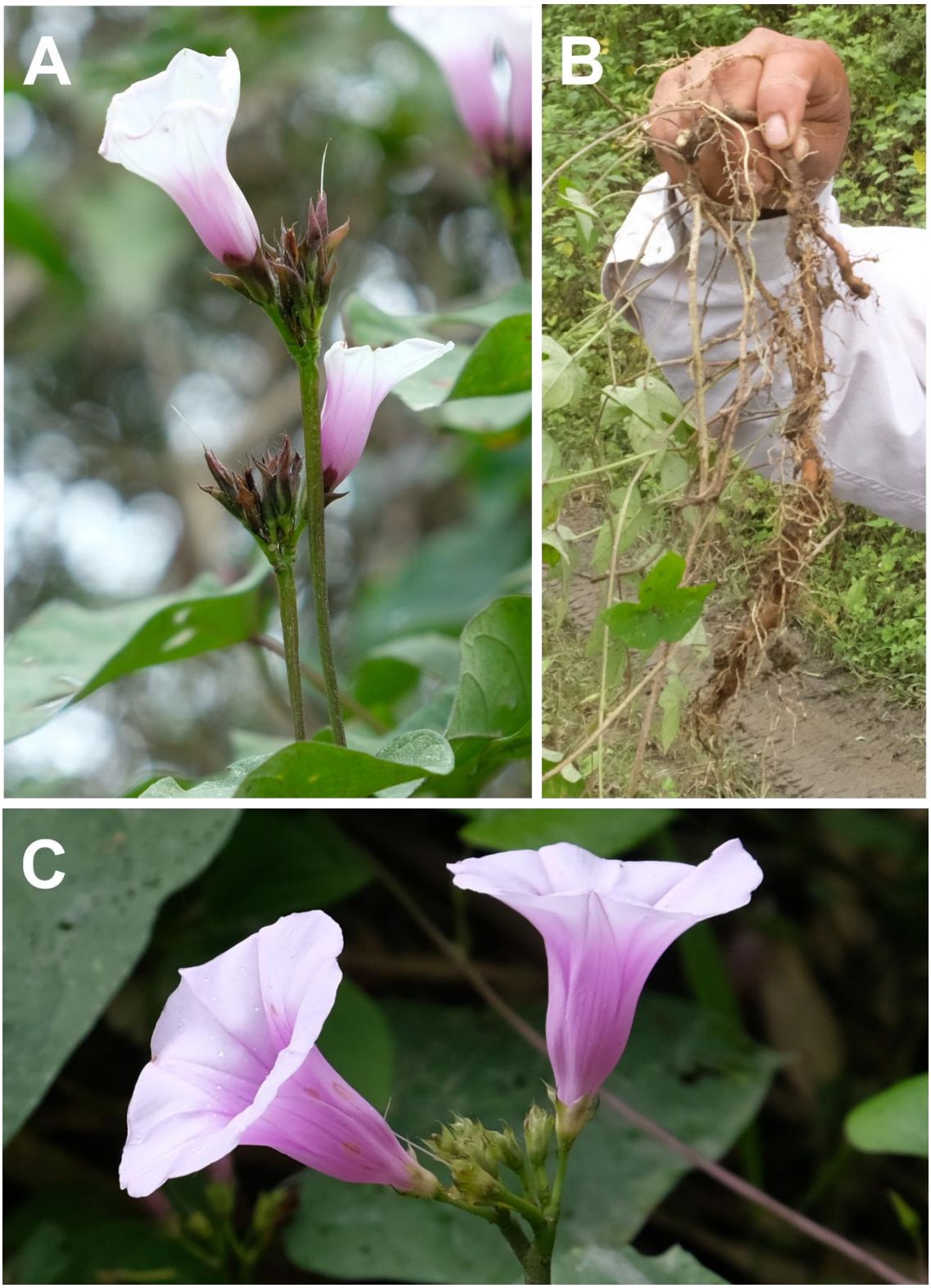


Figure 9: Phtoographs of specimens collected in June 2022 in western Ecuador. Showing **A:** A specimen of feral sweet potato (*I. batatas*) growing by the roadside in the southern province of Loja. With Dark purple, highly unequal, oblong sepals, and pale pink to off-white corolla tube. **B:** small storage roots of *I. aequatoriensis* specimen collected in Agua Blanca reserve, near Machalilla National Park, Manabi. 1-1.5cm in diameter, with white flesh and light brown skin. **C:** Inflorescence of *I. aequatoriensis* collected by the roadside near Quininde in Esmeraldas province. With chartacesous, pale green to off-white sepals, with mucronate tip. The corolla tube pale pink to white, with constriction just above calyx.

Variability in leaf form is common, even on the same plant, but leaves with more than 3 lobes/teeth are extremely rare. Many collections both in the field and greenhouse were observed to produce large cordate leaves on climbing stems, and smaller, trilobed ones on trailing ones. Similarly leaves can be dark to vivid green above and below, often varying with age. Two extrafloral nectaries are located either side of the leaf base, just above the petiole – a character shared with *I. batatas.*

Inflorescences are consistently dense umbels, with a long peduncle, branching once or twice just below the pedicels, with a total of up to 26 flowers, and rarely fewer than 8. The majority of specimens assessed tended to have short, chartaceous sepals, <6mm long, with blunt apices (>80º), often with the top third of the sepal drying brown and sometimes folded back on itself (Fig. 3). Wood et al. (2020) described this form from material collected within Esmeraldas province in the northwest of Ecuador. Germplasm collections such as *Rick CH71.3/K300* made in the same area also match this description (Fig. 10), as do multiple other herbarium and germplasm specimens collected further south in the provinces of Manabí, Guayas, El Oro and Loja, and new collections made there and in Azuay, Cañar, Pichincha and Chimborazo in 2022. Colombian specimens also have similarly short, blunt, chartaceous sepals, with a single, dark vein and mucronate tip (Fig. 11A), again matching this predominant Ecuadorian morphotype.

Plants collected in central Ecuador, especially around the province of Los Rios, were found to share a similar habit and inflorescence structure to the above material, with dense cymes that often contain upwards of 20 flowers (Figs. 4 & 12A). Both their sepals and corolla often tended to differ somewhat however. The former in being predominantly foliaceous rather than chartaceous, up to 10mm long, and with more acuminate tips (~50º). And the latter in being both longer and potentially with a wider limb (this character is often hard to assess in pressed herbarium specimens).



Figure 10: Germplasm collection *Rick CH71.3/K300* (*I. aequatoriensis*), collected near Quininde, Esmeraldas. Grown from seeds in the greenhouse at University of Oxford Department of Biology. Showing **A:** twining stems and cordate leaves with palmate venation. **B:** small storage roots, 1-1.5cm in diameter, with white flesh and brown skin. **C:** an inflorescence with short sepals, pilose to ciliate and drying brown towards the tip; the corolla pale pink to white, and narrowing towards the base, but with an only partially open limb.



Figure 11: Germplasm collection *Kobayashi K500/CH80.3* (*I. aequatoriensis*), collected near Cali, Colombia. Grown from seeds in the greenhouse at University of Oxford Department of Biology. Showing **A:** an inflorescence with short sepals, pilose to ciliate and drying brown towards the tip; the corolla pale pink to white, and narrowing towards the base, but with an only partially open limb. **B:** small storage roots, 1-1.5cm in diameter, with white flesh and brown skin. **C:** trilobed leaves with palmate venation.

This form corresponds to that described by Wood et al. (2020) as possessing a larger corolla than the material from Esmeraldas (see above & Fig. 4). Most of this area of central Ecuador was not visited during the 2022 collecting trip, but two specimens resembling the morphotype were collected from either end of its range. One on the border between Santo Domingo and Manabí (*Wells, 45*), and the other in Loja (*Munoz, 46-49*) (Fig. 8A-B). These two new collections were also observed to possess large cordate leaves, asperous to hispid all over, a character later confirmed for other herbarium material from the same area. In contrast, leaves in the more common morphotype tend to be more membranous, occasionally subtly trilobed, and more or less glabrous to sparsely pubescent. The germplasm specimen *Austin 7817* from El Oro also appears to fit this less common morphotype, with largely foliaceous sepals, and a long, narrow corolla (Fig. 10). It also has both large, cordate, hispid leaves and some that are membranous, more or less glabrous, and subtly trilobed. Despite the sepals in this less common morphotype being longer, more acuminate and foliaceous, they are also paler and approaching chartaceous towards the tips, and they are frequently reflexed and dry brown (Figs. 12A), as in the more common morphotype.

Sepal hairiness was variable across both morphotypes, with the back of the sepal densely pilose to more or less glabrous. Intermediate forms between the two morphotypes have also been observed (e.g. *Dodson 4334* [QCA]; *Harling 24997* [QCA]), and plants of both forms are recorded from the same areas of El Oro and Loja.

Wood et al (2020) also described a third, more distinct morphotype based on a specimen from Manabí (*Asplund 15966* [ST], Fig. 5) with few-flowered inflorescences, long acuminate sepals, long, straight corolla tubes, and deeply trilobed leaves. They included with this another specimen from Guayaquil (*Ruiz and Pavon s.n.* [FHO]), but closer analysis suggests the two differ.

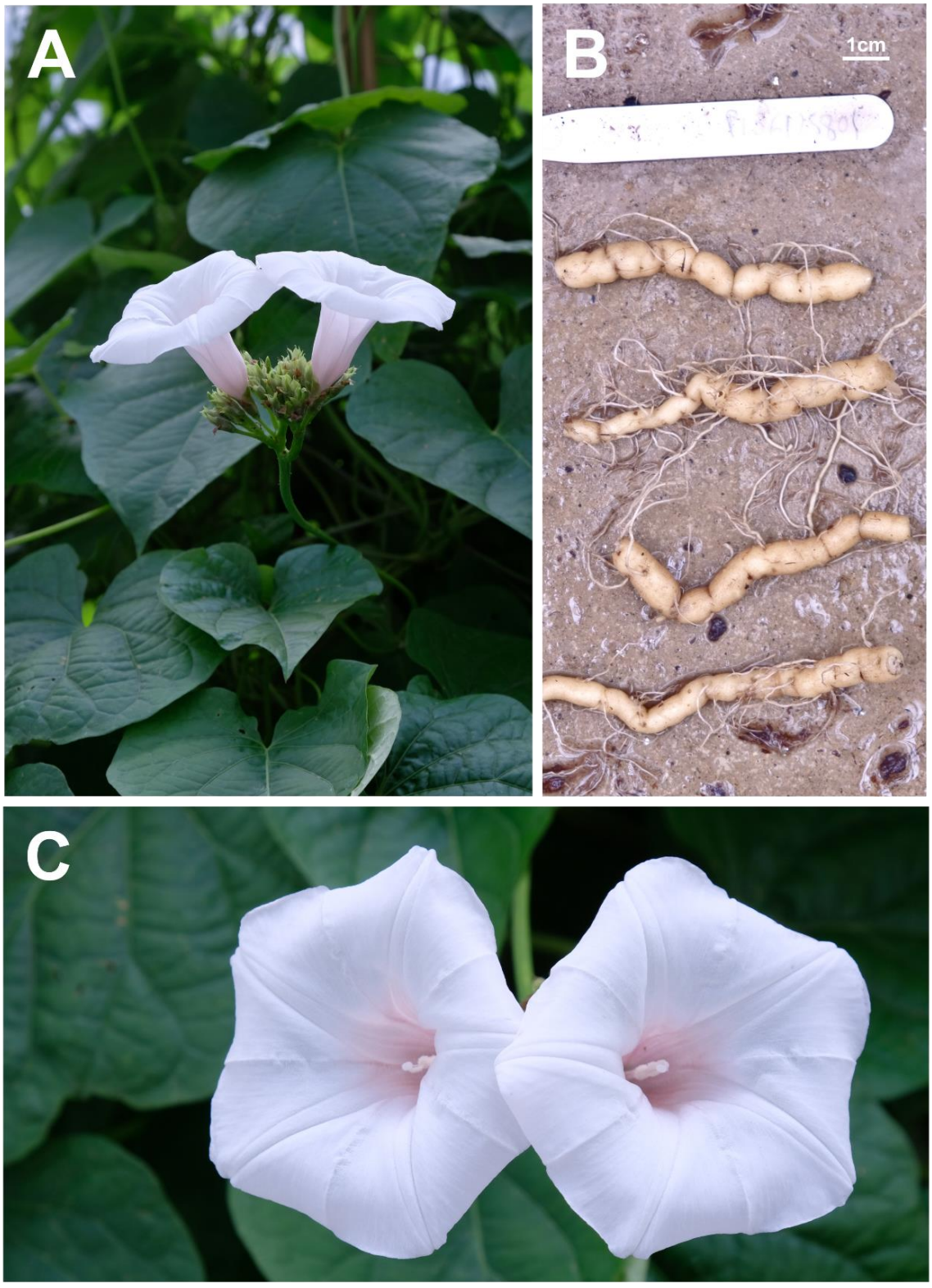


Figure 12: Germplasm collection *Austin 7817* (*I. aequatoriensis*), collected near El Guabo in Ecuador's southwest province of Los Rios. Grown from seeds in the greenhouse at University of Oxford Department of Biology. Showing **A:** broadly cordate leaves, and an inflorescence with more foliaceous sepals, still drying brown towards the tip; the corolla very pale pink to white, and narrowing towards the base, with a broad, open limb. **B:** small storage roots, 1-1.5cm in diameter, with white flesh and off-white skin. **C:** Broad corolla limbs and white stigma.

In particular, the corolla in the Ruiz and Pavon specimen only slightly exceeds the length of the sepals. Several specimens collected in coastal Manabí, Guayas and Loja were found to resemble *Asplund 15966*, though length of corolla, sepals and extent of leaf lobes were often less pronounced. This third morphotype was excluded from the analyses here however, due to its variability, more distinct morphology and the relative lack of collections. No examples of this morphotype were encountered in 2022, despite visits to the same localities as previous collections. It may represent an extreme form, a sporadic hybrid with another wild species or with cultivated sweet potato, or a separate species altogether, but further analysis will be required to resolve its exact status.

Despite the variability described here, the clustering analyses conducted in the previous chapter found specimens of the two main morphotypes described above to form a cohesive cluster distinct from closest relatives *I. trifida* and *I. batatas* (Chapter 3: Fig. 13). Even where *I. batatas* is found growing as a presumed escape from cultivation in the same areas of Western Ecuador (e.g. Fig. 9A), it can be readily distinguished from *I. aequatoriensis* by its highly unequal inner and outer sepals, the longest and innermost of which usually exceed 12mm in length.

Sepals of *I. batatas* are also generally pronouncedly oblong, with a variable tip from obtuse to acuminate, and often tinged purple (Fig. 9A). While these escapes are sometimes observed climbing on fences or in low trees and shrubs, their stems are consistently thicker than those of *I. aequatoriensis* (>3mm diameter), and their leaves are often highly lobed and tinged purple. Differences in corolla are more subtle and harder to quantify, but those of *I. batatas* tend to have a wider tube and less pronounced limb, making them appear more campanulate than those of *I. aequatoriensis*, which tend to have a wider limb, and to narrow towards the base of the tube (Fig. 2B-E).

Sepal form is perhaps the most important taxonomic character in *Ipomoea* (Austin 1978; Wood et al. 2020). As such, the variability in sepal morphology observed within *I. aequatoriensis* itself deserves consideration as to whether it is sufficient to justify further subdivision. Between the predominant morphotype with shorter more chartaceous sepals on the one hand, and those form west central Ecuador with longer, more acuminate, foliaceous sepals, and a longer, narrower corolla on the other. This was therefore considered with reference to the results of both the molecular phylogenetic analyses and the qualitative assessment of distribution and ecological data for the two morphotypes.

Genotypic Cohesion

The heterozygosity-aware phylogenetic analysis for a comprehensive set of 388 nuclear loci recovered all tetraploid samples from Ecuador and Colombia monophyletic, sister to hexaploid *I. batatas*, and distinct from other tetraploid material collected in Mexico (see previous chapter). A representative sample of the morphotype with foliaceous sepals from west-central Ecuador (*Austin 7817*) was recovered nested within the more prevalent morphotype from coastal Ecuador and Colombia. These results therefore provide support for a common ancestry and genotypic cohesion across both morphotypes.

The analysis of chloroplast data recovered the Colombian sample sister to *I. batatas* Lineage 2, and the Ecuadorian samples with Lineage 1. The exact explanation for this topology will require further analysis with a greater number of samples, but could be a result of introgression and chloroplast capture between *I. aequatoriensis* and the different lineages of hexaploid *I. batatas*. This could also explain why some Ecuadorian samples were recovered nested within Lineage 1 *I. batatas,* while others were sister to it.

An alternative explanation could be that the two *I. batatas* lineages arose independently from ancestors from separate populations in Ecuador and Colombia. This would not explain the monophyly of *I. batatas* in the nuclear analyses however, or the sister relationship between Lineage 2 and *I. trifida* in the chloroplast topology.

The incongruent and only partially resolved results of the chloroplast analysis and the fact that the tetraploid samples were recovered nested within hexaploid *I. batatas* in the denovo-assembly based results of the nuclear phylogenies reinforce the close and complex relationship between this material and *I. batatas.* As a result, it could potentially be argued that this material should continue to be referred to as a tetraploid race of *I. batatas*. However, incongruence between nuclear and chloroplast phylogenies is relatively common across plant species (Soltis 1991; Folk et al. 2017), and the denovo-assembled analyses fails to account for potentially divergent subgenomes in this group of mixed-ploidy taxa with a complex evolutionary history.

The difference in ploidy is of further relevance here because allopolyploid species (such as *I. batatas* appears to be) arguably require new names to differentiate them from their distinct progenitors. This is the case for fellow allohexaploid crop bread wheat (*Triticum aestivum* L.) for example, which is a hybrid between two wild species, *T. turgidum* L. and *Aegilops tauschii* Coss., themselves both wild occurring hybrids (Marcussen et al. 2014). In this sense, if *I. batatas* is an allohexaploid hybrid between an ancestor of the Ecuadorian and Colombian tetraploids and an ancestor of *I. trifida*, then simply continuing to refer to the descendants of one of those ancestors as a tetraploid race of *I. batatas* fails to clearly and accurately reflect that evolutionary history. An alternative option would be to also include diploid *I. trifida* in one large, mixed ploidy species, similar to what has previously been suggested by Japanese authors (Kobayashi 1984) (see thesis introduction).

But *I. trifida* is even more distinct from *I. batatas* than the Ecuadorian and Colombian material is, in terms of genotypic, phenotypic and ecotypic cohesion (see previous chapter and Muñoz-Rodríguez et al. 2018; Wood et al. 2020), making this option even harder to justify. Furthermore, despite some evidence of potential interploidy geneflow, the difference in ploidy is also likely to have led to a strong degree of reproductive isolation between these diploids, tetraploids and hexaploids (Ramsey and Schemske 1998; Tayalé and Parisod 2013). Analysis of heterozygosity patterns in the previous chapter also show that the Ecuadorian material is likely autotetraploid, with little evidence of past hybridization or geneflow. Subsequent attempts to assemble its genome have confirmed this, revealing a high degree of similarity between its two subgenomes (pers. comm. Dr. Muñoz-Rodríguez).

The Ecuadorian and Colombian tetraploids can therefore confidently be described as possessing clear evidence of genotypic cohesion, despite the potentially complex and only partially resolved evolutionary history for this group of taxa. The evidence presented here does not support independent genotypic cohesion for each of the two predominant morphotypes identified above, though population level-sampling would be required to definitively confirm that conclusion.

Ecotypic Cohesion:

Notes on habitat and ecology from existing herbarium specimens again revealed some variability across the collections from Colombia and Ecuador. As is common for the Batatas group and wider genus *Ipomoea*, the majority of collections are recorded from disturbed habitats, such as roadsides, riverbanks, secondary forest, and field edges. Beyond that, the habitats within which they have been collected are variously described as dry, humid, wet or transitional, semi-deciduous or evergreen forest or scrub, or herbaceous pasture and understorey.

The vast majority of collections made in 2022 were collected within a few metres of the roadside, most often amongst other herbaceous roadside weeds and small shrubs (Fig. 8A). Some effort was made to seek out collections in more natural habitats however, such as National Parks and regional reserves. Collections were made in Cerro Blanco and Manglare Churute for example, two reserves on east and southwest of the city of Guayaquil respectively.

These specimens were predominantly found growing in open areas such as clearings, in semi-evergreen seasonally dry tropical forest (SDTF). Here, they generally trailed along the ground or scrambled over fallen and low-growing vegetation, or up exposed slopes, sometimes reaching heights of 5-6m (Fig. 8D). Collections were also sought in the even drier deciduous SDTF of Machalilla National Park in the coastal province of Manabí, but only a few specimens were found, again by the roadside. Discussion with guides at the nearby community-run reserve of Agua Blanca led to the revelation that the plants – which are known locally as “Camotillo” – are common and highly abundant in the low-scrub and SDTF of the National Park, but only produce leaves and flowers in the wet season (December-April). The rest of the year, they are able to survive dormant, thanks to their small storage roots (c.1-2cm in diameter), which were also said to be popular food for browsing deer and goats. This likely explained our failure to collect any specimens in the park itself on our June visit at the height of the dry season.

Machalilla and its surroundings are characterized by low hills, with rapid changes in vegetation as elevation increases. Clouds coming in from the Pacific Ocean maintain relatively lush evergreen forest above ~200msl, and we therefore decided to look for material further up one of the hills in Agua Blanca. Plants were found alongside a rough vehicle tack leading to a telecoms mast at around 400msl (Fig. 9B). No plants were found even a few metres inside the humid forest either side of the track however.

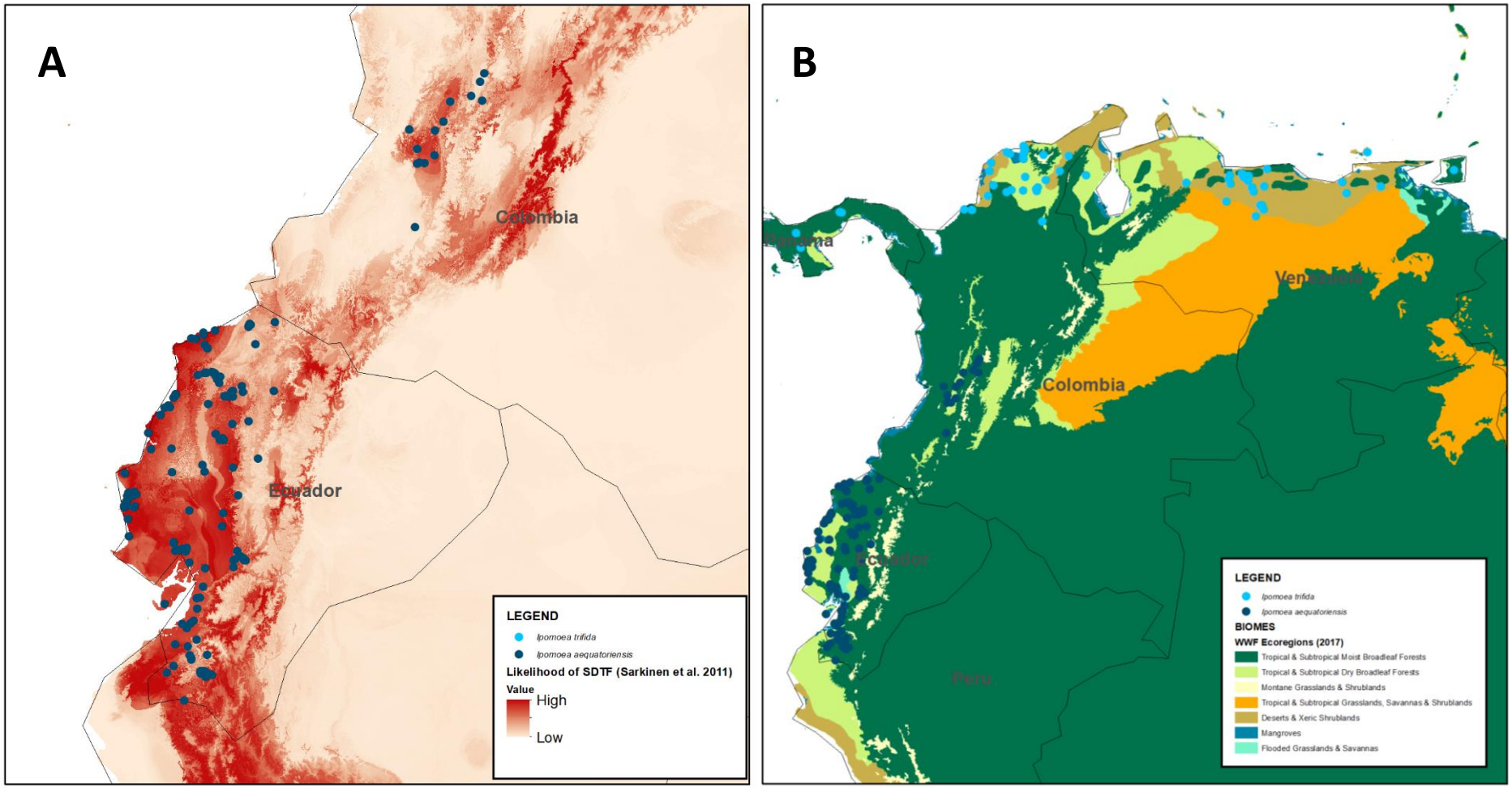


Figure 13: Biome maps based on **A:** the probablility of SDTF (Sarkinen et al. 2011); and **B:** Global Ecoregions sensu Dinerstein et al. (2017). Showing that specimens of *I. aequatoriensis* are most often collected in areas of high probability of SDTF, and that this may expalin the apparent disjunction between Ecuadorian and Colombian sepcimens, given the relative lack of SDTF in southwest Colombia.

It seems possible therefore that “Camotillo” originated as a disturbance specialist in SDTF, where its proto-storage roots allow it to survive seasonal drought, sprouting and flowering with the rains. Its prevalence in other areas outside this habitat may be a result of the disturbance caused by road building and agriculture allowing it to spread as a common weed. Here it can flower year-round in areas of greater moisture, which it would usually be unable to colonise due to competition and a more closed canopy.

This conclusion is also lent support by comparison of collection sites with the detailed map of predicted SDTF biome produced by Särkinen et al. (2011) (Fig. 13A). Collections from both Colombia and Ecuador are found almost exclusively in areas of high probability for SDTF. When the collection sites were instead compared to broader biome maps, such as that produced by (Dinerstein et al. 2017), the distribution had been harder to explain, given the presence of specimens in areas listed as both moist and dry broadleaf tropical forest (Fig. 13B). These broader biome maps are known to struggle with disturbed and fragmented habitats such as those found in lowland western Ecuador (Särkinen et al. 2011), which is characterised by a mosaic of different forest types, and has faced extensive deforestation over the last century (Dodson and Gentry 1991).

The steep altitudinal gradient of the Andes also appears to be a major factor in the distribution of these plants. No records exist east of the Andes, and Colombian collections are only recorded from the inter-Andean Valle de Cauca (Fig. 6). These Colombian collections – along with some from the southern Ecuadorian state of Loja – reach altitudes of up to 2000masl, but collection sites elsewhere are rarely above 400masl. Both of these high-elevation regions remain relatively dry however (Fig. 13), which again supports an apparent preference for drier habitats. At higher elevations with greater rainfall, no records are found, suggesting that presence of moist evergreen forest is a greater limiting factor than elevation itself.

The apparent disjunct distribution between Colombia and Ecuador may also be explained by this preference for drier habitats, given the relative lack of SDTF in southwest Colombia (Fig. 13), which is dominated by the Choco lowland rainforests – one of the wettest places on earth (Pérez-Escobar et al. 2019).

Variability in listed and observed habitat was also found to be common across the different morphotypes described above. One specimen of the less common morphotype collected in Ecuador in 2022 was found by the roadside in an area of wet forest in the Andean foothills of Loja (Fig. 8A). This morphotype has also previously been collected in Rio Palenque Science Station (Dodson and Gentry 1978), which was established by the University of Miami in 1971. Rio Palenque is described as one of the few remnants of primary wet forest in the upper Guayas basin, receiving around 3000mm of annual rainfall. However, the collections made here were still listed from the roadside and other disturbed areas rather than in the forest itself, and other specimens of this morphotype are collected from drier areas further south (e.g. *Austin 7817* in Loja). There is therefore only limited evidence for this less common morphotype being associated with a wetter habitat than the more prevalent one.

The production of thickened storage roots was also assessed in both existing germplasm accessions (Figs. 10B, 11B & 12B) and new collections made in Ecuador in June 2022 (Fig. 9B). Small storage roots, roughly 1-2cm in diameter were observed in all germplasm collections and in some new collections, though they were not always found across all newly collected material. It is possible that production of storage roots is mediated by environmental conditions, as has been observed in cultivated sweet potato (Meyers et al. 2013).

The fact that they were found in all the assessed germplasm material – including specimens of both morphotypes, and which were subjected to identical

environmental conditions – suggests that the potential for their production is common across *I. aequatoriensis*. These roots likely facilitate the adaptation to drier habitats proposed above, adding supporting evidence for ecotypic cohesion in *I. aequatoriensis*.

Taxonomic Treatment:

On the basis of the above assessments of phenotypic, genotypic and ecotypic cohesion, the material from Ecuador and Colombia was found to merit circumscription as a separate species. In particular, Colombian material was found to warrant inclusion here, on the basis that it falls within the morphological variation of that from Ecuador and that similar elevations are reached in southern Ecuador.

The disjunction in distribution is likely due to the lack of suitable habitat in southern Colombia's wet Choco. Defining characteristics of *I. aequatoriensis* are summarised below with minor updates based on analysis of the newly collected material presented here.

Ipomoea aequatoriensis T.Wells & P.Muñoz

TYPE: ECUADOR. *Austin, D. F. 7803*, (holotype FTG, isotype CIP).

Diagnosis: This species resembles *I. batatas* in terms its dense sub-umbellate inflorescence and pubescent ovary, but differs in possessing sepals that are consistently shorter (outer: <7mm vs > 7mm; inner: <10mm vs >12mm) and stems that are thinner (1-3mm vs. 2-6mm diam.) with longer internodes (6-16cm vs. 2-10cm), consistent with a twining (rather than trailing) habit. It also closely resembles *I. trifida*, particularly in its twining habit and somewhat chartaceous sepals, but differs in having generally obtuse sepals (50-160^{o} vs 20-70^{o}) and laxer, more obviously umbellate inflorescences with a greater number of flowers (5-24 vs 2-12) and mostly entire, generally larger leaves (4-14cm vs. 2-10cm long).

Description: Perennial twining herb to 3 metres. Roots fibrous or often somewhat thickened, forming small storage roots 1-2cm in diameter (Figs. 9-12B). Young stems terete, 1–3mm diameter, green, occasionally tinged purple above, drying reddish-brown on exposed surface, glabrous, pubescent or thinly to densely pilose, especially at nodes and towards tips, hairs sometimes asperous. Older stems angular, grey or grey-green, often verrucose, rooting at nodes where in contact with ground. Petioles 2–12.5 cm, thinly to densely pilose, with a pair of glands at junction with leaf base. Leaves petiolate, 4–14 × 3–13.5 cm, ovate, entire, or sometimes 3-lobed (Figs. 9B & 11B), or intermediate between the two states, with only small blunt teeth on widest point of lamina (Fig. 3); leaf apex shortly acuminate to cuspidate, with obtuse, mucronate tip; base broadly cordate, often deeply so; margins subentire to somewhat undulate; veins palmate, with prominent central rib; bright to deep green adaxially, paler abaxially; glabrous to thinly pubescent on both surfaces, often with hairs along veins, occasionally asperous to hispid across whole lamina. Inflorescence axillary, formed of solitary, pedunculate, usually dense, umbellate cymes with (5–)8–26 flowers; peduncles straight to slightly curved, rarely forked upwards, (3–)9–30 cm, glabrous or thinly pilose; short secondary and tertiary peduncles often present, 4-10 mm, noticeably thicker than pedicels; bracteoles c. 1–1.5 x 0.5 mm, ovate-deltoid, acuminate, chartaceous, caducous; pedicels (3–)6–12(–17)mm, slender, glabrous. Sepals unequal, the outermost 3–7mm long, the innermost 7–10mm long, oblong-elliptic to obovate, obtuse and mucronate, or less commonly acuminate, generally somewhat chartaceous, especially at margins and towards tip, the central vein dark and prominent, glabrous, ciliate or densely pilose, especially towards sepal base, often drying brown at anthesis, and folded or reflexed in apical third. Corolla tube 3–4.5 cm long, funnel-shaped to salver-form, with wide limb and tube narrowing towards base, pale or rarely vivid pink link and upper tube, with darker centre, glabrous on the exterior. Stamen filaments unequal, two long and three shorter, glabrous with the exception of glandular-hairs in basal third of filament; longest filament c. 15 mm,

shortest filament c. 6 mm; anthers 2.5 mm long, linear with sagittate base; style c. 20 mm long, glabrous; ovary spherical to slightly flattened, pubescent, often densely so. Capsules c. 5 x 6 mm, appressed ovoid, pilose; seeds 3 x 2 mm, dark brown to black, often somewhat mottled, entirely glabrous.

Habitat: Common in disturbed areas such as roadsides, secondary forest, clearings and riverside scrub. Especially in areas of seasonally dry tropical forest. 0–750(–2200)masl.

Distribution: Western provinces of Ecuador including Azuay, Bolivar?, Cañar, Chimborazo, El Oro, Esmeraldas, Guayas, Loja, Los Ríos, Manabí, Santo Domingo, and Santa Elena?. Southwestern provinces of Colombia including Cauca, Quindio, Risaralda, Tolima and Valle de Cauca.

Etymology: The name *aequatoriensis* was chosen for its distribution either side of the Equator in Ecuador and Colombia.

See Appendix 4 for full specimen citation.

CONCLUSION:

Wild-collected tetraploid plants from Ecuador and southern Colombia were found to warrant description as a new species, separate from sweet potato (*I. batatas*) and close CWR *I. trifida*. Analysis of existing and new collections shows that these plants occupy a distinct distribution and habitat on the west side of the Andes, where they are common to disturbed areas, particularly in SDTF. While they resemble a weedier form of *I. batatas*, they are also notably distinct in possessing shorter, less unequal sepals, often with a blunt tip, and generally chartaceous for all or part of their length.

While resolving their exact relationship to *I. batatas* remains complicated by the difference in ploidy and possibility of recent gene flow, this material is monophyletic, sharing a recent common ancestor, and it is autotetraploid with little evidence of recent geneflow. *I. aequatoriensis* can therefore be described as a distinct evolutionary lineage, with combined evidence of genotypic, phenotypic and ecotypic cohesion.

Muñoz-Rodríguez et al. (2022) concluded that the Colombian material required further analysis before deciding whether to include it within *I. aequatoriensis*. In particular, they cited the apparent disjunction in distribution and the high elevation of Colombian collection sites for this decision. Further analysis here leads me to believe both appear to be explained by a preference for seasonally dry habitats. The lack of collections from the very south west of Colombia makes sense when viewed in the context of probability of SDTF (Fig. 13A). Meanwhile, existing and new collections of *I. aequatoriensis* made in the southern Ecuadorian state of Loja were found at similar altitudes up to 2000m. Loja is noticeably drier than at similar elevations further north in Ecuador, thus resembling the inter-Andean Cauca valley of Colombia.

Variability in sepal morphology within west-central Ecuador was not found to warrant recognition as a separate species. Population-level sampling and further collecting within the region will be required to confirm this categorically, but it appears unlikely that this variability warrants recognition beyond the rank of variety or subspecies, if at all.

An additional consideration that appears especially pertinent here is a desire for nomenclatural stability and greater clarity in discussions of the delimitation and evolutionary relationships among sweet potato and its CWRs. In Chapter 2, I outlined the extent of past taxonomic confusion in relation to a heterogeneous set of

tetraploid germplasm collections that had previously been associated with sweet potato and hypothesised to be its progenitor. This past confusion – particularly in terms of the names applied to different material – has complicated communication between different authors, and comparison among different studies, hindering research and crop improvement breeding programmes. Treating the autotetraploid collections from Ecuador and Colombia as a race of *I. batatas* or *I. trifida* has failed to distinguish them both from the cultivated crop and other material of likely feral or hybrid origin collected elsewhere in Latin America, from which they are clearly distinct. It is hoped therefore, that their description as a separate species and the discussion of their diversity presented here will alleviate this confusion, thereby facilitating improved research into the origin and contemporary diversity of sweet potato.

REFERENCES:

Austin D.F. 1977. Hybrid polyploids in *Ipomoea* section batatas. J. Hered. 62:259–260.

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

Austin D.F. 1982. Convolulaceae. In: Harling G., Sparre B., editors. Flora of Ecuador. Stockholm: Swedish Research Councils. p. 30–69.

Austin D.F., Jarret R.L., Tapia C., de la Puente F. 1993. Collecting tetraploid *I. batatas* (L.) Lam. in Ecuador. FAO/IBPGR Plant Genet. Resour. Newsl. 91/92:33–35.

Bachmann K. 1998. Species as units of diversity: an outdated concept. Theory Biosci. 117.

Bauman D., Fortunel C., Cernusak L.A., Bentley L.P., McMahon S.M., Rifai S.W., Aguirre-Gutiérrez J., Oliveras I., Bradford M., Laurance S.G.W., Delhaye G., Hutchinson M.F., Dempsey R., McNellis B.E., Santos-Andrade P.E., Ninantay-Rivera H.R., Chambi Paucar J.R., Phillips O.L., Malhi Y. 2022. Tropical tree growth sensitivity to climate is driven by species intrinsic growth rate and leaf traits. Glob. Chang. Biol. 28:1414–1432.

Bohac J.R., Austin D.F., Jones A. 1993. Discovery of Wild Tetraploid Sweetpotatoes. Econ. Bot. 47:193–201.

Boyd R. 1991. Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. Philos. Stud. 61:127–148.

Boyd R. 1999. Homeostasis, Species and Higher Taxa. In: MIT Press, editor. Species: New interdisciplinary essays. p. 1–21.

Brundin L. 1972. Phylogenetics and Biogeography. Syst. Biol. 21:69–79.

de Candolle A. 1855. Géographie botanique raisonnée &c. Paris: Masson.

Darwin C. 1859. On the orign of species under natural selection, or Preservation of favoured races in the struggle for life. London: John Murray.

Dayrat B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85:407–415.

Dinerstein E., Olson D., Joshi A., Vynne C., Burgess N.D., Wikramanayake E., Hahn N., Palminteri S., Hedao P., Noss R., Hansen M., Locke H., Ellis E.C., Jones B., Barber C.V., Hayes R., Kormos C., Martin V., Crist E., Sechrest W., Price L., Baillie J.E.M., Weeden D., Suckling K., Davis C., Sizer N., Moore R., Thau D., Birch T., Potapov P., Turubanova S., Tyukavina A., De Souza N., Pintea L., Brito J.C., Llewellyn O.A., Miller A.G., Patzelt A., Ghazanfar S.A., Timberlake J., Klöser H., Shennan-Farpón Y., Kindt R., Lillesø J.P.B., Van Breugel P., Graudal L., Voge M., Al-Shammari K.F., Saleem M. 2017. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. Bioscience. 67:534–545.

Dodson C.H., Gentry A.H. 1978. FLORA OF THE RIO PALENQUE SCIENCE CENTER: Los Rios Province ECUADOR. Selbyana. 4:301–302.

Dodson C.H., Gentry A.H. 1991. Biological Extinction in Western Ecuador. Source Ann. Missouri Bot. Gard. 78.

Dodson C.H., Gentry A.H., Valverde F.M. 1985. La Flora de Jauneche, Los Rios, Ecuador. Quito: Banco Central del Ecuador.

Ereshefsky M. 1998. Species Pluralism and Anti-Realism. Philos. Sci. 65:103–120.

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

Freudenstein J. V., Broe M.B., Folk R.A., Sinn B.T. 2017. Biodiversity and the Species Concept - Lineages are not Enough. Syst. Biol. 66:644–656.

Gao L., Rieseberg L.H. 2020. While neither universally applicable nor practical operationally, the biological species concept continues to offer a compelling framework for studying species and speciation. Natl. Sci. Rev. 7:1398–1400.

Ghiselin M.T. 1987. Species concepts, individuality, and objectivity. Biol. Philos. 2:127–143.

Gigerenzer G., Brighton H. 2009. Homo Heuristicus: Why Biased Minds Make Better Inferences. Top. Cogn. Sci. 1:107–143.

Gilmour J.S.L. 1958. The Species: Yesterday and To-morrow. Nature. 181:379–380.

Hendry A.P., Vamosi S.M., Latham S.J., Heilbuth J.C., Day T. 2000. Questioning species realities. Conserv. Genet. 1:67–76.

Hillebrand H. 2003. On the Generality of the Latitudinal Diversity Gradient. Am. Nat. 163:192–211.

Hillis D.M. 2020. The Detection and Naming of Geographic Variation Within Species. Herpetol. Rev. 51:52–56.

Hughes C.E., Pennington R.T., Antonelli A. 2013. Neotropical Plant Evolution: Assembling the Big Picture. Bot. J. Linn. Soc. 171:1–18.

Hull. 1978. A Matter of Individuality. Philos. Sci. 45:335–360.

Hull D.L. 1976. Are species really individuals? Syst. Zool. 25:174–191.

Hull D.L. 1997. The ideal species concept - and why we can't get it. In: Claridge M.F., Dawah H.A., Wilson M.R., editors. Species: The Units of Biodiversity1. London: Chapman & Hall. p. 357–377.

von Humboldt A. 1805. Essai sur la geographie des plantes; accompagne d'un tableau physique des régions equinoxiales. Paris: Levrault.

Kobayashi M. 1984. The *Ipomoea trifida* Complex Closely Related to Sweet Potato. Proceedings, Sixth Symposium of the International Society for Tropical Root Crops, hosted by CIP in Lima, Peru. p. 561–568.

Kulkarni D., Simon H.A. 1988. The Processes of Scientific Discovery: The Strategy of Experimentation. Cogn. Sci. 12:139–175.

Landrum L.R. 2001. What has happened to descriptive systematics? What would make it thrive? Syst. Bot. 26:438–442.

Lee M.S.Y. 2000. A worrying systematic decline. Trends Ecol. Evol. 15:346.

Maclaurin J., Sterelny K. 2008. What is Biodiversity? Chicago: The University of Chicago Press.

Mallet J. 2020. Alternative views of biological species: reproductively isolated units or genotypic clusters? Natl. Sci. Rev. 7:1400–1401.

Marcussen T., Sandve S.R., Heier L., Spannagl M., Pfeifer M., The International Wheat Genome Sequencing Consortium, Jakobsen K.S., Wulff B.B.H., Steuernagel B., Mayer K.F.X., Olsen O.-A. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. Science (80). 345:1285–1288.

Marewski J.N., Gigerenzer G. 2012. Heuristic decision making in medicine. Dialogues Clin. Neurosci. 14:77–89.

Martin F.W., Jones A., Ruberte R.M. 1974. A Wild *Ipomoea* Species Closely Related to the Sweet Potato. Econ. Bot. 28:287–292.

Mayr E. 1963. Animal Species and Evolution. Cambridge, MA: Belknap Press of Harvard University.

Meyers S.L., Arancibia R.A., Shankle M.W., Main J., Gajanayake B., Reddy K.R. 2013. Sweet Potato Storage Root Initiation. Publ. 2809. Publicatio.

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., Rausher M.D., Kelly S., Liston A., Scotland R.W. 2018. Reconciling Conflicting Phylogenies in the Origin of Sweet Potato and Dispersal to Polynesia. Curr. Biol. 28:1246-1256.e12.

Muñoz-Rodríguez P., Carruthers T., Wood J.R.I., Williams B.R.M., Weitemier K., Kronmiller B., Goodwin Z., Sumadijaya A., Anglin N.L., Filer D., Harris D., Rausher M.D., Kelly S., Liston A., Scotland R.W. 2019. A taxonomic monograph of *Ipomoea* integrated across phylogenetic scales. Nat. Plants. 5:1136–1144.

Muñoz-Rodríguez P., Wells T., Wood J.R.I., Carruthers T., Anglin N.L., Jarret R.L., Scotland R.W. 2022. Discovery and characterization of sweetpotato's closest tetraploid relative. New Phytol. 234:4.

Myers N., Mittermeier R.A., Mittermeier C.G., da Fonseca G.A.B., Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature. 403:853–858.

Padial J.M., De la Riva I. 2020. A paradigm shift in our view of species drives current trends in biological classification. Biol. Rev.

Padial J.M., De La Riva I. 2010. A response to recent proposals for integrative taxonomy. Biol. J. Linn. Soc. 101:747–756.

Paterson H.E.H. 1985. The recognition concept of species. In: Vrba E.S., editor. Species and Speciation. Pretoria: Transvaal Museum Monograph. p. 21–29.

Pérez-Escobar O.A., Lucas E., Jaramillo C., Monro A., Morris S.K., Bogarín D., Greer D., Dodsworth S., Aguilar-Cano J., Sanchez Meseguer A., Antonelli A.

2019. The Origin and Diversification of the Hyperdiverse Flora in the Chocó Biogeographic Region. Front. Plant Sci. 10:1–9.

Pigliucci M. 2003. Species as family resemblance concepts: The (dis-)solution of the species problem? BioEssays. 25:596–602.

Pleijel F. 1999. Phylogenetic taxonomy, a farewell to species, and a revision of *Heteropodarke* (Hesionidae, Polychaeta, Annelida). Syst. Biol. 48:755–789.

Pleijel F., Rouse G.W. 2000. Least-inclusive taxonomic unit: A new taxonomic concept for biology. Proc. R. Soc. B Biol. Sci. 267:627–630.

de Queiroz K. 1998. The General Lineage Concept of Species, Species Criteria, and the Process of Speciation: A conceptual unification and terminological recommendations. In: Howard D.J., Berlocher S.H., editors. Endless Forms: Species and Speciation. New York: Oxford University Press. p. 57–75.

de Queiroz K. 1999. The General Lineage Concept of Species and the Defining Properties of the Species Category. In: Wilson R.A., editor. Species: New interdisciplinary essays. Cambridge, Massachusetts: p. 49–89.

de Queiroz K. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

Ramsey J., Schemske D.W. 1998. Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. Annu. Rev. Ecol. Syst. 29:467–501.

Roullier C., Duputié A., Wennekes P., Benoit L., Fernández Bringas V.M., Rossel G., Tay D., McKey D., Lebot V. 2013. Disentangling the Origins of Cultivated Sweet Potato (*Ipomoea batatas* (L.) Lam.). PLoS One. 8.

Särkinen T., Iganci J.R.V., Linares-Palomino R., Simon M.F., Prado D.E. 2011. Forgotten forests - issues and prospects in biome mapping using Seasonally Dry Tropical Forests as a case study. BMC Ecol. 11.

Simpson G.G. 1951. The Species Concept. Evolution. 5:285–298.

Soltis D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants. 5:65–84.

Sukumaran J., Knowles L.L. 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. U. S. A. 114:1607–1611.

Tayalé A., Parisod C. 2013. Natural pathways to polyploidy in plants and consequences for genome reorganization. Cytogenet. Genome Res. 140:79–96.

Templeton A.R. 1989. The Meaning of Species and Speciation. In: Endler J.A., Otte D., editors. Speciation and its Consequences. Sinauer Associates. p. 3–27.

Vrana P., Wheeler W. 1992. Individual Organisms As Terminal Entities: Laying the Species Problem To Rest. Cladistics. 8:67–72.

Wallace A.R. 1876. The geographical distribution of animals. London: Macmillan.

Wang X., He Z., Shi S., Wu C.-I. 2020. Genes and speciation: is it time to abandon the biological species concept? Natl. Sci. Rev. 7:1387–1397.

Wells T., Carruthers T., Muñoz-Rodríguez P., Sumadijaya A., Wood J.R.I., Scotland R.W. 2021a. Species as a Heuristic: Reconciling Theory and Practice. Syst. Biol. 71(5):1233-1243.

Wells T., Carruthers T., Scotland R.W. 2022. Heuristics, species, and the analysis of systematic data. Trends Plant Sci. 27:524–527.

Wheeler Q.D. 2020. A taxonomic renaissance in three acts. Megataxa 001:4–8. https://doi.org/10.11646/megataxa.1.1.2

Will K.W., Mishler B.D., Wheeler Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54:844–851.

Wilson E.O. 2000. On the future of conservation biology: Editorial. Conserv. Biol. 14:1–3.

Wilson R., Barker M., Brigandt I. 2007. When Traditional Essentialism Fails: Biological Natural Kinds. Philos. Top. 35:189–215.

Wittgenstein L. 1953. Philosophical Investigations. New York: Macmillan.

Wood J.R.I., Muñoz-Rodríguez P., Williams B.R.M., Scotland R.W. 2020. A foundation monograph of *Ipomoea* (Convolvulaceae) in the New World. PhytoKeys. 143:1–823.

Yeates D.K., Seago A., Nelson L., Cameron S.L., Joseph L., Trueman J.W.H. 2011. Integrative taxonomy, or iterative taxonomy? Syst. Entomol. 36:209–217.

Zachos F.E. 2016. Species Concepts in Biology. Spinger International Publishing Switzerland

Zografou K., Swartz M.T., Adamidis G.C., Tilden V.P., McKinney E.N., Sewall B.J. 2021. Species traits affect phenological responses to climate change in a butterfly community. Sci. Rep. 11:1–14.

CHAPTER 5 (THESIS CONCLUSION): The identification and description of sweet potato's closest wild relative paves the way for new research on its evolution and diversity

In the introduction to this thesis, I outlined the centrality of tetraploid plants resembling wild forms of sweet potato (*I. batatas* (L.) Lam.) for resolving outstanding questions about the evolution and contemporary diversity of this important global crop. Understanding the origin of sweet potato's hexaploid genome, its thickened storage roots and its two separate chloroplast lineages is dependent on resolved phylogenetic relationships and improved taxonomy for these tetraploid plants. Despite its global agricultural, dietary and economic importance, it has remained unclear whether sweet potato is auto- or allo-hexaploid, how or when it evolved its thickened storage roots, or why it has two separate chloroplast lineages. The failure to identify the tetraploid progenitor of hexaploid *I. batatas* and distinguish it from potential hybrids has thus impeded understanding on these issues for almost a century.

In Chapter 2, I reviewed the reasons for the lack of progress in the identification of this important crop wild relative (CWR), with reference to the confusion surrounding a heterogeneous set of tetraploid germplasm collections from across Mexico, Central and South America. These collections had been identified under multiple different names and variously suggested to be involved in the origin of hexaploid *I. batatas*, but only limited and circumstantial evidence existed in support of any of these conclusions. Details of their morphology, genomic configuration and phylogenetic relationships remained partial, often contradictory, and therefore unresolved. The failure to cross-reference and accurately cite these different accessions had also obscured the diversity present across these different collections.

Subsequent to being incorrectly assigned to various species within the Batatas group of sweet potato's CWRs, they had latterly begun to be treated as a single homogenous entity, often referred to as either "*I. trifida* 4X" or "*I. batatas* 4X". This simplification had both obscured the different characteristics and likely evolutionary origins of the different tetraploid collections, and undermined the resulting conclusions drawn on the basis of different analyses. Resolving this confusion – the aim of this thesis – was thus essential to progress in sweet potato research and breeding programs. I uncovered some evidence for differentiation between material collected in Ecuador and southern Colombia, and that from Central America and Mexico. This distinction had not been widely discussed or accepted however, and was based only on partial evidence, such as morphological differences, karyology, and Sanger barcode markers.

In Chapter 3, I used a combination of molecular phylogenetic, genomic and morphometric analyses to conclusively distinguish between two distinct entities of tetraploid germplasm collections. The results showed that those from Ecuador and Colombia are autotetraploid, monophyletic and the closest extant relative of hexaploid *I. batatas*, and that they are morphologically distinct from *I. batatas, I. trifida* and other tetraploid collections from Mexico. By contrast, tetraploid accessions from Mexico are most likely the product of recent hybridization between hexaploid *I. batatas* and diploid *I. trifida*. Other tetraploid germplasm material and herbarium specimens will need to be included in subsequent analyses to confirm this result for all Central American and Mexican tetraploids.

The genomic analyses will also need to be repeated with higher coverage sequencing in order to verify their apparent possession of two divergent subgenomes with more confidence. What is clear however, is that these two groups of tetraploids originated separately and likely by different means. Ecuadorian and Colombian tetraploids are most likely the descendants of sweet potato's tetraploid

progenitor, constituting a vital discovery in crop biology and CWR research that will dramatically improve efforts to understand its evolution and contemporary diversity.

In Chapter 4, I reviewed all known information about this newly described species *I. aequatoriensis* T.Wells & P.Munoz, based on analysis of existing herbarium and germplasm collections, and new collections made in 2022. This analysis showed that *I. aequatoriensis* is likely a disturbance specialist, predominantly from seasonally dry tropical forests (SDTF) in western Ecuador and southern Colombia, with its small storage roots enabling it to survive periods of drought, and its twining and trailing stems allowing it to seek out light in spatiotemporal canopy gaps. *Ipomoea aequatoriensis* is readily distinguished from *I. trifida* by its short, generally blunt sepals, thicker stems that turn grey-green and verrucose with age, and by its large, generally cordate leaves, and its dense umbellate cymes. Its thinner stems and twining habit differentiate it from *I. batatas*, as do its generally at least partially chartaceous sepals, which are shorter than 1cm in length and often dry brown and reflex towards the tip. Its phylogenetic position, distinctive morphology, apparent adaptation to SDTF, and contiguous distribution distinct from that of *I. trifida,* together provide evidence for clear genotypic, phenotypic and ecotypic cohesion. The use of names like “Camotillo” and “Batatilla” in the region also suggest a recognition both of its close relationship to *I. batatas*, and its distinction from it.

That *I. aequatoriensis* is closely related to *I. batatas* is underlined by the continued difficulty of recovering a fully resolved topology in phylogenetic analysis of nuclear data, and the fact that some Ecuadorian samples are recovered nested within *I. batatas* chloroplast Lineage 1. The possession of a Lineage 2 chloroplast by the single Colombian specimen analysed further illustrates this close relationship and the fact that more research will be required to resolve exactly how *I. batatas* evolved from a tetraploid ancestor shared with *I. aequatoriensis*. Despite these

unresolved issues it is now possible to discuss the three questions I outlined in the introduction to this thesis with greater confidence and clarity.

The first of these was how sweet potato evolved its hexaploid genome from a diploid ancestor shared with *I. trifida*. The phylogenetic and genomic results presented in Chapter 3 provide tentative support for an allohexaploid origin through hybridization between the ancestors of *I. trifida* and *I. aequatoriensis*.

First, the autotetraploid nature of the *I. aequatoriensis* genome and its position sister to, or nested within hexaploid *I. batatas* (depending on assembly method) strongly suggest that it is the direct descendant of sweet potato's tetraploid progenitor.

Second, the topology of the results from phylogenetic analysis of the denovo-assembled loci indicate that this method struggles to resolve the relationships in a manner that is potentially suggestive of shared ancestry for *I. batatas* derived from both *I. trifida* and *I. aequatoriensis*. This is because the position of *I. aequatoriensis* nested within *I. batatas* is likely the result of the failure to account for heterozygosity between different *I. batatas* subgenomes. If *I. batatas* was autohexaploid and possessed three identical subgenomes, then this issue would not arise, and there would be no incongruence between the results of the denovo-assembled and genotyped loci. The fact that *I. aequatoriensis* is nested within *I. batatas* in the denovo-assembled results suggests that the two share a large number of alleles. But the fact that a heterozygosity-aware genotyping pipeline instead produces a topology where the two are reciprocally monophyletic, suggests the denovo-assembly method is obscuring important heterozygosity. The fact that when this heterozygosity is artificially masked in this manner, *I. aequatoriensis* is recovered nested within *I. batatas* indicates that the majority of *I. batatas* subgenomes likely strongly resemble those of *I. aequatoriensis*.

While this is far from conclusive evidence (and see below for an alternative explanation), it does suggest that *I. batatas* possesses important heterozygosity among its subgenomes, and that two of them are highly similar to those of *I. aequatoriensis*, while the third differs, consistent with an allohexaploid origin.

Further, the results of the genomic analysis in Chapter 3 for putative hybrid *I. tabascana* also provide potential support for an allohexaploid *I. batatas* genome. This is because *I. tabascana* and other Mexican tetraploid germplasm accessions are found to have a genomic structure resembling that of an allotetraploid. Since *I. tabascana* is not thought to be a true allotetraploid, but rather a recent interploidy hybrid between *I. batatas* and *I. trifida*, this suggests that the structure of the *I. batatas* genome is most likely that of an allohexaploid (AABBBB). This would have to be the case, such that a hybrid between AABBBB and the descendant of its presumed diploid progenitor (*I. trifida*) AA, will produce a tetraploid (ABB + A = AABB), which will artificially resemble a true allotetraploid in these analyses. This evidence is again circumstantial and based on somewhat circular logic, but it is suggestive, since it is difficult to provide an alternative explanation for the phylogenetic position and genomic structure observed in *I. tabascana* and other Mexican tetraploid material.

Most importantly, the discovery of autotetraploid plants sister to hexaploid *I. batatas* in heterozygosity-aware phylogenetic analyses provides the opportunity to produce a more accurate assembly of the different hexaploid *I. batatas* subgenomes and resolve this debate once and for all. Once the genome of *I. aequatoriensis* has been assembled, partially assembled contigs of *I. batatas* can then be compared to it and to the already assembled genome of *I. trifida.* This should facilitate an accurate chromosome-level assembly of all three subgenomes. Sweet potato is currently one of the few globally important crops without an accurately assembled

genome, and the results outlined in Chapter 3 represent an important step towards producing one.

Progress has also been made towards a better understanding of the origin of large, edible storage roots in hexaploid *I. batatas*. In Chapter 4 I provide photographs and descriptions of small storage roots observed in wild plants and cultivated germplasm accessions of *I. aequatoriensis*. These organs are in no sense comparable to those produced by cultivated hexaploid *I. batatas,* which indicates that those of *I. batatas* most likely originated after their ancestors diverged. This divergence between tetraploid *I. aequatoriensis* and hexaploid *I. batatas* has not yet been dated, but this result shows that the storage roots of sweet potato are likely to be younger than 800,000 years (Muñoz-Rodríguez et al. 2018), and that they most likely originated after the transition to hexaploidy. The close relationship between *I. aequatoriensis* and *I. batatas* recovered in Chapter 3 suggests that this transition may in fact have been relatively recent, in that there appears to have been little time for divergence to occur between the two.

The assembly of the *I. aequatoriensis* genome may also facilitate the search for relevant genes in storage root development and the comparison between these genes in *I. trifida*, *I. aequatoriensis* and *I. batatas*. This will in turn provide a more detailed understanding of the development of this important organ, grounded in a more complete understanding of the evolutionary history of *I. batatas* than was previously available. This represents a fundamental improvement on the situation I outlined in Chapter 2, where previous such studies had relied solely on the genome of *I. trifida*, in combination with a flawed interpretation of phylogenetic results based on unidentified polyploid germplasm material (e.g. Feng et al. 2018; Li et al. 2019).

Finally, the implications of the results presented here for the origin of sweet potato's two chloroplast lineages are less clear. Support for the chloroplast capture

hypothesis is potentially undermined by the possession by the Colombian specimen (K500) of a Lineage 2 chloroplast, the fact that this specimen and the putative Mexican hybrids are sister to Lineage 2 *I. batatas* rather than nested within it, and the apparent predominance of maternal inheritance of the chloroplast in putative hybrids between *I. batatas* and *I. trifida.* Alternative hypotheses are also difficult to reconcile with these results however.

Taken in isolation, the results of the chloroplast analysis in Chapter 3 appear to suggest a dual origin for hexaploid *I. batatas* from two different tetraploid progenitors. "Southern" or Lineage 1 *I. batatas* from an ancestor shared with the Ecuadorian tetraploids, and "Northern" or Lineage 2 *I. batatas* from one shared with the Colombian or Mexican tetraploids. Ignoring for a moment the apparent difficulty of reconciling this hypothesis with the monophyly of hexaploid *I. batatas* in the results of the phylogenetic analyses of nuclear data, there are two possible ways in which this could have occurred.

The first is that both lineages are allohexaploid, but from crosses between the diploid ancestor of *I. trifida* and the ancestors of two different populations of *I. aequatoriensis* (Ecuador for Lineage 1, and Colombia for Lineage 2). In Chapter 4, I showed that the disjunct distribution between these two populations of contemporary *I. aequatoriensis* appears to be based on a preference for an SDTF habitat, which is largely absent from the lowland Chocó in southwest Colombia. The contemporary Chocó biome is believed to have originated by the early Pleistocene (~2.7Mya) (Pérez-Escobar et al. 2019), meaning that these populations may have been isolated for some time. While I found no indication of any divergence in morphology and ecology between these two populations, in Chapter 3 I uncovered some evidence indicating this isolation may have led to genetic divergence between them.

In the results for the phylogenetic analysis of the genotyped loci for example, the Colombian sample (K500) is recovered sister to all Ecuadorian samples. Meanwhile, in the denovo-assembled results, K500 is instead recovered sister to a well-supported clade containing the Ecuadorian tetraploids and several specimens of hexaploid *I. batatas*. These analyses will need to be repeated with a higher number of Colombian and Ecuadorian samples, but it suggests there may be some genetic structure within *I. aequatoriensis*, corresponding to its disjunct distribution between the two countries. What this hypothetical divergence in the nuclear genome cannot explain however, is why chloroplast Lineage 1 is recovered sister to a well-supported clade containing Lineage 2 and *I. trifida*. If tetraploid *I. aequatoriensis* diverged from an ancestor shared with diploid *I. trifida*, before then itself diverging into distinct Colombian and Ecuadorian populations – each of which then gave rise to a separate lineage of hexaploid *I. batatas* – we would expect the two chloroplast lineages to be recovered sister to each other. That they are not undermines this hypothesis.

An alternative hypothesis is that Lineage 1 arose as an autohexaploid, entirely from an ancestor shared with *I. aequatoriensis*, while Lineage 2 is an allohexaploid derived from a hybridization event between an ancestor of *I. aequatoriensis* and an ancestor of *I. trifida* (See Chapter 1: Fig. 2B-C). If the chloroplast was donated by the ancestor of *I. trifida*, then this could explain the separate chloroplast lineages, and the fact that Lineage 2 is more closely related to *I. trifida* than it is to Lineage 1.

Potential support for this scenario can again be found in the topology of the nuclear phylogenetic results presented in Chapter 3. Using Chapter 3: Fig.1, I explained the topology of the denovo-assembled results as likely to be an artefact of the assembly process, resulting from the rejection of less common alleles derived from the hypothetical AA subgenome.

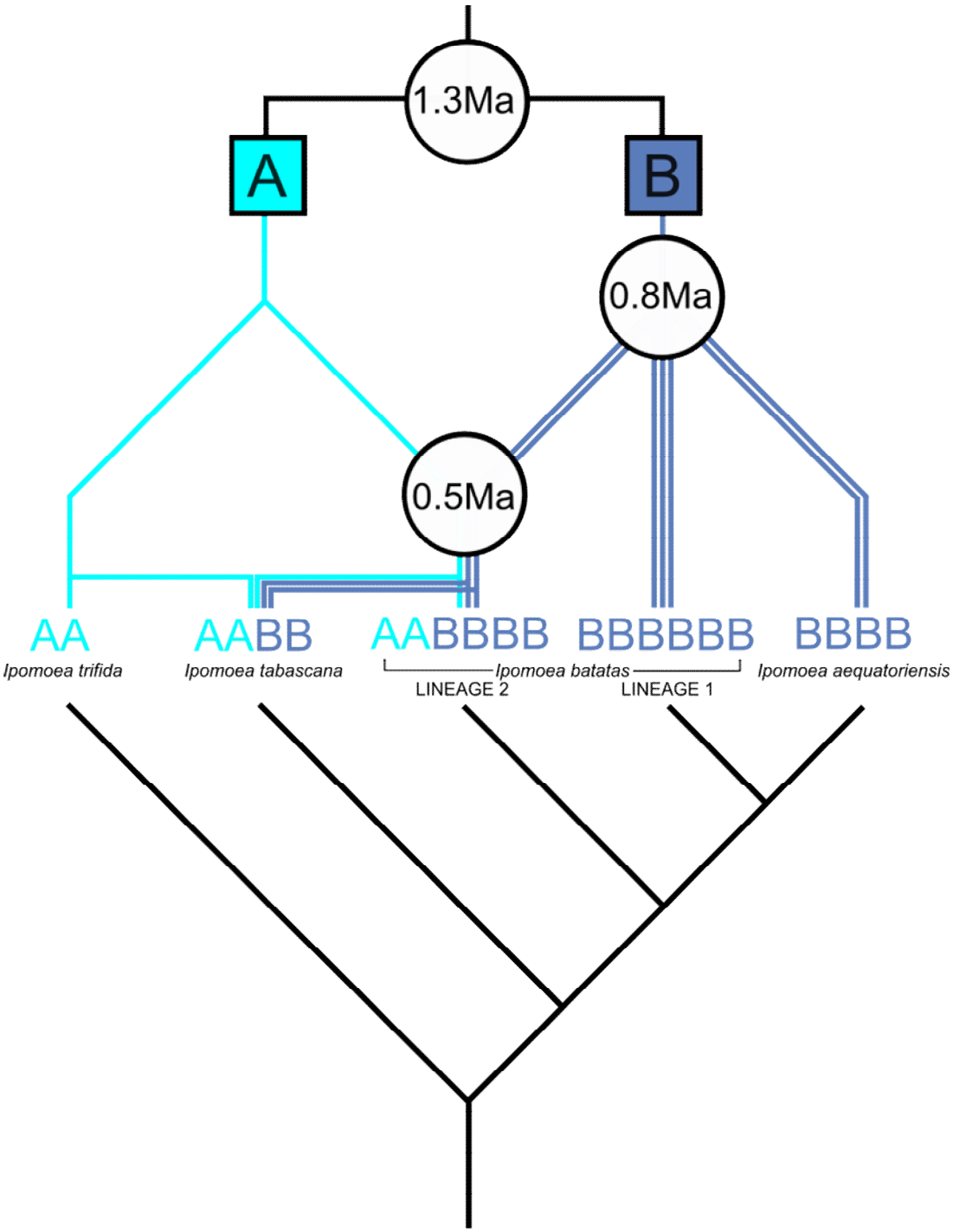


Figure 1: Top half showing potential evolutionary relationships if *I. batatas* evolved twice, as both an auto- and an allo-hexaploid. Bottom half showing the likely topology recovered by phylogenetic analysis of unphased nuclear consensus sequences if this were the case. Auto tetraploid *I. aequatoriensis* would most likely be recovered in a clade with its autohexaploid descendant (*I. batatas* Lineage 1), and this clade would be nested within Lineage 2 samlpes of hexaploid *I. batatas* as a result of its possession of some alleles inherited from an ancestor shared with diploid *I. trifida*. Likewise, putative hybrids like *I. tabascana* will be recovered in a similar position on the tree, between diploid *I. trifida* and hexaploid *I. batatas*. This resembles the topology recovered in the denovo-assembly based phylogentetic analyses in Chapter 3.

However, if we update this figure to fit the hypothesis of separate auto- and allopolyploid origins for hexaploid *I. batatas* (see Fig. 1 above), we can see that an alternative explanation could be that *I. aequatoriensis* is in fact nested within hexaploid *I. batatas* because some *I. batatas* accessions possess only alleles from the B subgenome, while others possess both A and B subgenome alleles.

Thus, in the same way that putative hybrid *I. tabascana* is recovered between its two presumed parents on the phylogenetic tree, the A subgenome alleles "pull" the putative allohexaploid *I. batatas* samples towards *I. trifida* (Fig. 1). This would then also potentially explain why *I. aequatoriensis* is recovered in a well-supported clade of predominantly chloroplast Lineage 1 *I. batatas* accessions in the results of the denovo-assembly based analyses in Chapter 3.

However, in order to explain the possession of a Lineage 2 chloroplast by Colombian specimen K500, this scenario would also require some geneflow and chloroplast capture between *I. aequatoriensis* and Lineage 2 *I. batatas*. It is also again reliant on paternal inheritance of the chloroplast, no evidence of which was recovered here. Both hypotheses would also mean that the status of the Mexican tetraploids as recent hybrids may require further analysis and revision, since they are also recovered sister to *I. batatas* Lineage 2, rather than nested within it in the chloroplast analysis. One future possibility would be to test chloroplast inheritance in crosses between hexaploid *I. batatas* and diploid *I. trifida*. This could be achieved using the *rpl32-TrnL*(UAG) region as a barcode marker for each parent and the F1 offspring.

In summary none of these hypotheses about the origin of sweet potato's two chloroplast lineages receives conclusive support from the results I have presented here. Ultimately, resolving the origin of the two chloroplast lineages is inextricably tied to understanding the route to hexaploidy in *I. batatas*. Whether they arose

through chloroplast capture, repeated allohexaploid speciation from two different populations of *I. aequatoriensis*, or separate auto- and allohexaploid speciation events, the definitive identification and description of the tetraploid progenitor of *I. batatas* will expedite the process of resolving these important inter-related questions. Given the difficulty of resolving these potentially reticulate relationships, a population genetics approach should be used in future studies, with an increased sampling across the taxa. In particular, more samples of the Colombian and Central American tetraploids should be included, along with herbarium material collected in the same areas. Another major priority is phasing the subgenomes of the different taxa, which may require long-read sequencing.

Taken together, I consider that the results of this thesis demonstrate the vital importance and immense power of taxonomy and natural history collections for contemporary biological research. This conclusion was crystalized for me in the work summarised in Chapter 4, when interrogating the decision to recognise *I. aequatoriensis* as a distinct species. With an increasingly fine-grain understanding of the process of evolution arising from dramatic technological advances, the notion of distinct "species" has itself come under attack in recent years. Partly as a result of this, many biologists are increasingly ambivalent or dismissive about the term, and perhaps also about taxonomy, which is sometimes characterised as an exercise in mere "stamp-collecting" (e.g. Clifford et al. 1990).

In reality, the continued necessity of accurately and consistently delimited species as fundamental units of biological research is testament to their power as a heuristic tool for understanding complex biological phenomena that would otherwise be more or less incomprehensible (Wells et al. 2021, 2022). In this sense, while modern sequencing and bioinformatic technologies undoubtedly represent dramatically improved analytic tools, their power is too often fundamentally compromised by poor taxonomy and a reliance on a small number of representative

samples. For many years, that has been the case for sweet potato, with the resultant impact I outlined in the first half of this thesis.

In resolving that taxonomic confusion and highlighting the diversity of material held in global natural history collections, I believe I have demonstrated what can be achieved when these modern techniques and more “traditional” methods are combined. Well-resolved species-level taxonomy – based on the integrative analysis of representative herbarium and germplasm collections – has the power to transform our understanding of the natural world, at a time when doing so has arguably never been more pressing.

REFERENCES:

Clifford H.T., Rogers R.W., Dettmann M.E. 1990. Where now for taxonomy? Nature. 346:602.

Feng J.Y., Li M., Zhao S., Zhang C., Yang S.T., Qiao S., Tan W.F., Qu H.J., Wang D.Y., Pu Z.G. 2018. Analysis of evolution and genetic diversity of sweetpotato and its related different polyploidy wild species *I. trifida* using RAD-seq. BMC Plant Biol. 18:1–12.

Li M., Yang S., Xu W., Pu Z., Feng J., Wang Z., Zhang C., Peng M., Du C., Lin F., Wei C., Qiao S., Zou H., Zhang L., Li Y., Yang H., Liao A., Song W., Zhang Z., Li J., Wang K., Zhang Y., Lin H., Zhang J., Tan W. 2019. The wild sweetpotato (*Ipomoea trifida*) genome provides insights into storage root development. BMC Plant Biol. 19:1–17.

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., Rausher M.D., Kelly S., Liston A., Scotland R.W., Williams B.R.M., Kronmiller B., Harris S.A., Anglin N.L., Ellis D., Rausher M.D., Kelly S., Liston A., Scotland R.W. 2018. Reconciling Conflicting Phylogenies in the Origin of Sweet Potato and Dispersal to Polynesia. Curr. Biol. 28:1246-1256.e12.

Muñoz-Rodríguez P., Carruthers T., Wood J.R.I., Williams B.R.M., Weitemier K., Kronmiller B., Goodwin Z., Sumadijaya A., Anglin N.L., Filer D., Harris D., Rausher M.D., Kelly S., Liston A., Scotland R.W. 2019. A taxonomic monograph of *Ipomoea* integrated across phylogenetic scales. Nat. Plants. 5:1136–1144.

Pérez-Escobar O.A., Lucas E., Jaramillo C., Monro A., Morris S.K., Bogarín D., Greer D., Dodsworth S., Aguilar-Cano J., Sanchez Meseguer A., Antonelli A. 2019. The Origin and Diversification of the Hyperdiverse Flora in the Chocó Biogeographic Region. Front. Plant Sci. 10:1–9.

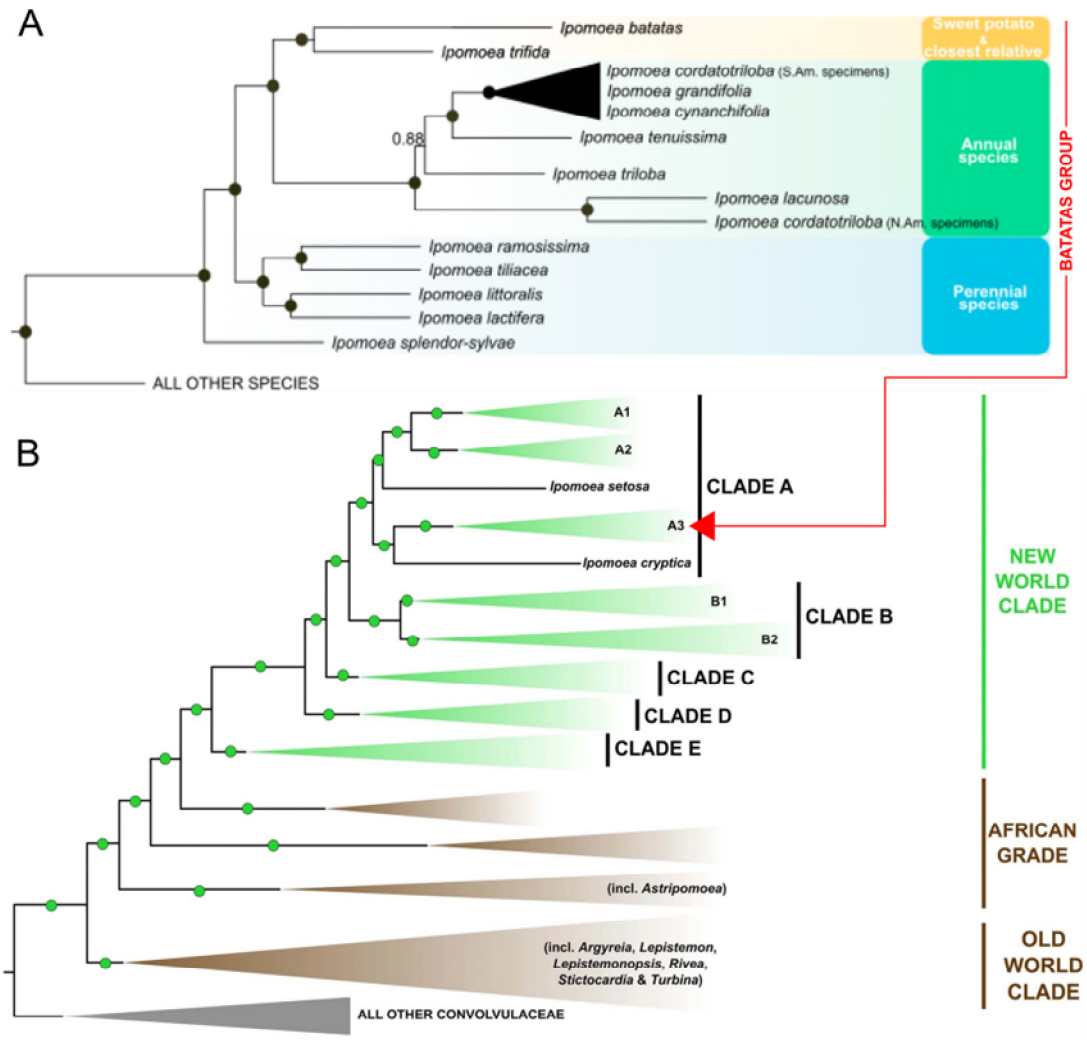
Wells T., Carruthers T., Muñoz-Rodríguez P., Sumadijaya A., Wood J.R.I., Scotland R.W. 2021. Species as a heuristic: reconciling theory and practice. Syst. Biol. 71(5):1233–1243.

Wells T., Carruthers T., Scotland R.W. 2022. Heuristics, species, and the analysis of systematic data. Trends Plant Sci. 27:524–527.

APPENDICES:

APPENDIX 1: Phylogenetic relationships in Batatas Group and wider genus

A taken from Muñoz-Rodríguez et al. (2018) and showing phylogenetic relationships within the Batatas group of sweet potatoes closest CWR. **B** taken from Muñoz-Rodríguez et al. (2019), showing the position of the Batatas group (clade A3) within the wider genus *Ipomoea*. Both phylogenies based on analysis of HybSeq nuclear sequence data.



APPENDIX 2: Morphometric measurements by sample

Morphometric measurements for each specimen included in analyses in Chapter 3. Each sample was given a Unique Identifier composed as follows: Collection Number _ Country ISO code _ Collection Year. Institution codes according to Index Herbariorum (https://sweetgum.nybg.org/science/ih/), with the exception of CIP, NARO, and USDA, which refer to the International Potato Centre (Peru), the National Agriculture and Food Research Organization (Japan), and the United States Department of Agriculture (USA) respectively.

Institutions	Collector	Unique Identifier	species	internode diameter (mm)	internode length (cm)	Lamina length (cm)	Lamina width (cm)	No. lamina lobes	pedicel length (mm)	No. flowers /infl.	outer sepal length (mm)	Inner sepal length (mm)	Sepal apex angle (°)	Corolla length (mm)	Corolla tube width (mm)
K	Hancock, W.	31_MEX_1882	2X trifida	1.13	7	4	3.5	3	10	3	6.89	8.31	40	34.62	6.35
BM; OXF	Miller, O.O.	77_VEN_1901	2X trifida	1.17	6	6	4	3	7.5	4	6.39	9.23	40	47.36	6.92
BM	Burnham, R.	92_MEX_1984	2X trifida	1.52	10	7	5	1	2	2	5.24	9.62	40	30.4	6.62
BM; OXF	Pineda, M.	97_HND_1989	2X trifida	0.95	5	2	1.5	3	5	4	6.13	8.75	40	34.89	5.84
BM; OXF	Lopez, E.	98_CRI_1996	2X trifida	1.38	11	8	7	1	10	6	6.75	9.93	30	18.37	4.14
BM	Wilkin, P.	108_CRI_1989	2X trifida	1.02	4	3	2.5	3	5	2	6.27	10.04	40	37.81	7.59
BM; OXF	Wilkin, P.	109_CRI_1989	2X trifida	1.38	5	4.5	3.5	5	2	2	5.96	10.31	20	38.66	7.55
BM	Wilkin, P.	110_CRI_1989	2X trifida	1.11	5	4	4	3	5	2	7.61	10.8	40	33.66	6
BM	Gutierrez, C.	111_CRI_1984	2X trifida	1.69	8	3.5	2	3	10	4	5.46	7.04	90	29	5.12
K	Palmer, E.	123_MEX_1894	2X trifida	1.19	9	5	5	5	12	3	6.43	9.07	30	35.38	6.59
BM; K	Seeman	175_PAN_SN	2X trifida	1.14	10	4	4.5	3	10	3	4.95	6.35	130	19.56	6.54
BM; OXF	Wilkin, P.	416_CRI_1990	2X trifida	1.26	7	3.5	3	5	6	6	6.93	9.91	30	30.97	8.58
BM; OXF	Wilkin, P.	417_CRI_1990	2X trifida	0.71	4.5	4.5	3	3	7	4	8.24	12.29	30	36.01	7.57
K	Tucker, J.M.	444_SLV_1941	2X trifida	1.21	7	5	4	3	7	4	4.58	7.88	60	30.57	6.44
BM	Wilkin, P.	455_CRI_1990	2X trifida	1.28	7	4.5	4.5	3	7	4	6.99	8.71	50	34.72	6.25
BM; OXF	Wilkin, P.	457_CRI_1990	2X trifida	1.65	10	6.5	5	3	9	8	4.51	8.26	60	23.65	6.35
BM	Wilkin, P.	492_CRI_1990	2X trifida	1.32	5	2.5	2	3	5	2	8.06	10.36	70	37.11	8

BM; OXF	Sidwell, K.	510_SLV_1998	2X trifida	2.44	7	2	1.5	3	17	4	3.73	6.89	40	23.94	5.44
BM; OXF	Palmer, E.	543_MEX_1910	2X trifida	1.35	6	4	4	3	3.5	2	8.71	10.28	30	37.75	4.97
K	Fendler, A.	583_TTO_1881	2X trifida	1.34	7	4.5	3	3	8	6	6.61	8.45	30	22.07	5.15
K	Chavarria, M.	601_CRI_1995	2X trifida	1.12	7	3.5	2.5	1	4	4	7.25	9.83	40	28.85	4.54
K	Langlasse	630_MEX_1898	2X trifida	1.11	6	4	3	3	6	3	6.46	8.85	50	30.84	6.32
BM; OXF	Cutiss, A.H.	632_CUB_1905	2X trifida	1.12	5	5	4.5	3	6	5	5.88	9.52	30	30.18	7.37
K	Hunter, A.A.	723_PAN_1935	2X trifida	1.1	6	2.5	2	3	6	4	4.51	6.45	60	22.6	5.01
K	Palmer, E.	978_MEX_1890	2X trifida	1.23	9	10	10	3	2.5	5	7.01	8.74	40	35.86	5.91
K	Chavarria, M.M.	1027_CRI_1995	2X trifida	1.86	7	5.5	4.5	3	9	4	5.84	7.54	40	21.48	5.3
K	Chavarria, M.M.	1051_CRI_1995	2X trifida	1.39	7	7	6	3	9	4	7.83	8.82	50	35.99	7.05
K	Cascante, A.	1382_CRI_1998	2X trifida	1.39	7	6	4	3	5	4	7.41	10.04	50	39.62	8.51
BM	Espinoza, R.	1406_CRI_1995	2X trifida	1.04	4	2.5	3	5	5	3	6.37	9.28	40	15.26	4.58
BM; OXF	Jimenez, A.	1553_CRI_1964	2X trifida	1.35	9	3.5	3	1	9.5	3	6.76	9.82	40	23.25	5.9
BM; OXF; K; E	Smith, H.H.	1569_COL_1899	2X trifida	1.34	7	6	6	3	18	8	7.76	9.86	40	37.44	7.68
K; BM; OXF	Smith, H.H.	1570_COL_1899	2X trifida	2.05	10	11	10.5	1	22	12	8.9	10.05	NA	40.19	8.76
K	Matuda, E.	2154_MEX_1937	2X trifida	1.29	7	8	8	1	8	6	4.84	6.21	20	26.68	6.56
K	Jeffrey, C.	2351_VEN_1977	2X trifida	1.2	6	7	6	1	17	7	5.48	6.84	60	23.3	4.83
BM; OXF	Pittier, H.	2677_PAN_1911	2X trifida	1	3	3	2.5	5	3	3	5.78	7.68	40	29.54	4.72
K	Abana, B.	3075_CUB_1904	2X trifida	1.74	9.5	7.5	6.5	5	15	5	5.87	9.35	40	29.31	9.52
K	Hinton, G.B.	3434_MEX_1933	2X trifida	1.08	7	3	3	1	5	4	7.62	10.53	40	44.24	7.64
K	Haught, O.	3875_COL_1943	2X trifida	1.72	10	7.5	6.5	1	11	8	6.38	9.17	50	30.11	8.37
BM	Stevens, W.D.	4570_NIC_1977	2X trifida	1.49	7	5.5	4.5	3	10	4	6.79	9.71	20	27.43	6.57
FL; OXF	Austin, D.F.	5060_MEX_1972	2X trifida	1.16	5	5.5	6	3	7	5	8.87	11.28	40	28.2	5.25
BM	Stevens, W.D.	5191_NIC_1977	2X trifida	2.12	10	5	5	3	4	4	5.7	8.28	40	19.72	4.18
BM; OXF	Seymour, F.C.	5436_NIC_1971	2X trifida	1.28	3	3	2	3	3	2	5.39	8.47	40	30.48	6.8
BM; OXF	Ekman, E.L.	6636_CUB_1915	2X trifida	1.87	10.5	7	6	3	9	5	7.23	9.47	30	32.16	6.6
BM	Purpus, C.A.	6918_MEX_1913	2X trifida	1.96	7	4.5	5	1	4	4	6.37	9.23	40	25.26	6.63

K	Mexia, Y.	8865_MEX_1937	2X trifida	1.45	8	6	6	3	10	5	7.67	9.82	50	30.86	7.36
K	Wilson, P.	9281_CUB_1910	2X trifida	1.18	8	5	4	3	10	4	6.8	8.9	60	27.19	7.65
BM	Stevens, W.D.	9428_NIC_1978	2X trifida	1.76	10	5	4	3	9	3	7.46	11.03	50	30.49	7.73
K	Hinton, G.B.	9784_MEX_1936	2X trifida	1.82	7	7	7	1	15	5	7.95	11.67	40	39.75	6.53
BM	Grayum, M.H.	12319_CRI_2004	2X trifida	0.96	6	3	3	3	4	2	7.12	9.33	60	28.28	6.38
BM	Gomez, L.D.	18661_CRI_1982	2X trifida	1.08	11	6.5	5.5	3	14	4	8.24	11.98	50	37.51	7.52
BM	Stevens, W.D.	19709_NIC_1981	2X trifida	1.31	6	8.5	3	1	6.5	4	6.69	8.02	90	21.81	6.71
BM	Hammel, B.	19715_CRI_1994	2X trifida	1.44	10	5	3.5	3	8	6	4.65	8.24	50	18.13	4.49
K	Fendler, A.	2074B_VEN_1856	2X trifida	1.41	8	5	4	3	7	4	6.04	9.08	40	30.63	6.14
K	Sinclair	SN_MEX_1845	2X trifida	0.95	5	2.5	2	5	5	24.86	6.77	8.23	40	28.32	5.66
K	Sinclair	SN_NIC_SN	2X trifida	1.48	4.5	4	3	3	7	3	7.19	10.09	40	30.34	5.48
XAL	Castillo, G.	7997_MEX_1991	2X trifida	1.51	7	5.5	4.5	3	12	4	5.04	11.47	40	29.95	5.51
XAL	Castillo, G.	7144_MEX_1991	2X trifida	1.64	10	7	4.5	3	12	7	4.99	11.35	30	33.31	7.82
XAL	Carranza, E.	6604_MEX_2003	2X trifida	1.41	10	6	5	3	10	5	6.67	9.35	40	44.87	7.81
XAL	Castillo, G.	7228_MEX_1991	2X trifida	1.45	9	5	4	1	12	5	8.17	12.83	40	35.37	6.58
FTG	Neville Jones, G.	23156_MEX_1958	4X_MEX	1.89	6	3	4.5	9	6.8	2	6.91	10.51	100	32.13	10.85
FTG	Austin, D.F.	7480_MEX_1987	4X_MEX	0.78	5	2	2	5	9.38	2	7.28	9.89	120	33.03	9.56
FTG	Orcutt, C.R.	3157_MEX_1910	4X_MEX	1.37	5.5	2	2	7	7.73	3	8.31	12.27	120	42.18	9.24
K	Nogueira, T.D.	295_MEX_1838	4X_MEX	1.31	NA	6	7.5	5	6.71	4	10.85	10.89	120	29.89	7.97
USDA	Austin, D.F.	2908_MEX_1987	4X_MEX	1.09	7	3.5	3.5	3	6.36	2	7.92	10.52	160	40.69	11.76
FTG	Austin, D.F.	7505_MEX_1990	4X_MEX	1.47	4	7	5	3	6.29	3	7.09	7.7	90	30.99	7.49
USDA	Kobayashi	80.3_COL_1976	4X ECU/COL	2.1	9	9	8	3	7.48	8	7.47	9.07	100	41.78	8.14
USDA	Rick, C.M.	71.3_ECU_1972	4X ECU/COL	2.09	8	13	13	1	7.52	22	7.06	8.96	90	36.31	7.61
CIP	De La Puente, F.	1818_COL_1996	4X ECU/COL	1.78	14.5	7.5	6	3	12.3	12	5.62	7.05	110	35.4	7.45
CIP	De La Puente, F.	1843_COL_1987	4X ECU/COL	1.21	3.58	9	6	5	8.29	7	4.42	6.54	110	35.31	5.32
CIP	De La Puente, F.	1874_COL_1986	4X ECU/COL	2.1	8.5	6.5	5.5	3	5.4	NA	4.65	7.24	105	28.05	5.51

QAP	Montedosca, M.	677_ECU_1992	4X ECU/COL	2.85	9	11	9	1	11	12	5.48	8.72	120	26.22	6.38
QAP	Ceron, C.	18775_ECU_1992	4X ECU/COL	2.55	7.5	9	8	3	11	16	6.85	8.49	140	36.21	7.26
QCNE	Bass, M.S.	208_ECU_1994	4X ECU/COL	1.88	12	11	9	1	12	18	6.19	8.73	80	40.75	
QCNE; GUAY	Cornejo, X.	304_ECU_1993	4X ECU/COL	1.69	11	8	7	1	8	18	5.1	7.97	80	24.73	5.15
QCNE;QCA	Kvist, L.P.	40476_ECU_1982	4X ECU/COL	1.73	13	12	8.5	1	11	8	5.93	7.48	80	NA	NA
QCNE;QCA	Madsen, J.E.	64032_ECU_1987	4X ECU/COL	2.18	4	8	6.5	1	11	12	6.65	8.91	90	39.39	9.02
QCNE; QAC	Madsen, J.E.	50113_ECU_1984	4X ECU/COL	3.03	10	9.5	9	1	15	18	5.12	7.46	140	30.85	6.12
GUAY	Madsen, J.E.	64032_ECU_1987	4X ECU/COL	1.98	6.5	8	6	1	6	10	5.1	9.1		34.1	10
GUAY	Cornejo, X.	2283_ECU_1994	4X ECU/COL	1.19	6	7	5.5	1	9.5	8	6.1	9.23		34.89	7.9
GUAY	Cornejo, X.	1867_ECU_1994	4X ECU/COL	1.9	12	10	9	3	10	12	5.79	7.1		35	8.49
QAC	Dodson, C.H.	5112_ECU_1972	4X ECU/COL	1.81	6	8.5	7	1	7	8	5.89	8.2	50	36.19	4.87
QAC	McMahon, M.P.	4207_ECU_1971	4X ECU/COL	1.8	8	8.5	8	1	12	18	6.84	8.5	30	NA	NA
QAC	Besse, L.	2305_ECU_1985	4X ECU/COL	1.86	10	6	6	1	7.5	20	4.75	7.02	90	33.95	3.87
QAC	Balslev, H.	3131_ECU_1982	4X ECU/COL	1.65	10	7	6	1	8	16	5.09	6.73	110	30.18	7.32
QAC	Harling, G.	24997_ECU_1985	4X ECU/COL	2.14	8	9	8.5	1	16	14	6.62	9.15	50	30.46	3.66
QAC	Hoover, W.S.	4129_ECU_1988	4X ECU/COL	2.35	12	13	12	1	17	10	4.95	6.39	90	NA	NA
QAC	Dodson, C.H.	4334_ECU_1972	4X ECU/COL	1.89	12	11	11	3	12	6	6.71	9.42	40	51.44	8.41
ST	Sparre, B.	15341_ECU_1967	4X ECU/COL	1.75	8	7	6	3	6.65	14	5.29	7.42	110	42.03	4.36
ST	Sparre, B.	15517_ECU_1967	4X ECU/COL	1.81	10	10	8	1	5.64	12	4.11	5.507	130	35.61	3.51
AAU;FTG	Holm-Nielsen, L.	25318_ECU_1980	4X ECU/COL	1.21	6	7	6	3	5.16	2	5.79	7.202	110	45.48	5.373
ST	Sparre, B.	15308_ECU_1967	4X ECU/COL	2.32	10	12	10	1	6.4	14	5.62	7.063	90	42.73	4.54

ST	Jativa, C.	1191_ECU_1967	4X ECU/COL	2.65	10	10	8	1	5.31	22	5.38	7.34	100	44.71	5.37
ST	Sparre, B.	15286_ECU_1967	4X ECU/COL	2.22	11	9	7	1	4.8	14	4.72	6.05	120	29.17	3.404
ST	Asplund, E.	16545_ECU_1955	4X ECU/COL	3.25	12	12	11	3	5.36	24	4.39	4.93	140	34.07	2.835
US	Hudson, J.	730_ECU_1974	4X ECU/COL	2.18	10	7	4	1	3.01	18	3.77	4.67	110	30.19	2.81
K; ST	Asplund, E.	15676_ECU_1955	4X ECU/COL	2.53	9	10	9	3	6.4	14	6.94	9.314	120	35.809	8.114
ST	Holmgren, I.	88_ECU_1919	4X ECU/COL	1.02	11	8	7	1	4.05	10	6.24	6.9	130	30.49	5.17
FTG;CIP	Austin, D.F.	7800_ECU_1992	4X ECU/COL	2.505	13	6	5	1	6.71	16	4.99	6.64	100	39.745	3.58
FTG;CIP	De La Puente, F.	5282_ECU_1992	4X ECU/COL	2.21	7	5	3	1	3.2	12	4.41	6.613	110	36.797	2.45
FTG;CIP	De La Puente, F.	5283_ECU_1992	4X ECU/COL	2.02	10	10	8	1	6.41	18	6.05	7.2	150	44.02	2.84
FTG;CIP	De La Puente, F.	5284_ECU_1992	4X ECU/COL	2.6	10	7	6	1	6.44	16	5.022	6.86	130	43.45	2.91
FTG;CIP	Austin, D.F.	7804_ECU_1992	4X ECU/COL	2.01	11	8	7	1	9.62	12	6.78	8.845	90	43.2	4.031
FTG;CIP	Austin, D.F.	7806_ECU_1992	4X ECU/COL	1.54	7	5	4	1	3.54	8	4.78	8.34	120	42.68	4.79
FTG;CIP	Austin, D.F.	7810_ECU_1992	4X ECU/COL	2.83	10	13	11	1	4.87	16	6.23	8.87	90	36.52	4.93
FTG;CIP	Austin, D.F.	7812_ECU_1992	4X ECU/COL	2.12	8	7	6	1	4.75	24	5.49	8.78	110	43.58	4.64
FTG;CIP	Austin, D.F.	7815_ECU_1992	4X ECU/COL	1.758	12	8	7	3	4.1	14	6.66	8.17	120	41	3.57
FTG;CIP	Austin, D.F.	7817_ECU_1992	4X ECU/COL	2.07	9	11	10	1	5.67	12	5.78	8.05	100	38.43	2.82
FTG;CIP	Austin, D.F.	7820_ECU_1992	4X ECU/COL	1.289	7	6	4	3	4.132	8	4.91	7.18	110	44.63	4.57
OXF	Austin, D.F.	78172_ECU_1992	4X ECU/COL	2.11	16	12	11	3	8.16	28	7.76	9.62	100	48.91	5.38
OXF	De La Puente, F.	5293_ECU_1992	4X ECU/COL	2.03	12	9	8	1	6.04	16	5.56	8.42	100	34.26	7.75
FTG;CIP	Austin, D.F.	7809_ECU_1992	4X ECU/COL	1.979	10	14	12	1	8.865	10	8.959	10.915	40	45.672	3.91
FTG;CIP	Austin, D.F.	7811_ECU_1992	4X ECU/COL	1.877	10	11	10	1	4.507	10	8.73	9.99	80	44.104	2.769

FTG;CIP	Austin, D.F.	7816_ECU_1992	4X ECU/COL	1.61	10	9	6	1	7.389	24	7.675	9.635	80	38.096	4.965
ST	Asplund, E.	15966_ECU_1955	4X ECU/COL	1.61	6	9	7.5	3	11	4	11.63	16.16	30	58.13	10.46
K	Grubb, P.	1299_ECU_1960	6X batatas	3.87	7.5	14	14	1	14.34	18	10.78	12.67	50	46.44	9.7
ST	Sparre, B.	19190_ECU_1967	6X batatas	3.45	4	11	10	1	8.41	6	13.06	14.12	60	49.46	10.19
K	Wood, J.R.I.	27928_BOL_2014	6X batatas	2.1	6	8	8	5	7.73	10	10.9	13.72	40	42.83	9.97
K	Beck, St.G.	29746_BOL_2006	6X batatas	2.34	3	8	7	1	NA	NA	NA	NA	NA	NA	NA
K	Balee, W.L.	893_BRA_1985	6X batatas	2.7	6	11	12	5	7.35	9	9.51	9.91	80	32.08	11.85
K	Balee, W.L.	805_BRA_1985	6X batatas	2.22	8	9	8	3	9.91	8	10.24	11.77	50	44.36	12.62
K		1771_BRA_SN	6X batatas	3.74	9	9	10	1	6.74	16	8.89	9.89	60	18.35	8.17
K	Burchell, W.J.	3613_BRA_SN	6X batatas	NA	6	11	11	5	3.82	4	8.58	10.14	110	47.06	9.26
K	Prance, G.T.	18015_BRA_1973	6X batatas	2.25	3.5	6	5	1	4.2	4	6.48	8.74	100	27.87	8.75
K	Lehmann, F.C.	5092_COL_SN	6X batatas	2.64	5.5	7.5	7	3	10.05	8	7.91	10.33	80	34.98	9.54
K	Von Sneidern, K.	1220_COL_1937	6X batatas	3.56	8	12	10	3	6.76	8	9.93	10.78	100	46.19	10.96
K	Lawrence	544_COL_1932	6X batatas	3.23	NA	13	11	1	5.96	12	6.92	9.29	100	34.12	8.76
K	Fuchs, H.P.	21758_COL_1967	6X batatas	2.29	10	8	8	5	6.19	10	7.83	9.21	90	29.63	6.96
BM	Solis, A.	1147_CRI_1994	6X batatas	3.87	8	12	10	5	8.39	10	7.64	8.92	40	38.02	9.49
BM	Dobbeler, P.	1619_CRI_1989	6X batatas	6.14	3	10	9	1	NA	NA	NA	NA	NA	NA	NA
BM	Whitefoord, C.	7017_DMA_1992	6X batatas	4.96	3.5	12	13	7	7.75	10	9.49	12.12	50	43.76	8.78
K	Grubb, P.	1299_ECU_1960	6X batatas	3.87	7.5	14	14	1	14.34	18	10.78	12.67	50	46.44	9.7
ST	Sparre, B.	19190_ECU_1967	6X batatas	3.45	4	11	10	1	8.41	6	13.06	14.12	60	49.46	10.19
K	Clements	230_GUY_1971	6X batatas	2.18	5	9	7	1	11.27	6	7.84	11.69	80	47.86	12.56

BM	House, P.	37_HND_1994	6X batatas	3.84	4	11	10.5	5	9.47	5	10.62	11.81	80	42.64	8.01
BM	Proctor, G.R.	18380_JAM_1958	6X batatas	4.03	2.5	11	11	5	7.56	6	9.65	10.72	140	48.24	10.41
K	Lamb, F.H.	538_MEX_1895	6X batatas	3.38	2	7	7.5	5	5.61	9	8.61	11.51	70	39.31	9.29
K	Hinton, G.B.	2009_MEX_1932	6X batatas	2.49	11	7	7	3	7.85	5	9.95	13.56	60	29.97	6.28
FTG	Austin, D.F.	7672_MEX_1990	6X batatas	2.23	10	8	6	1	4.4	6	8.39	9.87	60	38.56	8.49
BM;E;K	Hassler, E.	11762_PRY_1913	6X batatas	4.35	3.5	8	8	1	7.51	7	8.42	11.02	NA	37.36	10.49
K	Kuntze, E.	2384_PER_SN	6X batatas	3.26	5	12	10	3	9.08	8	9.21	10.46	60	NA	NA
K	Koemar, S.	21_SUR_1999	6X batatas	2.24	10	9	8.5	5	12.43	6	6.9	11.27	80	48.01	11.05
K	Lister	149_VEN_1976	6X batatas	2.86	6	11	10	1	9.87	12	8.84	10.2	160	40.34	8.65
K	Colchester	2003_VEN_1975	6X batatas	2.66	10	8	9	3	10.97	8	9.17	11.01	50	31.39	5.5
K	Colchester	2152_VEN_1976	6X batatas	2.23	2	4	5	3	8.17	6	7.99	10.11	140	41.07	7.81
BM	Alston, A.H.G.	5435_VEN_1938	6X batatas	3.52	3	6	6.5	5	5.18	3	9.75	9.94	40	37.75	8

APPENDIX 3: GENOMESCOPE2.0 KMER & HETEROZYGOSITY STATS

	kmcov	error	Fit	haploid genome length		heterozygosity		aabb		aaab	
sample				min	max	min	max	min	max	min	max
I_apiculata4X_DLP2908_MEX	16	0.24%	**92.4%**	430,679,468	436,329,416	5.52%	27.69%	**5.52%**	**13.49%**	0%	2.85%
I_apiculata4X_K233_MEX	18	0.40%	**91.8%**	403,740,481	408,479,796	8.08%	24.55%	**8.08%**	**11.22%**	0%	1.83%
I_sp4X_DLP2923_MEX	19	0.35%	**92.3%**	388,643,966	392,459,163	7.33%	25.82%	**7.33%**	**12.94%**	0%	2.81%
I_tabascana4X_DLP2946_MEX	18	0.42%	**92.1%**	392,187,428	397,128,348	7.53%	25.26%	**7.53%**	**11.73%**	0%	2.77%
I_sp4X_DLP5284_ECU	19	0.07%	**96.8%**	596,335,479	609,771,573	2.22%	6.56%	0.76%	1.59%	**2.22%**	**2.37%**

APPENDIX 4: SPECIMEN CITATION FOR *I. AEQUATORIENSIS*

Institution codes according to Index Herbariorum (https://sweetgum.nybg.org/science/ih/), with the exception of CIP, NARO, and USDA, which refer to the International Potato Centre (Peru), the National Agriculture and Food Research Organization (Japan), and the United States Department of Agriculture (USA) respectively. Latitude and Longitude given as Decimal Degrees, using WGS84 projection.

Institution	Accession	Barcode	Collectors	Field Number	Collection Year	Country	Locality	Latitude	Longitude
K			Hind	s.n.	1841	Ecuador	Atacamas	0.863996	-79.845532
K			?	78	1854	Ecuador	Huayaguil [Guayaquil]	-2.205842	-79.907948
US	1335946	03152194	Eggers	14483	1892	Ecuador		-2.884705	-79.748862
K			Eggers	15105	1897	Ecuador	Manabi		
P		P03866778	Eggers	15105	1897	Ecuador	Manabi		
ST			Holmgren, I	88	1919	Ecuador	GUAYAS: San Ignacio, On river bank	-2.363781	-79.855901
US	1195411	03152196	Hitchcock, A.S.	20464	1923	Ecuador	Guayas: Teresita, 3km West of Bucay	-2.2	-79.16666667
US	1195838	0315295	Hitchcock, A.S.	20152	1923	Ecuador	Guayaquil, Guayas	-2.156987	-79.89212
F	955481		Mille, L.	1053	1928	Ecuador	Guayaquil	-2.176285	-79.914457
F	1739608		Steyermark, J A	54073	1943	Ecuador	Prov. El Oro: along Rio Amarillo, upstream from Portovelo	-3.703946	-79.598647
F	1536910		Acosta Solis, M.		1945	Ecuador	Entre Slaina y Chone. Prov. Manabi	-0.713679	-80.114501
UNL			Espinosa, R.	1185	1946	Ecuador	Torata (camino a Santa Rosa)	-3.584444444	-79.87833333
F	1543409		Acosta Solis, M.	13897	1949	Ecuador	Km 170-175, via Sto. Domingo - Quininde.	0.290661	-79.453665
S			Fagerlind, F.; G, W.	362	1952	Ecuador	Guayas: Las Americas, in clearing	-1.05	-79.63333306
K			Asplund, E	15676	1955	Ecuador	EL ORO: Arenillas	-3.553987	-80.066816
S			Asplund, E	16545	1955	Ecuador	Esmeraldas: Esmeraldas.	0.966232	-79.654853
ST			Asplund, E	15676	1955	Ecuador	EL ORO: Arenillas	-3.553987	-80.066816
US	2428437	03152202	Gilmartin, A.J.	705	1962	Ecuador	12km Guayaquil to Salina Playa	-2.199591	-80.061351
ST			Jativa, C; Epling, C	182	1962	Ecuador	LOS RIOS: near La Guadalupe	-1.61364	-79.855551

ST			Sparre, B	17916	1967	Ecuador	LOS RIOS: Hacienda Clementia, between Babahoyo and Montalvo, secondary monsoon forest and cultivated land	-1.842658	-79.383367
ST			Sparre, B	17329	1967	Ecuador	COTOPAXI: Rio Guapara, +- 20Km NW of El Corazon Monsoon forest	-0.987551	-79.223392
S			Sparre, B	15286	1967	Ecuador	Esmeraldas: Hacienda Timbre, +- 25km S Esmerladas, cultivated land and secondary forest vegetation	0.740789	-79.592727
S			Sparre, B	15517	1967	Ecuador	Esmeraldas: Hacienda Guayas, +- 20km S Esmeraldas, open secondary vegetation	0.789547	-79.634482
ST			Sparre, B	14810	1967	Ecuador	PICHINCHA: road Aloag-Santo Domingo, Alluriquin, at confluence between Rio Alluriquin and Rio Toachi, tropical rainforest, c. 600m	-0.320455	-78.994754
F			Dodson, C.H.;Williams, N.H.;Adams, R.P.	3811	1967	Ecuador	PINCHINCHA: Along Rio Baba about 10 km south of Santo Domingo	-0.359499	-79.231552
S			Sparre, B	15341	1967	Ecuador	Esmeraldas: Pacific shore, +- 2km N Atacames	0.882704	-79.830429
S			Sparre, B	15308	1967	Ecuador	Esmeraldas: Hacienda Timbre, +- 25km S Esmerladas, cultivated land and secondary forest vegetation	0.737134	-79.596009
S			Jativa, C.; Epling, E.	1191	1967	Ecuador	Esmeraldas: New Deforno Camp, Borbon	1.063716	-78.996535
USDA	PI 355830	SAM 2614	Rick, C M	71.3	1970	Ecuador	Esmeraldas: Quininde	0.327213	-79.434375
OXF			Rick, C M	71.3	1970	Ecuador	Esmeraldas: Quininde	0.327213	-79.434375
OXF		K300	Rick, C M	71.3	1970	Ecuador	Esmeraldas: Quininde	0.327213	-79.434375
NARO		K300	Rick, C M	71.3	1970	Ecuador	Esmeraldas: Quininde	0.327213	-79.434375
OXF			Rick, C M	71.3	1970	Ecuador	Esmeraldas: Quininde	0.327213	-79.434375
DAV	DAV185320		Rick, C M	2639	1970	Ecuador	Ecuador, Province Guayas, Ecuador: Province Guayas. Hills NE of Los Ceibos, suburb of Guayaquil.	-2.175853	-79.948079
QAC			McMahon, M.P.	4207	1971	Ecuador	LOS RIOS: Rio Palenque Biological Station. Km 56 Rd. Quevedo - Sto. Domingo.	-0.583333333	-79.41666667
US	2690994	03152208	McMahon, M.P.	4207	1971	Ecuador	Rio Palenque Biological Station. Km 56 Quevedo to Sto. Domingo. Prov. Los Rios.	-0.565035	-79.375247

US			Dodson, C.H.; McMhon, M.	5112	1972	Ecuador	LOS RIOS: Rio Palenque Biological Station. Km 56 Rd. Quevedo - Sto. Domingo.	-0.588189	-79.363132
QAC			Dodson, C.H.; Vrieze, J.M.	4334	1972	Ecuador	LOS RIOS: Rio Palenque Biological Station. Km 56 Rd. Quevedo - Sto. Domingo. Along edge of road from highway.	-0.589726	-79.365382
QAC			Dodson, C.H.; McMhon, M.	5112	1972	Ecuador	LOS RIOS: Rio Palenque Biological Station. Km 56 Rd. Quevedo - Sto. Domingo.	-0.588189	-79.363132
US			Hudson, J.	730	1974	Ecuador	Esmeraldas: About 1km S of the city of Esmeraldas. Scrubby area.	0.937857	-79.653444
QAC			Dodson, C.H.	5510	1974	Ecuador	LOS RIOS: Rio Palenque Biological Station. Km 56 Rd. Quevedo - Sto. Domingo.	-0.587494	-79.362177
OXF			Rick, C M	81.2	1980	Ecuador	Loja: Cangonomá	-3.963333	-79.701389
USDA	CH81.2	SAM 5316	Rick, C M	81.2	1980	Ecuador	Loja: Cangonomá	-3.963333	-79.701389
OXF			Rick, C M	81.2	1980	Ecuador	Loja: Cangonomá	-3.963333	-79.701389
QCNE			Dodson, C.H.; Gentry, A.	9540	1980	Ecuador	LOS RIOS: Rio Palenque Biological Station. Km 56 Rd. Quevedo - Sto. Domingo.	-0.588985	-79.362514
ARIZ			Holm-Nielsen, L.;Jaramillo, J.;Coello, F.;Asanza, E.	25318	1980	Ecuador	Esmeraldas: Rio Cayapas, Cooperativa la Concordia, half an hour upstream from Borbon. Riverside forest.	1.05	-79
AAU			Holm-Nielsen, L.;Jaramillo, J.;Coello, F.;Asanza, E.	25318	1980	Ecuador	Esmeraldas: Rio Cayapas, Cooperativa la Concordia, half an hour upstream from Borbon. Riverside forest.	1.05	-79
ST			Sparre, B	19970	1981	Ecuador	MANABI: ca 10 Km SW of Jipijapa, near Olia	-1.37789	-80.625438
QCNE			Werling, L.; Leth-Nissen, S.	558	1981	Ecuador	STO. DOMINGO DE LOS COLORADOS: Sto. Domingo-Quevedo road km 35. Haciendda Margarita. Along small river and field edges.	-0.584611	-79.373092
QAC			Werling, L.; Leth-Nissen, S.	558	1981	Ecuador	STO. DOMINGO DE LOS COLORADOS: Sto. Domingo-Quevedo road km 35. Haciendda Margarita. Along small river and field edges.	-0.584611	-79.373092
QCNE			Kvist, L.P.	40476	1982	Ecuador	ESMERALDAS: Rio Cayapa. Zapallo Grande. Farmland close to the village	0.8	-78.9
OXF			Balslev, H.; Steere, W.C.	3131	1982	Ecuador	ESMERALDAS: Carretera Chontaduro-Rio Verde km 3-10. Selva Tropical y pastorales	1	-79.48333333
QAC			Balslev, H.; Steere, W.C.	3131	1982	Ecuador	ESMERALDAS: Carretera Chontaduro-Rio Verde km 3-10. Selva Tropical y pastorales	1	-79.48333333

QCA			Kvist, L.P.	40476	1982	Ecuador	ESMERALDAS: Rio Cayapa. Zapallo Grande. Farmland close to the village	0.8	-78.9
Q			Dodson, C.H.; Embree, A.	13399	1982	Ecuador	COTOPAXI: Tenefuerste. Rio Pilalo, km 52-53 Quevedo-Latacunga	-0.859771	-78.859671
QCNE			Dodson, C.H.; Embree, A.	13399	1982	Ecuador	COTOPAXI: Tenefuerste. Rio Pilalo, km 52-53 Quevedo-Latacunga	-0.859771	-78.859671
QAC			Madsen, J.E.	50113	1984	Ecuador	GUAYAS: Road Salinas - Manglaralto, at crossroad Ayangue. Dry forest dominated by Pilosocereus tweedyanus.	-1.983333333	-80.73333333
QCNE			Madsen, J.E.	50113	1984	Ecuador	GUAYAS: Road Salinas - Manglaralto, at crossroad Ayangue. Dry forest dominated by Pilosocereus tweedyanus.	-1.983333333	-80.73333333
DAV	DAV185318		Rick, C M	2613	1985	Ecuador	Ecuador, Province Esmeraldas, Ecuador: Province Esmeraldas. Several km E of Quinindé.	0.328032	-79.404969
QAC			Harling, G.; Andersson, L.	24997	1985	Ecuador	MANABI: 11km E of San Placido on road Portoviejo-Pinchincha, 400m, dry evergreen forest.	-1.058593	-80.105646
QAC			Harling, G.; Andersson, L.	22636	1985	Ecuador	LOJA: At crest above Macara on road to Loja, 1200m, secondary scrub	-4.366666667	-79.93333333
QAC			Besse, L.; Luther, H.; Halton, J.	2305	1985	Ecuador	ESEMERALDAS: vic. Borbon, 50m, tropical wet forest	1.1	-78.98333333
QCA			Madsen, J.E.	64032	1987	Ecuador	GUAYAS: Isla Puná. Vicintity of Cerro Zambapala. Semi-deciduous forest canopy with Ceiba pentandra and Bombax ruizii and evergreen understorey. Further numerous epiphytes and vines.	-2.966666667	-80.21666667
GUAY			Madsen, J.E.	64032	1987	Ecuador	GUAYAS: Isla Puná. Vicintity of Cerro Zambapala. Semi-deciduous forest canopy with Ceiba pentandra and Bombax ruizii and evergreen understorey. Further numerous epiphytes and vines.	-2.966666667	-80.21666667
QCNE			Madsen, J.E.	64032	1987	Ecuador	GUAYAS: Isla Puná. Vicintity of Cerro Zambapala. Semi-deciduous forest canopy with Ceiba pentandra and Bombax ruizii and evergreen understorey. Further numerous epiphytes and vines.	-2.966666667	-80.21666667

QAC			Hoover, W.S.; Lorentzen, R.A.; Gelpi, P.	4129	1988	Ecuador	ESMERALDAS: Further down Rio Palavi km's camp. Exposed area on sandy bank, 5m from river;	1.116666667	-78.61666667
F			Austin, D F; De La Puente, F	7821/5302	1990	Ecuador	Cangonoma, Loja Province	-3.976859	-79.69301
CIP	DLP5302		Austin, D F; De La Puente, F	7821/5302	1990	Ecuador	Cangonoma, Loja Province	-3.976859	-79.69301
USDA	PI 561260		Austin, D F; De La Puente, F	7821/5302	1990	Ecuador	Cangonoma, Loja Province	-3.976859	-79.69301
QCNE			Rubio, D.; Tipaz, G.; Herzog, W.; Nunez, T.	1838	1991	Ecuador	GUAYAS: Guayaquil Canton. Bosque Protector Cerro Blanco. Carretera a Salinas, km 15. Bosque seco Tropical. Suelo derivado de rocacaliza.	-2.166666667	-79.96666667
F			Austin, D F; De La Puente, F	7820/5301	1992	Ecuador	Machala, El Oro Province	-3.258608	-79.960535
F	DFA7806	PI 561250	Austin, D F; De La Puente, F	7806/5287	1992	Ecuador	Manabi?		
CIP	DLP5301		Austin, D F; De La Puente, F	7820/5301	1992	Ecuador	Machala, El Oro Province	-3.258608	-79.960535
F	DFA7812		Austin, D F; De La Puente, F	7812/5293	1992	Ecuador	Guayaquil Desvio Julio Moreno, Guayaquil, Guayas Province	-2.083333	-80.083333
CIP	DLP5298		Austin, D F; De La Puente, F	7817/5298	1992	Ecuador	El Guabo, El Oro Province	-3.215167	-79.793455
CIP	CIP 403552		Austin, D F; De La Puente, F	7802/5283	1992	Ecuador	Entr A Sua D Esmeralda, Sua, Esmeraldas, Province	0.879394	-79.766944
OXF			Austin, D F; De La Puente, F	7817/5298	1992	Ecuador	El Guabo, El Oro Province	-3.215167	-79.793455
OXF			Austin, D F; De La Puente, F	7802/5283	1992	Ecuador	Entr A Sua D Esmeralda, Sua, Esmeraldas, Province	0.879394	-79.766944
QAP			Cerón, C.E.; Montalvo, C.; Torres, N.	18775	1992	Ecuador	MANABI: Jipijapa. Parroquia: Pto. Lopez. Parque Nacional Machalilla. Isla Salango. Matorral desertico Tropical.	-1.566666667	-80.8
QCNE			Yánez, A.P.; Keel, S.; M, G.; Hernandez, C.; Marchán, N.	1122	1992	Ecuador	MANABI: Parque Nacional Machalilla, Canton Jipijapa, Sector Salaite, bosque seco. 3-6km al Oeste de la Guardiania de Salaite	-1.4	-80.68333333
UNL			Hernandez, C.; Yanez, A.P.; Quezad, N.; Marchan, N.	75	1992	Ecuador	MANABI: Canton Jipijapa. Parroquia Machalilla, camino de San Sebastian hasta el rio Guacharaco. Bosque humedo.	-1.566666667	-80.66666667
USDA	PI 561255 01		Austin, D F; De La Puente, F	7815/5296	1992	Ecuador	Via Julio Moreno, Guayaquil, Guayas Province	-2.06667	-80.0833
F			Austin, D F; De La Puente, F	7817/5298	1992	Ecuador	El Guabo, El Oro Province	-3.215167	-79.793455

OXF			Austin, D F; De La Puente, F	7812/5293	1992	Ecuador	Guayaquil Desvio Julio Moreno, Guayaquil, Guayas Province	-2.083333	-80.083333
OXF			Austin, D F; De La Puente, F	7803/5284	1992	Ecuador	Km 5 Ant D Esmeralda Quininde, Esmeralda, Esmeraldas Province	0.377948	-79.493458
CIP	DLP5285		Austin, D F; De La Puente, F	7804/5285	1992	Ecuador	Km 74.5 Esmeralda-Quininde, Quininde, Esmeraldas Province	-0.5	-79.474226
USDA	PI 561261 01		Austin, D F; De La Puente, F	7823/5304	1992	Ecuador	Puyango, Loja Province	-3.866856	-80.08333
USDA	PI 561246 01		Austin, D F; De La Puente, F	7801/5282	1992	Ecuador	Entr Esmeralda D Rosa Zarate, Esmeralda, Esmeraldas Province	-0.95	-79.666667
OXF			Austin, D F; De La Puente, F	7815/5296	1992	Ecuador	Via Julio Moreno, Guayaquil, Guayas Province	-2.06667	-80.0833
USDA	PI 561254		Austin, D F; De La Puente, F	7812/5293	1992	Ecuador	Guayaquil Desvio Julio Moreno, Guayaquil, Guayas Province	-2.083333	-80.083333
CIP	CIP403553		Austin, D F; De La Puente, F	7803/5284	1992	Ecuador	Km 5 Ant D Esmeralda Quininde, Esmeralda, Esmeraldas Province	0.377948	-79.493458
USDA	PI 561249		Austin, D F; De La Puente, F	7804/5285	1992	Ecuador	Km 74.5 Esmeralda-Quininde, Quininde, Esmeraldas Province	-0.5	-79.474226
F	DFA 7802		Austin, D F; De La Puente, F	7802/5283	1992	Ecuador	Entr A Sua D Esmeralda, Sua, Esmeraldas, Province	0.879394	-79.766944
F	DFA7804		Austin, D F; De La Puente, F	7804/5285	1992	Ecuador	Km 74.5 Esmeralda-Quininde, Quininde, Esmeraldas Province	-0.5	-79.474226
USDA	PI 561248 01		Austin, D F; De La Puente, F	7803/5284	1992	Ecuador	Km 5 Ant D Esmeralda Quininde, Esmeralda, Esmeraldas Province	0.377948	-79.493458
QAC			Yánez, A.P.; Keel, S.; M, G.; Hernandez, C.; Marchán, N.	1122	1992	Ecuador	MANABI: Parque Nacional Machalilla, Canton Jipijapa, Sector Salaite, bosque seco. 3-6km al Oeste de la Guardiania de Salaite	-1.4	-80.68333333
OXF			Austin, D F; De La Puente, F	7823/5304	1992	Ecuador	Puyango, Loja Province	-3.866856	-80.08333
QCNE			Hernandez, C.; Keel, S.; Yanez, A.P.; Gavilanes, M.; Marchan, N.	224	1992	Ecuador	MANABI: Canton Jipijapa. Parroquia Machalilla, Las Goteras Bosque seco perturbado.	-1.55	-80.65
F	PI 561259		Austin, D F; De La Puente, F	7820/5301	1992	Ecuador	Machala, El Oro Province	-3.258608	-79.960535
CIP	CIP 403554		Austin, D F; De La Puente, F	7804/5285	1992	Ecuador	Km 74.5 Esmeralda-Quininde, Quininde, Esmeraldas Province	-0.5	-79.474226
CIP	CIP 403559		Austin, D F; De La Puente, F	7812/5293	1992	Ecuador	Guayaquil Desvio Julio Moreno, Guayaquil, Guayas Province	-2.083333	-80.083333

OXF			Austin, D F; De La Puente, F	7803/5284	1992	Ecuador	Km 5 Ant D Esmeralda Quininde, Esmeralda, Esmeraldas Province	0.377948	-79.493458
F	DFA7804		Austin, D F; De La Puente, F	7803/5284	1992	Ecuador	Km 5 Ant D Esmeralda Quininde, Esmeralda, Esmeraldas Province	0.377948	-79.493458
CIP	CIP 403553		Austin, D F; De La Puente, F	7803/5284	1992	Ecuador	Km 5 Ant D Esmeralda Quininde, Esmeralda, Esmeraldas Province	0.377948	-79.493458
F	DFA 7815		Austin, D F; De La Puente, F	7815/5296	1992	Ecuador	Via Julio Moreno, Guayaquil, Guayas Province	-2.06667	-80.0833
OXF			Austin, D F; De La Puente, F	7801/5282	1992	Ecuador	Entr Esmeralda D Rosa Zarate, Esmeralda, Esmeraldas Province	-0.95	-79.666667
FTG	DFA7801		Austin, D F; De La Puente, F	7801/5282	1992	Ecuador	Entr Esmeralda D Rosa Zarate, Esmeralda, Esmeraldas Province	-0.95	-79.666667
CIP	CIP 403551		Austin, D F; De La Puente, F	7801/5282	1992	Ecuador	Entr Esmeralda D Rosa Zarate, Esmeralda, Esmeraldas Province	-0.95	-79.666667
QAC			Hernandez, C.; Yanez, A.P.; Quezad, N.; Marchan, N.	75	1992	Ecuador	MANABI: Canton Jipijapa. Parroquia Machalilla, camino de San Sebastian hasta el rio Guacharaco. Bosque humedo.	-1.566666667	-80.66666667
CIP	CIP 403552		Austin, D F; De La Puente, F	7802/5283	1992	Ecuador	Entr A Sua D Esmeralda, Sua, Esmeraldas, Province	0.879394	-79.766944
USDA	PI 561258		Austin, D F; De La Puente, F	7817/5298	1992	Ecuador	El Guabo, El Oro Province	-3.215167	-79.793455
F	DFA7810	PI561252?	Austin, D F; De La Puente, F	7810/5291	1992	Ecuador	Bolivar?		
OXF			Austin, D F; De La Puente, F	7812/5293	1992	Ecuador	Guayaquil Desvio Julio Moreno, Guayaquil, Guayas Province	-2.083333	-80.083333
F	DFA7811		Austin, D F; De La Puente, F	7811/5292	1992	Ecuador	Bolivar?		
F	DFA7809	PI561251?	Austin, D F; De La Puente, F	7809/5290	1992	Ecuador	Los Rios		
OXF			Cerón, C.E.; Montalvo, C.; Torres, N.	18775	1992	Ecuador	MANABI: Jipijapa. Parroquia: Pto. Lopez. Parque Nacional Machalilla. Isla Salango. Matorral desertico Tropical.	-1.566666667	-80.8
F	DFA7816		Austin, D F; De La Puente, F	7816/5297	1992	Ecuador	Naranjal, Guayas	-2.720455	-79.659522
OXF			Austin, D F; De La Puente, F	7815/5296	1992	Ecuador	Via Julio Moreno, Guayaquil, Guayas Province	-2.06667	-80.0833
USDA	PI 561247		Austin, D F; De La Puente, F	7802/5283	1992	Ecuador	Entr A Sua D Esmeralda, Sua, Esmeraldas, Province	0.879394	-79.766944
OXF			Austin, D F; De La Puente, F	7802/5283	1992	Ecuador	Entr A Sua D Esmeralda, Sua, Esmeraldas, Province	0.879394	-79.766944

CIP	CIP 403560		Austin, D F; De La Puente, F	7815/5296	1992	Ecuador	Via Julio Moreno, Guayaquil, Guayas Province	-2.06667	-80.0833
GUAY			Cornejo, X.	304	1993	Ecuador	EL ORO: Hcda. Daucay. La Vega Rivera. Bosque humedo Premontano.	-3.483333333	-79.75
QCNE			Cornejo, X.	304	1993	Ecuador	EL ORO: Hcda. Daucay. La Vega Rivera. Bosque humedo Premontano.	-3.483333333	-79.75
GUAY			Cornejo, X; Bonifaz, C	1867	1994	Ecuador	ESMERALDAS: Borbon. Bosque humedo Tropical. Bosque secundario.	1.083333333	-78.96666667
QCNE			Bass, M.S.; Pitman, N.	208	1994	Ecuador	ESMERALDAS: Quininde. Entrance on road to Herrara and Bilsa Biological Station.	0.35	-79.73333333
GUAY			Cornejo, X; Bonifaz, C	2283	1994	Ecuador	GUAYAS: 10 km al N de Montanita. Bosque transicionl seco-humedo Tropical. Bosque secundario.	-1.733333333	-80.75
AAU			Cornejo, X; Bonifaz, C	2283	1994	Ecuador	GUAYAS: 10 km al N de Montanita. Bosque transicionl seco-humedo Tropical. Bosque secundario.	-1.733333333	-80.75
OXF			Stahl, B.;Knudsen, J.T.	1293	1994	Ecuador	LOS RIOS: Hacienda Clementina, ca 38km NE of Babahoyo	-1.65	-79.36666667
UNL			Klitgaard, B.B.; Lozano, P.; Delgado, T.; Aguirre, N.; Aleyandro, H.; Tapia, R.	547	1997	Ecuador	LOJA: Hacienda Banderones, which is 2 hours walk from km5 on road El Limo-Casaderos. Roadside scrub.	-3.9685	-80.18333306
GUAY			Bonitaz, C.; Cornejo, X.	4022	2000	Ecuador	GUAYAS: Canton Pedro J. Montero, km 12 via Puerto Inca Bypass.Est. 11 Bosque seco tropical. Intervenido.	-2.433333333	-79.21666667
GUAY			Bonitaz, C.; Cornejo, X.	4177	2001	Ecuador	GUAYAS: Chongon Km 27 via Guayaquil-Salinas. Bosque seco tropical	-2.183333333	-80
QCNE			Vargas, H.; Defas, W.; Gomez, N.	5032	2005	Ecuador	BOLIVAR: Guaranda. Recorrido entre el recinto Chazo Juan hasta el km 3 del sendero a la Chorrera de Moras. Bosque Humedo Pre-Montano.	-1.393333333	-79.14972222
QCA;OXF	DEO351		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	351	2022	Ecuador	Vía Operativa Refinería del Pacífico, 15,5 km desde Manta	-1.07085	-80.79175
QCA;OXF	PMR24		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	24	2022	Ecuador	Salida de Puerto López a Puerto Cayo, junto a gasolinera PetrolRios	-1.5473	-80.80523

QCA;OXF	PMR28		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	28	2022	Ecuador	Ruta del Spondylus, entre Puerto Cayo y Pueble Nuevo, frente a Restaurante "Islote El sombrerito"	-1.40242	-80.7541
QCA;OXF	DEO366		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.	366	2022	Ecuador	Vía a Cangonamá, 1,6 km después de la entrada, desde carretera E50	-3.9433	-79.666
QCA;OXF	TOWE30		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	30	2022	Ecuador	South side of road to Laguna after turning at km80 on Quininde-Esmeraldas. Growing on Cassava plants beside farmhouse.	0.39925	-79.52993
QCA;OXF	TOWE68		Wells, T.; Muñoz, P.; Espinel, D.	68	2022	Ecuador	Just south of roundabout on road to Machala, in wide ditch.	-3.24708	-79.84563
QCA;OXF	PMR12		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	12	2022	Ecuador	Camino a la salida sur de Pedernales por Ruta del Spondylus	0.0656	-80.0531
QCA;OXF	PMR8		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	8	2022	Ecuador	Carretera de Puerto Quito a Independencia (Esmeraldas), pasado La Abundancia	0.04427	-79.28603
QCA;OXF	TOWE44		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	44	2022	Ecuador	Just off main Quininde-Esmeraldas road at km80, ~1km towards Laguna. By farmhouse.	0.39998	-79.53193
QCA;OXF	PMR15		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	15	2022	Ecuador	Ruta del Spondylus, entre Pedernales y Coaque, dentro del recinto del hotel Safari, junto a la puerta de entrada	0.02915	-80.0789
QCA;OXF	PMR52		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.	52	2022	Ecuador	Carretera Panamericana E50/585, junto a Casatoma	-3.99447	-79.52907
QCA;OXF	PMR38		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Navarro, J.; Wells, T.; Wood, J.R.I.	38	2022	Ecuador	Guayaquil, Reserva Bosque Protector Cerro Blanco	-2.18092	-80.01542
QCA;OXF	TOWE42		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	42	2022	Ecuador	Either side of road to Laguna for about 50m.	0.38332	-79.61117
QCA;OXF	DEO340		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	340	2022	Ecuador	Canton Puerto Quito, via La Independencia-Puerto Quito, via E28, aprox. 2.6 km despues de cruzar el río Blanco	0.04302	-79.3446
QCA;OXF	DEO355		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	355	2022	Ecuador	Parque Nacional Machalilla, cerca de Antenas CTL, entrando por carretera E15, a 210 m de las antenas	-1.54562	-80.77452

QCA;OXF	DEO342		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	342	2022	Ecuador	Cantón Puerto Quito, vía La Independencia-Puerto Quito, vía E28, aprox. 4.2 km después de cruzar el río Blanco	0.04875	-79.33277
QCA;OXF	PMR5		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	5	2022	Ecuador	Carretera de Puerto Quito a Independencia (Esmeraldas), pasado La Abundancia	0.0403	-79.2863
QCA;OXF	PMR17		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	17	2022	Ecuador	Ruta del Spondylus, desde Pedernales a Coaque, pasado hotel Safari	0.0244	-80.08253
QCA;OXF	TOWE39		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	39	2022	Ecuador	On roadside East of Laguna. By entrance to Reserva Forestal y Biologica Rancho Raquetita and La Cascada del Amor.	0.38867	-79.62262
QCA;OXF	PMR47		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.	47	2022	Ecuador	Carretera Transversal Sur, entre Saracay y Balsas, pasado Siniestro José Valdivieso	-3.71703	-79.849
QCA;OXF	PMR45		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.	45	2022	Ecuador	Carretera Transversal Sur, entre Saracay y Balsas, pasado Siniestro José Valdivieso	-3.71603	-79.85062
QCA;OXF	TOWE64		Wells, T.; Muñoz, P.; Espinel, D.	64	2022	Ecuador	Manglare Churute Reserve, by Car Park. Next to and over broken down truck.	-2.45332	-79.62765
QCA;OXF	TOWE58		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	58	2022	Ecuador	On road to Huigra from Milagro.	-2.32603	-79.0655
QCA;OXF	PMR10		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	10	2022	Ecuador	Finca abandonada junto a carretera de Pedernales a Cojimies, frente a la costa	0.07982	-80.05403
QCA;OXF	DEO367		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.	367	2022	Ecuador	Vía a Cangonamá, 2,15 km después de la entrada, desde carretera E50	-3.9442	-79.6706
QCA;OXF	PMR42		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.	42	2022	Ecuador	Carretera de Machala a Loja, a la salida de Machala	-3.32047	-79.89408
QCA;OXF	TOWE66		Wells, T.; Muñoz, P.; Espinel, D.	66	2022	Ecuador	Roadside south of Naranjal.	-2.88008	-79.70452
QCA;OXF	DEO336		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	336	2022	Ecuador	Cantón Quinindé, vía Quinindé-Esmeraldas, vía E20, aprox. 2 km después de cruzar el río Quinindé	0.34678	-79.47555
QCA;OXF	PMR34		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	34	2022	Ecuador	Carretera de Puerto Cayo a Jipijapa	-1.34262	-80.65688

QCA;OXF	DEO360		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	360	2022	Ecuador	Vía El Triunfo-Huigra, carretera E47, pasando ciudadela San Isidro, 33,25 km después de El Triunfo	-2.27037	-79.13048
QCA;OXF	PMR4		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	4	2022	Ecuador	Carretera de Puerto Quito a Independencia (Esmeraldas), pasado La Abundancia	0.04448	-79.28577
QCA;OXF	DEO338		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	338	2022	Ecuador	Cantón Quinindé, vía Quinindé-Esmeraldas, vía E20, aprox. 2 km después de cruzar el río Quinindé	0.34698	-79.47547
QCA;OXF	DEO361		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.	361	2022	Ecuador	Vía El Triunfo-Huigra, carretera E47, pasando ciudadela San Isidro, 41,7 km después de El Triunfo	-2.29228	-79.0972
QCA;OXF	DEO352		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	352	2022	Ecuador	Parque Nacional Machalilla, vía a la playa los Frailes, a 1 km desde la caseta de ingreso	-1.49302	-80.7801
QCA;OXF	TOWE52		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	52	2022	Ecuador	Puerto Cayo. Nextdoor to Hotel Poseidon.	-1.35632	-80.73752
QCA;OXF	TOWE36		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	36	2022	Ecuador	By roadside East of Laguna.	0.3833	-79.63795
QCA;OXF	TOWE31		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	31	2022	Ecuador	North side of road to Laguna after turning at km80 on Quininde-Esmeraldas. Opposite farmhouse.	0.39927	-79.52992
QCA;OXF	PMR54		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.	54	2022	Ecuador	A la salida de Velacruz hacia Catacocha	-3.9805	-79.58047
QCA;OXF	DEO369		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.	369	2022	Ecuador	Vía a Cangonamá, 7,4 km después de la entrada, desde carretera E50, en Cruz del Agua Fría	-3.94458	-79.70118
QCA;OXF	TOWE37		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	37	2022	Ecuador	By roadside East of Laguna.	0.3833	-79.63795
QCA;OXF	DEO354		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	354	2022	Ecuador	Parque Nacional Machalilla, cerca de Antenas CTL, entrando por carretera E15, a 150 m de las antenas	-1.546	-80.7742

QCA;OXF	DEO343		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	343	2022	Ecuador	Cantón Puerto Quito, vía La Independencia-Puerto Quito, vía E28, aprox. 4.3 km después de cruzar el río Blanco	0.04893	-79.33203
QCA;OXF	DEO337		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	337	2022	Ecuador	Cantón Quinindé, vía Quinindé-Esmeraldas, vía E20, aprox. 2 km después de cruzar el río Quinindé	0.34678	-79.47555
QCA;OXF	PMR55		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.	55	2022	Ecuador	Camino a Naranjo Panto, cerca del desvío desde la carretera Troncal de la Sierra (E35)	-4.02873	-79.61833
QCA;OXF	DEO365		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.	365	2022	Ecuador	Vía a Cangonamá, 500 m después de la entrada, desde carretera E50	-3.9456	-79.65817
QCA;OXF	PMR49		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.	49	2022	Ecuador	Carretera a la salida de Balsas hacia Loja	-3.77312	-79.82568
QCA;OXF	TOWE29		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	29	2022	Ecuador	South side of road past Puerto Quito, West towards Quininde. On bend past large layby.	0.10767	-79.08263
QCA;OXF	PMR9		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	9	2022	Ecuador	Cupa, al norte de Quinindé, en la intersección entre carretera principal y Vía 5 de Agosto	0.37145	-79.48547
QCA;OXF	DEO348		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	348	2022	Ecuador	Vía Canoa-Boca de Briceño, carretera E15, 300 m después de la rotonda	-0.48275	-80.4507
QCA;OXF	TOWE50		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	50	2022	Ecuador	Machalilla National Park. Agua Blanca Rserve. Track to Antenna.	-1.54533	-80.77475
QCA;OXF	TOWE38		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	38	2022	Ecuador	By roadside East of Laguna.	0.38352	-79.63775
QCA;OXF	TOWE27		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	27	2022	Ecuador	South side of road past Puerto Quito, West towards Quininde. On bend opposite parking area/layby	0.108	-79.25027
QCA;OXF	PMR20		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	20	2022	Ecuador	Vía a Camarones desde carretera Troncal del Pacífico, cerca de la Reserva Bosque Seco Lalo Loor	-0.08997	-80.15903
QCA;OXF	TOWE25		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	25	2022	Ecuador	Roadside East of Puerto Quito on road from Pedro Vicente Maldonaldo.	0.12182	-79.22733

QCA;OXF	TOWE65		Wells, T.; Muñoz, P.; Espinel, D.	65	2022	Ecuador	Roadside south of Naranjal.	-2.88008	-79.70452
QCA;OXF	DEO335		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	335	2022	Ecuador	Cantón Quinindé, vía Quinindé-Esmeraldas, vía E20, aprox. 2 km después de cruzar el río Quinindé	0.3471	-79.4754
QCA;OXF	TOWE59		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	59	2022	Ecuador	On road to Huigra from Milagro.	-2.336	-79.0425
QCA;OXF	DEO349		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	349	2022	Ecuador	Vía Bahía de Caraquez-San Jacinto, carretera E15, aprox. 10,5 km después de salir de la ciudad	-0.7188	-80.4171
QCA;OXF	DEO344		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	344	2022	Ecuador	Vía Pedernales-Jama, 8,9 km después de la entrada a Lalo Loor	-0.1216	-80.21305
QCA;OXF	TOWE34		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	34	2022	Ecuador	North side of road to Laguna after turning at km80 on Quininde-Esmeraldas. Opposite farmhouse.	0.3994	-79.53025
QCA;OXF	PMR48		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.	48	2022	Ecuador	Entrada a sitio Las Acacias, borde de carretera	-3.75035	-79.82987
QCA;OXF	TOWE67		Wells, T.; Muñoz, P.; Espinel, D.	67	2022	Ecuador	Roadside opposite Broken Rock Night Club, near Camillo Ponce Enriquez.	-3.03615	-79.7415
QCA;OXF	PMR32		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Ventura, L.; Wells, T.; Wood, J.R.I.	32	2022	Ecuador	Carretera principal entre Parque Nacional Machalilla y Puerto López. Camino de subida a las antenas desde carretera principal	-1.5445	-80.7761
QCA;OXF	PMR14		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	14	2022	Ecuador	Junto a camino a la salida sur de Pedernales por Ruta del Spondylus	0.05097	-80.06
QCA;OXF	PMR1		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	1	2022	Ecuador	Carretera de Puerto Quito a Independencia (Esmeraldas), pasado La Abundancia	0.04425	-79.28602
QCA;OXF	PMR2		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	2	2022	Ecuador	Carretera de Puerto Quito a Independencia (Esmeraldas), pasado La Abundancia	0.04427	-79.28602
QCA;OXF	PMR13		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	13	2022	Ecuador	Junto a camino a la salida sur de Pedernales por Ruta del Spondylus	0.05077	-80.06005

QCA;OXF	TOWE23		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	23	2022	Ecuador	North side of road West of La Celica. On frence at entrance to house.	0.19272	-79.09717
QCA;OXF	DEO363		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.	363	2022	Ecuador	Vía Olmedo-Loja, a 100 m de la "Y" de Olmedo, cerca del momumento	-3.94298	-79.6344
QCA;OXF	TOWE70		Wells, T.; Muñoz, P.; Espinel, D.	70	2022	Ecuador	11km north of Machala.	-3.24708	-79.84563
QCA;OXF	TOWE32		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	32	2022	Ecuador	North side of road to Laguna after turning at km80 on Quininde-Esmeraldas. Opposite farmhouse.	0.3993	-79.52997
QCA;OXF	TOWE24		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	24	2022	Ecuador	Roadside East of Puerto Quito on road from Pedro Vicente Maldonaldo.	0.03787	-79.22717
QCA;OXF	PMR41		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Navarro, J.; Wells, T.; Wood, J.R.I.	41	2022	Ecuador	Guayaquil, Reserva Bosque Protector Cerro Blanco	-2.18138	-80.01682
QCA;OXF	DEO346		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	346	2022	Ecuador	Vía Jama-Canoa, 2,8 km después de Jama, camino a Canoa	-0.229	-80.27568
QCA;OXF	PMR31		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Ventura, L.; Wells, T.; Wood, J.R.I.	31	2022	Ecuador	Carretera principal entre Parque Nacional Machalilla y Puerto López. Camino de subida a las antenas desde carretera principal	-1.5445	-80.77612
QCA;OXF	DEO358		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	358	2022	Ecuador	Vía El Triunfo-Huigra, carretera E47, pasando ciudadela San Isidro, 19,5 km después de El Triunfo	-2.33552	-79.22087
QCA;OXF	PMR44		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.	44	2022	Ecuador	Carretera Transversal Sur, entre Saracay y Balsas, pasado Siniestro José Valdivieso	-3.71637	-79.85107
QCA;OXF	PMR35		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	35	2022	Ecuador	Carretera de Puerto Cayo a Jipijapa	-1.34462	-80.65647
QCA;OXF	TOWE43		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	43	2022	Ecuador	Either side of road to Laguna for about 50m.	0.38322	-79.61162
QCA;OXF	PMR26		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	26	2022	Ecuador	Junto a carretera dentro del Parque Nacional Machalilla	-1.45528	-80.75975

QCA;OXF	TOWE57		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	57	2022	Ecuador	On road to Huigra from Milagro.	-2.32633	-79.06562
QCA;OXF	PMR27		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	27	2022	Ecuador	Ruta del Spondylus, entre Puerto Cayo y Machalilla, junto a Pueble Nuevo	-1.4214	-80.74503
QCA;OXF	DEO368		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.	368	2022	Ecuador	Vía a Cangonamá, 5,2 km después de la entrada, desde carretera E50	-3.94037	-79.68693
QCA;OXF	DEO362		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.	362	2022	Ecuador	Vía Olmedo-Loja, a 100 m de la "Y" de Olmedo, cerca del momumento	-3.94298	-79.6344
QCA;OXF	PMR21		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	21	2022	Ecuador	Camino de tierra de Camarones a Reserva Jama Coaque	-0.11148	-80.13755
QCA;OXF	TOWE22		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	22	2022	Ecuador	North side of road West of La Celica	0.19247	-79.09788
QCA;OXF	DEO353		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.; Ventura-Albán, L.	353	2022	Ecuador	Parque Nacional Machalilla, vía Puerto López-Machalilla, 500 m después de la entrada a Comuna Agua Blanca	-1.51437	-80.77422
QCA;OXF	TOWE26		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	26	2022	Ecuador	Roadside East of Puerto Quito on road from Pedro Vicente Maldonaldo.	0.1212	-79.22717
QCA;OXF	TOWE33		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	33	2022	Ecuador	North side of road to Laguna after turning at km80 on Quininde-Esmeraldas. Opposite farmhouse.	0.39935	-79.53012
QCA;OXF	DEO347		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	347	2022	Ecuador	Vía Canoa-Boca de Briceño, carretera E15, 300 m después de la rotonda	-0.48275	-80.4507
QCA;OXF	PMR6		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	6	2022	Ecuador	Carretera de Puerto Quito a Independencia (Esmeraldas), pasado La Abundancia	0.04392	-79.28645
QCA;OXF	PMR3		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	3	2022	Ecuador	Carretera de Puerto Quito a Independencia (Esmeraldas), pasado La Abundancia	0.04432	-79.28593

QCA;OXF	PMR33		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	33	2022	Ecuador	Carretera de Puerto Cayo a Jipijapa	-1.34352	-80.65662
QCA;OXF	TOWE40		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	40	2022	Ecuador	On roadside East of Laguna. By entrance to Reserva Forestal y Biologica Rancho Raquetita and La Cascada del Amor.	0.38858	-79.62238
QCA;OXF	TOWE51		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	51	2022	Ecuador	Machalilla National Park. Agua Blanca Rserve. Track to Antenna. ~ 500m from top.	-1.54433	-80.77497
QCA;OXF	PMR46		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.	46	2022	Ecuador	Carretera Transversal Sur, entre Saracay y Balsas, pasado Siniestro José Valdivieso	-3.71603	-79.85062
QCA;OXF	TOWE28		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	28	2022	Ecuador	South side of road past Puerto Quito, West towards Quininde. Opposite large parking area/layby	0.108	-79.24983
QCA;OXF	TOWE60		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	60	2022	Ecuador	On road to Huigra from Milagro.	-2.336	-79.0425
QCA;OXF	PMR7		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	7	2022	Ecuador	Carretera de Puerto Quito a Independencia (Esmeraldas), pasado La Abundancia	0.04365	-79.28677
QCA;OXF	DEO341		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	341	2022	Ecuador	Cantón Puerto Quito, vía La Independencia-Puerto Quito, vía E28, aprox. 2.5 km después de cruzar el río Blanco	0.04308	-79.34537
QCA;OXF	DEO356		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	356	2022	Ecuador	Parque Nacional Machalilla, cerca de Antenas CTL, entrando por carretera E15, a 230 m de las antenas	-1.54545	-80.77468
QCA;OXF	DEO345		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	345	2022	Ecuador	Vía Jama-Canoa, 2,8 km después de Jama, camino a Canoa	-0.22922	-80.27585
QCA;OXF	TOWE45		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	45	2022	Ecuador	On roadside , main road to Pedernales, just over boarder into Manabi.	-0.07137	-79.58365
QCA;OXF	PMR19		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	19	2022	Ecuador	Vía a Camarones desde carretera Troncal del Pacífico, cerca de la Reserva Bosque Seco Lalo Loor	-0.0892	-80.15937

QCA;OXF	TOWE46		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	46	2022	Ecuador	On side of road to Pedernales. Just over boarder into Manabi.	-0.07138	-79.58372
QCA;OXF	PMR29		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	29	2022	Ecuador	Ruta del Spondylus, entre Puerto Cayo y Pueble Nuevo, frente a Restaurante "Islote El sombrerito"	-1.4033	-80.75375
QCA;OXF	DEO339		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	339	2022	Ecuador	Cantón Quinindé, vía Quinindé-La Independencia, vía E20, aprox. 12.5 km después de cruzar el río Quinindé	0.24063	-79.42003
QCA;OXF	PMR39		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Navarro, J.; Wells, T.; Wood, J.R.I.	39	2022	Ecuador	Guayaquil, Reserva Bosque Protector Cerro Blanco	-2.18117	-80.01612
QCA;OXF	TOWE62		Wells, T.; Muñoz, P.; Espinel, D.	62	2022	Ecuador	Manglare Churute Reserve, by Car Park.	-2.45393	-79.62783
QCA;OXF	PMR25		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	25	2022	Ecuador	A la salida de Puerto López, dentro del Parque Nacional Machalilla	-1.53588	-80.7965
QCA;OXF	DEO357		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	357	2022	Ecuador	Vía El Triunfo-Huigra, carretera E47, pasando ciudadela San Isidro, 19,5 km después de El Triunfo	-2.33538	-79.22053
QCA;OXF	PMR18		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	18	2022	Ecuador	Vía a Camarones desde carretera Troncal del Pacífico, cerca de la Reserva Bosque Seco Lalo Loor	-0.08943	-80.15925
QCA;OXF	TOWE41		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	41	2022	Ecuador	Either side of road to Laguna for about 50m.	0.38332	-79.61128
QCA;OXF	DEO359		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	359	2022	Ecuador	Vía El Triunfo-Huigra, carretera E47, pasando ciudadela San Isidro, 22,4 km después de El Triunfo	-2.3241	-79.19715
QCA;OXF	PMR53		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.	53	2022	Ecuador	A la salida de Velacruz hacia Catacocha	-3.9805	-79.58047
QCA;OXF	PMR11		Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wells, T.; Wood, J.R.I.	11	2022	Ecuador	Camino a la salida sur de Pedernales por Ruta del Spondylus	0.06512	-80.05278
QCA;OXF	DEO350		Espinel-Ortiz, D.; Muñoz-Rodríguez, P.; Wells, T.; Wood, J.R.I.	350	2022	Ecuador	Vía Operativa Refinería del Pacífico, 15,5 km desde Manta	-1.07085	-80.79175

QCA;OXF	TOWE56		Wells, T.; Muñoz-Rodríguez, P.; Espinel-Ortiz, D.; Wood, J.R.I.	56	2022	Ecuador	On road to Huigra from Milagro.	-2.30672	-79.07523
K			Ruiz; Pavon			Ecuador	Guayaquil	-2.159625	-79.911756
OXF			Ruiz; Pavon			Ecuador	Guayaquil	-2.159625	-79.911756
K			Holton, I F	540	1853	Colombia	La Paila	4.319133	-76.07278
K			Andre, E F	2604	1876	Colombia	Cartago	4.750615	-75.912887
K			Hugh-Jones, D L	25	1962	Colombia	Calima Valley. NW of Restrepo Valle. Lowlying grassland near river	3.916667	-76.666667
OXF			Martin, F. .W.;Jones, A.	73.1	1972	Colombia	Palmira	3.539444	-76.303611
OXF			Martin, F. .W.;Jones, A.	73.1	1972	Colombia	Cali	3.437222	-76.5225
OXF			Martin, F. .W.;Jones, A.	73.1	1972	Colombia	Palmira	3.539444	-76.303611
USDA			Martin, F. .W.;Jones, A.	73.1	1972	Colombia	Cali	3.437222	-76.5225
USDA			Martin, F. .W.;Jones, A.	73.1	1972	Colombia	Palmira	3.539444	-76.303611
OXF			Martin, F. .W.;Jones, A.	73.1	1972	Colombia	Cali	3.437222	-76.5225
CIP	CIP 460042	5682965	De La Puente, F	1859	1986	Colombia	0.5 N Rio Verde	4.4045	-75.7666
CIP			De La Puente, F	1843	1986	Colombia	1 Km E Cali	3.425934	-76.444852
CIP	CIP 403261		De La Puente, F	1842	1986	Colombia	0.7 Km N Montenegro	4.33	-75.61
OXF			De La Puente, F	1818	1986	Colombia	20 Km Ne Buga-buenaventura	4.0269	-76.1731
OXF			De La Puente, F	1874	1986	Colombia	1 Km SE Circacia-calarca	4.6102	-75.6436
CIP	CIP 403248		De La Puente, F	1818	1986	Colombia	20 Km Ne Buga-buenaventura	4.0269	-76.1731
CIP	CIP 403261		De La Puente, F	1843	1986	Colombia	1 Km E Cali	3.425934	-76.444852
CIP			De La Puente, F	1874	1986	Colombia	1 Km SE Circacia-calarca	4.6102	-75.6436
CIP	CIP 403270		De La Puente, F	1874	1986	Colombia	1 Km SE Circacia-calarca	4.6102	-75.6436
CIP			De La Puente, F	1818	1986	Colombia	20 Km Ne Buga-buenaventura	4.0269	-76.1731
FTG			Gentry, A.;Keating, P.	59527	1987	Colombia	VALLE: Bosque de Ema, ca. 60km N of Cali. Cordillera Occidental, east slope.	3.633333333	-76.55
OXF			Alvarez, Y.H.;Farrar, J.;Sells, S.	5	2018	Colombia	Risaralda, Pereira, Otun Quimbaya Sanctuary for Flora and Fauna: located on roadside 100m E from the entrance to Otun Quimbaya	4.728888889	-75.57805556

E			Alvarez, Y.H.;Farrar, J.;Sells, S.	5	2018	Colombia	Risaralda, Pereira, Otun Quimbaya Sanctuary for Flora and Fauna: located on roadside 100m E from the entrance to Otun Quimbaya	4.728888889	-75.57805556
K			Lehmann, F C	5519		Colombia	highlands of Popayan	2.5	-76.583333
OXF			Kobayashi, Y	80.3		Colombia	Cali	3.416667	-76.55
USDA	CH80.3		Kobayashi, Y	80.3		Colombia	Cali	3.416667	-76.55
NARO		K500	Kobayashi, Y	80.3		Colombia	Cali	3.416667	-76.55
K			Andre, E	468		Colombia	Buga	3.900889	-76.297829