

# A taxonomic monograph of *Ipomoea* integrated across phylogenetic scales

Pablo Muñoz-Rodríguez<sup>1,9</sup>, Tom Carruthers<sup>1,9</sup>, John R. I. Wood<sup>1,2</sup>, Bethany R. M. Williams<sup>1</sup>, Kevin Weitemier<sup>1,3</sup>, Brent Kronmiller<sup>4</sup>, Zoë Goodwin<sup>5</sup>, Alex Sumadijaya<sup>1</sup>, Noelle L. Anglin<sup>6</sup>, Denis Filer<sup>1</sup>, David Harris<sup>1,5</sup>, Mark D. Rausher<sup>1,7</sup>, Steven Kelly<sup>1,8</sup>, Aaron Liston<sup>8</sup> and Robert W. Scotland<sup>1\*</sup>

**Taxonomic monographs have the potential to make a unique contribution to the understanding of global biodiversity. However, such studies, now rare, are often considered too daunting to undertake within a realistic time frame, especially as the world's collections have doubled in size in recent times. Here, we report a global-scale monographic study of morning glories (*Ipomoea*) that integrated DNA barcodes and high-throughput sequencing with the morphological study of herbarium specimens. Our approach overhauled the taxonomy of this megadiverse group, described 63 new species and uncovered significant increases in net diversification rates comparable to the most iconic evolutionary radiations in the plant kingdom. Finally, we show that more than 60 species of *Ipomoea*, including sweet potato, independently evolved storage roots in pre-human times, indicating that the storage root is not solely a product of human domestication but a trait that predisposed the species for cultivation. This study demonstrates how the world's natural history collections can contribute to global challenges in the Anthropocene.**

When Joseph Banks and Daniel Solander travelled with Captain Cook on the *Endeavour* in 1768, the plants they collected were new species to science<sup>1</sup>. Similarly, when Robert Brown sailed to Australia in 1801, he too discovered and described a completely new flora with many new species<sup>2</sup>. More than 200 years later, however, the task of deciding whether a specimen represents a new species has become much more difficult because taxonomists need to work through the large number of specimens held in natural history collections, a number that has doubled since 1960<sup>3</sup>, as well as a massive accumulation of literature. The provisional nature of species curation adds to these difficulties, reflecting the fact that species-level taxonomy is incomplete and unsatisfactory for many taxa, especially insects and tropical plants<sup>3</sup>. These difficulties come at a time when improved taxonomic knowledge is an urgent priority for policymakers<sup>4</sup>, environmental scientists<sup>5</sup> and museum directors<sup>6</sup> throughout the world. The Global Strategy for Plant Conservation, for example, seeks to assess the conservation status of all plant species by 2020, but at present less than 25% of plant species have been assessed<sup>7</sup>, largely because of incomplete taxonomic information<sup>8</sup>. Many suggestions have been made to enhance the accuracy, speed, accessibility and relevance of taxonomy<sup>5–14</sup>; but, nevertheless, the pace of flowering plant taxonomy has remained unchanged for the last 30 years<sup>15</sup>. Finding ways to address these substantial issues in a realistic time frame is a recurring challenge<sup>4</sup>.

Much existing taxonomy is inaccurate because it is essentially country- or region-based and inevitably depends on limited specimen sampling<sup>16</sup>. The choice of a particular geographical area to document species is a pragmatic decision and reflects national priorities and funding constraints as well as the interests of policymakers and taxonomists who are focused on the plants and animals of

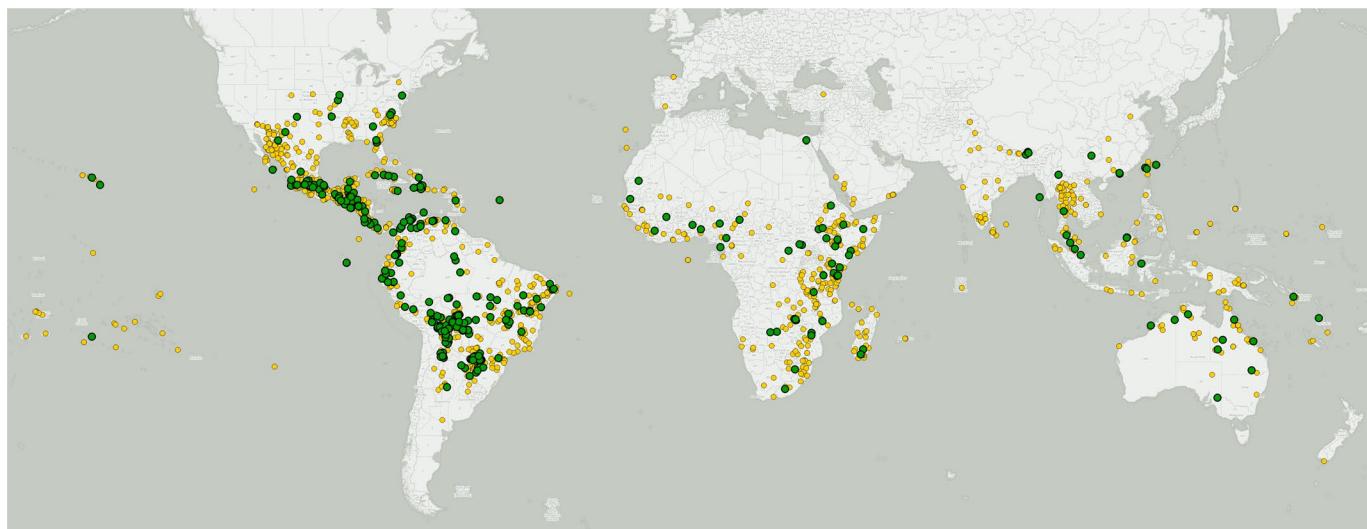
their region. However, species are often widely distributed with the result that the same species may be described on multiple occasions from different countries under different names (synonymy). Over time, issues of synonymy, when combined with misidentification and poor species-level sampling<sup>3,10</sup>, result in many tropical plants being so poorly known that they are invisible to modern ecological and conservation tools<sup>8</sup>. Furthermore, when existing taxonomy is so provisional, determining whether potential new species are different from existing species is highly problematic with the consequence that half of the world's natural history collections are incorrectly named<sup>3</sup>. An urgent priority is, therefore, to tackle the taxonomy of tropical plants from a global perspective.

DNA taxonomy was proposed 15 years ago as an alternative to morphology-based taxonomy<sup>17,18</sup>, which was dismissed as slow and over-reliant on a dwindling number of experts<sup>9</sup>. Since then, DNA has played an increasingly important role in phylogeny reconstruction and higher-level classifications of major lineages<sup>19,20</sup>, as well as in identification of existing species<sup>21,22</sup>, but it is being used only in an auxiliary capacity<sup>18</sup>, if at all, for taxonomic revisions and monographs. Studies integrating DNA and morphology are few and tend to avoid species-rich tropical groups where the greatest taxonomic problems lie<sup>7</sup>. Furthermore, there is no consensus on how DNA sequence data can be best used to solve taxonomic problems at the species level.

This paper describes the integration of molecular phylogenetics with the morphological study of living plants and herbarium collections to produce a taxonomic study of the megadiverse genus *Ipomoea* L. (Convolvulaceae)—with an emphasis on the 423 species described from the American continent. In parallel to the morphological study of herbarium specimens from 72 European and American institutions, we sequenced DNA from 1,560 of those

<sup>1</sup>Department of Plant Sciences, University of Oxford, Oxford, UK. <sup>2</sup>Royal Botanic Gardens, Kew, Richmond, UK. <sup>3</sup>Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR, USA. <sup>4</sup>Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA. <sup>5</sup>Royal Botanic Garden Edinburgh, Edinburgh, UK. <sup>6</sup>International Potato Center, Lima, Peru. <sup>7</sup>French Family Science Center, Duke University, Durham, NC, USA.

<sup>8</sup>Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA. <sup>9</sup>These authors contributed equally: Pablo Muñoz-Rodríguez, Tom Carruthers. \*e-mail: [robert.scotland@plants.ox.ac.uk](mailto:robert.scotland@plants.ox.ac.uk)



**Fig. 1 | Natural history collections facilitate biodiversity studies at a global scale.** A map showing where the 1,560 herbarium specimens sequenced during our study of *Ipomoea* were collected. The yellow dots indicate the collection locality of specimens sequenced for DNA barcoding; the green dots indicate the subset of specimens that were also sequenced using Hyb-Seq to obtain genomic-scale data. Basemap from OpenStreetMap.

specimens for several DNA barcodes. We also sequenced a subset of 384 samples, representing 211 species, for the whole chloroplast genome and 605 putative single-copy nuclear regions using Hyb-Seq<sup>23</sup> (Fig. 1). Integrating these two complementary sequencing strategies alongside a comprehensive morphological study enabled us to exploit the resources found in natural history collections and contribute to a diverse range of contemporary issues, including the origin of a major crop, the temporal and spatial dynamics of how the New World tropical flora was assembled, and the discovery of a substantial number of new species.

### Tackling megadiverse groups on a global scale

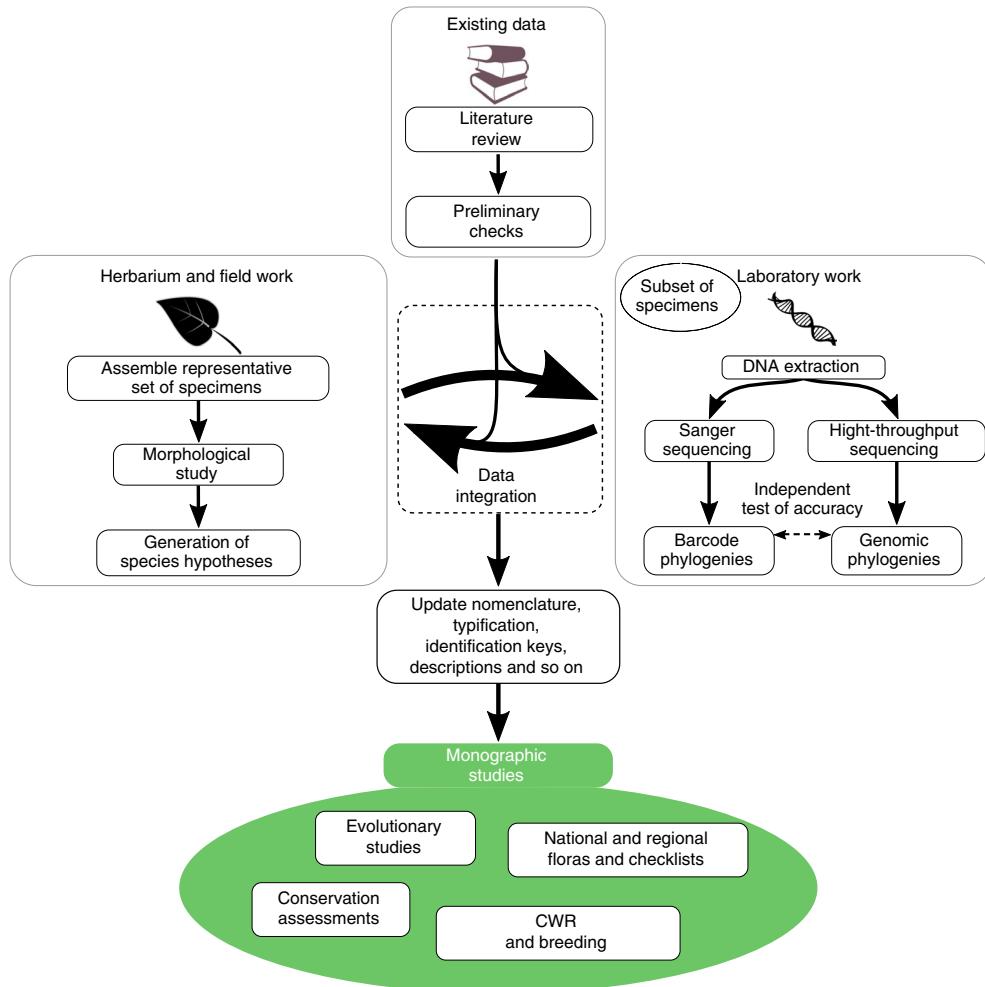
Present in all tropical and subtropical regions of the world, *Ipomoea* is among the largest genera of plants<sup>24</sup>. The taxonomic knowledge of the genus at the beginning of our project, in 2012, was relatively poor. The extensive body of literature and the existing taxonomy contained as much error as valuable information, reflected in the fact that more than 50% of *Ipomoea* names in the Global Biodiversity Information Facility, assigned to over 40,000 plant specimen records, are not currently accepted (Supplementary Data 1). Given this unsatisfactory situation, simple tasks such as identifying specimens, enumerating species from a particular country or preparing conservation assessments were problematic.

We based our approach to this comprehensive study of *Ipomoea* on the experience we had gained from a previous foundation monograph of *Convolvulus*<sup>25</sup>. We began our work by preparing a working checklist of all recognized species of *Ipomoea* (Supplementary Methods, Section 1) together with their commoner synonyms and their approximate distribution. On the basis of the distribution of individual species and their authors, we were able to predict which herbaria were likely to hold important collections of *Ipomoea*, including type specimens (Supplementary Methods, Sections 2 and 3). With a minimum estimate of 200,000 specimens of *Ipomoea* in the world's herbaria (Supplementary Methods, Section 2), obtaining all specimens on loan was neither practical nor necessary. Fortunately, we had ready access to large collections of *Ipomoea* at Kew Gardens (K) and the Natural History Museum in London (BM). By combining the study of specimens at these institutions with images in virtual herbaria and the insights of previous taxonomists (Supplementary Methods, Sections 3 and 4), we were able to determine important and useful taxonomic characters and thus begin to delimit species (Supplementary Methods, Section 5).

From the outset of the project, we aimed to integrate molecular and morphological data at all stages of the taxonomic process, each kind of data providing reciprocal illumination for many taxonomic decisions (Fig. 2).

Our approach was based on the idea that higher confidence for each species hypothesis is achieved when morphology and DNA barcodes—and genomic data when available—correlate, corroborating a species hypothesis. With this aim, and in parallel to our morphological studies, we started sequencing three DNA barcodes (nuclear *ITS* and chloroplast *matK* and *rbcL* regions) from specimens available to us from our own collections, from K and BM, an additional 45 other herbaria and individual sources (Supplementary Methods, Sections 6–8; Extended Data Fig. 1; and Supplementary Data 2). Our aim was to include, when possible, several specimens of every species in the phylogenies, as well as un-named specimens or specimens that we considered, from our morphological studies, to be interesting or puzzling. From this extensive sampling strategy, we gradually developed a provisional phylogenetic framework to inform species delimitation.

Given the time constraints and the large quantity of species we were trying to study, we were unable to optimize conditions for extracting and sequencing DNA from intractable specimens but, instead, opted to find alternative specimens or simply to move on. About one and a half years into the project, we decided to focus our barcode sequencing solely on *ITS* as it had provided most resolution and the highest success in extracting and sequencing DNA (about 60% of the specimens extracted were successfully amplified). We treated the *ITS* phylogeny (Supplementary Data 3) as a single taxonomic character and thus equivalent to a single morphological character<sup>26</sup> that might sometimes provide information for species delimitation and sometimes not (Extended Data Fig. 2). In many cases, the *ITS* phylogeny corroborated a species hypothesis based on morphology by showing it to be monophyletic. In other cases, the *ITS* phylogeny also revealed that specimens a priori thought to be the same species were, in reality, different taxa, in which case we re-evaluated the morphology and sequenced additional specimens where these were available. For other species, the *ITS* phylogeny provided little or no resolution, for example in the group of species most closely related to the sweet potato (sometimes spelled sweet-potato), *Ipomoea batatas* (L.) Lam. In these cases, we tested species hypotheses using genomic data<sup>27</sup> (see below). If no genomic data were available, we based our species delimitation on morphology



**Fig. 2 | Integrating morphology and DNA in global taxonomic studies is key to utilizing the resources of natural history collections.** The study of plant groups across their entire geographical distribution results in an accurate taxonomy that enables the assembly of national and regional checklists and floras, and also provides an essential framework for subsequent evolutionary studies, conservation assessments and research on CWRs and food security.

alone (Supplementary Discussion, DNA barcodes as another taxonomic character).

We were nevertheless aware of the many limitations of single-marker phylogenies<sup>28–30</sup> and of the inability of ITS to provide a robust and independent phylogenetic framework for *Ipomoea*<sup>31–33</sup>. Our whole approach to the interpretation of the ITS phylogeny was, therefore, one of extreme caution and, in addition, we had always planned to secure a greater amount of sequence data using high-throughput sequencing. We used Hyb-Seq<sup>23</sup> to obtain 605 nuclear regions and the whole chloroplast genome of 384 samples of *Ipomoea* representing 211 species (Supplementary Methods, Section 8). These data allowed us to obtain more robust phylogenies for *Ipomoea* (Extended Data Fig. 3 and Supplementary Data 4–8), to test the accuracy of the ITS phylogeny and to critically evaluate species delimitation in relation to the sweet potato and its closest relatives<sup>27</sup>. In summary, incorporating molecular phylogenetics into the taxonomic process provided a phylogenetic structure for *Ipomoea* as well as insights into species relationships, ultimately contributing to the taxonomic process at a number of levels (Box 1 and Fig. 2).

Species delimitation proceeds by looking for discrete and correlated characters that separate entities that are hypothesized to be ‘separately evolving metapopulation lineages’<sup>34</sup>. As the process of species delimitation is extended and complex, involving the

integration of morphology, DNA sequencing, previous literature, photographs and fieldwork, DNA sequencing alone is not sufficient to underpin taxonomic decisions. In contrast, when integrated with other sources of data, it can be extremely powerful. We provide eight examples to illustrate the process of species delimitation and taxonomic decision-making that underpinned this work (Supplementary Discussion, Species narratives).

#### Key taxonomic results

An accurate taxonomy of a plant group across its entire geographical distribution enables the assembly of checklists and floras at different scales. Figure 3a illustrates the power and importance of continental-scale taxonomy conducted against the backdrop of a global phylogenetic framework. This figure shows that the 109 species of *Ipomoea* known from Bolivia<sup>35–37</sup>—20 of them described as new species during this project—are dispersed across the entire phylogeny of the genus, underlining the limitations of geographically restricted studies.

The power of the global approach is also illustrated by the number of specimens that required a name change as a result of our studies—39% of specimens sequenced (Fig. 3b) (see specific examples of species delimitations and synonymy in Supplementary Discussion, Species narratives). In addition to the large number of new identifications provided, we described 63 new species, all of them dispersed

**Box 1 | Contribution of the DNA to the taxonomic decision process**

At the species-level taxonomy, DNA has:

1. Confirmed the monophyly of many species.
2. Drawn attention to the existence of unrecognized new species.
3. Shown some species thought to be distinct are conspecific with others from different geographical areas (for example, *I. acanthocarpa* from Africa with *I. piurensis* from America, or *I. lindenii* from mainland America with the Jamaican endemic *I. cyanantha*).
4. Shown that some species sometimes thought to be the same are distinct (for example *I. paludicola* and *I. asarifolia*, *I. huayllae* and *I. aristolochiifolia*, *I. jalapa* and *I. pterocaulis*, and so on).
5. Revealed wrongly identified specimens as they appear in parts of the phylogeny away from the clade with which they had been identified.
6. Provided a phylogenetic context to interpret morphology when specimens were poorly preserved.

Regarding evolutionary relationships between species, DNA has:

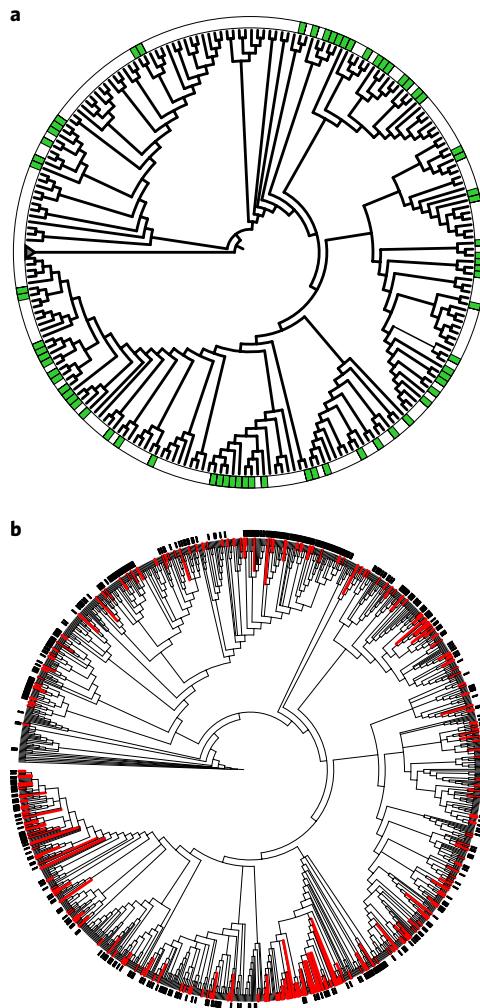
1. Revealed the existence of several clades and radiations.
2. Confirmed the monophyly of some groups previously recognized on morphological grounds such as *Pharbitis*, *Quamoclit*, *Astripomoea* and *Batatas*.
3. Shown that all previously recognized genera of the tribe Ipomoeae (*Argyreia*, *Stictocardia* and so on) are nested within *Ipomoea*, and all but *Astripomoea* are not monophyletic.
4. Demonstrated that *Rivea* is nested within the clade dominated by *Argyreia* species.
5. Shown that some groups previously recognized are monophyletic only if certain species are excluded (for example, *Arborescens* group).
6. Clarified the relationship between the sweet potato and its wild relatives, and discovered two new species within this group.

throughout the phylogenetic breadth of *Ipomoea*. Importantly, our contribution to the taxonomy of *Ipomoea* documented a 69% synonymy rate: seven out of every ten published names are synonyms<sup>38</sup>. In addition, we lectotypified 274 names and published 423 descriptions, 257 new illustrations, 43 distribution maps and 27 identification keys<sup>36–46</sup>.

Finally, our phylogenies confirm that many previously recognized segregated genera are nested within *Ipomoea*<sup>31,47</sup> (Extended Data Fig. 3) and that an expanded *Ipomoea* containing these species is necessary to make the genus monophyletic (Supplementary Discussion, Phylogeny of *Ipomoea*). New combinations for all names in other genera that need transferring into *Ipomoea* are provided in Supplementary Discussion, Nomenclatural changes.

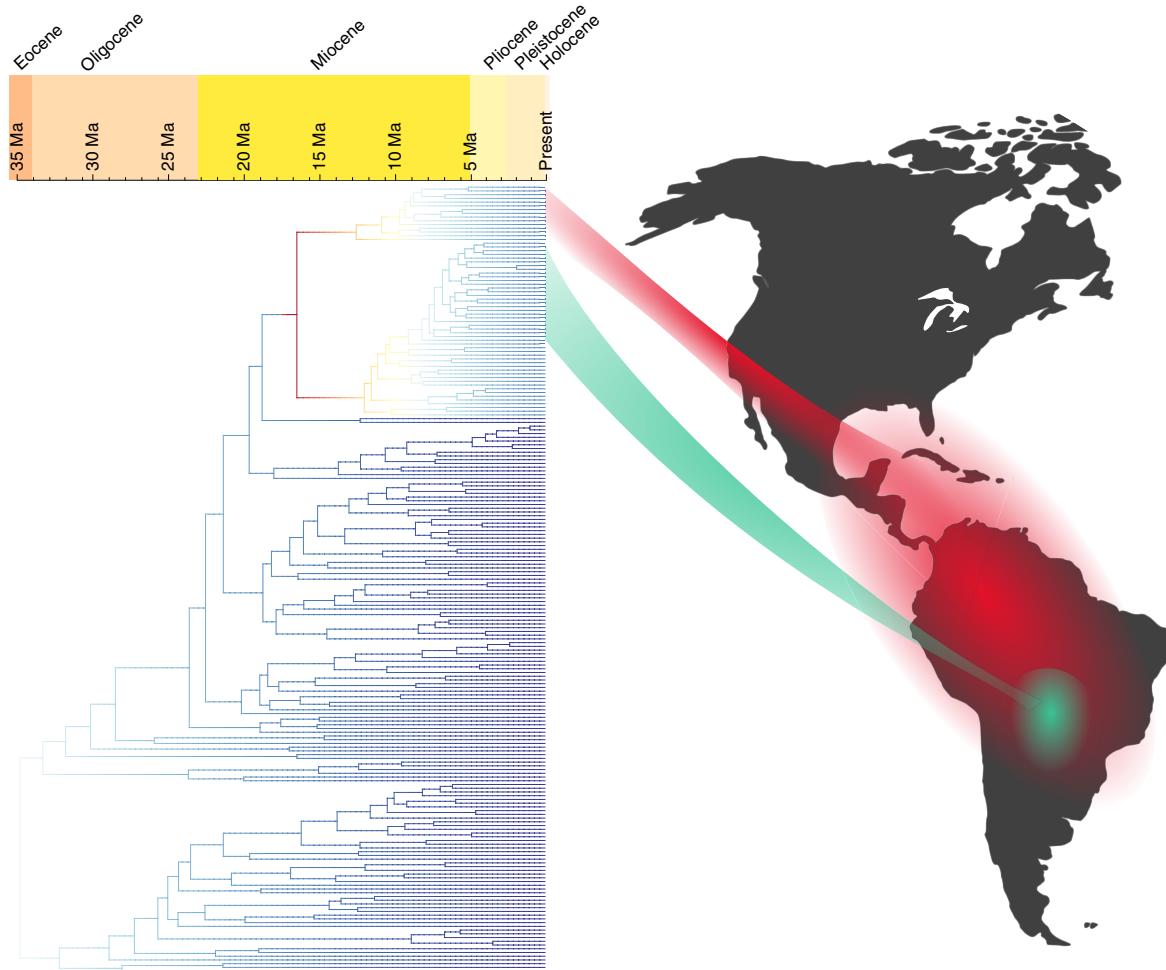
### Rapid radiations in *Ipomoea*

A by-product of our focus on species-level taxonomy and DNA sequencing was a comprehensively sampled phylogenetic framework for *Ipomoea* that provided valuable information at multiple levels. During our studies, we became aware of two very diverse clades within *Ipomoea* in which species morphologies overlap considerably and phylogenetic relationships are poorly resolved. One of these clades is concentrated in central South America (Paraguay, southeast Bolivia, southwest Brazil and northern Argentina), while the other is more widespread in the Americas but with a particularly



**Fig. 3 | Megadiverse plant groups demand a global approach.** **a**, Nuclear genomic phylogeny showing that the species recorded from Bolivia (green boxes) are scattered across the phylogeny of the genus, which has a global distribution. **b**, ITS phylogeny of *Ipomoea*. The red branches indicate specimens also sequenced using high-throughput sequencing. The black boxes indicate specimens that we sequenced that changed their identification during our studies, approximately 39% of them. Many more specimens not included in our molecular analyses also required a change of name.

high concentration of species in the Caribbean region. These two diverse clades are closely related in our nuclear and chloroplast phylogenies, although the exact relationship differs between the two datasets (Extended Data Fig. 3a,b). In view of the unique characteristics of these two clades, we constructed a time-calibrated phylogeny for *Ipomoea* and estimated diversification rates throughout the genus (Fig. 4 and Extended Data Figs. 4–6). This showed that diversification rates were relatively constant in most of the genus, except for the part of the phylogeny that contained these two diverse clades (and a small number of other species). In this part of the phylogeny, there was initially a greater than 5.5-fold increase in net diversification rates compared to the background rate across the rest of the tree (an increase from 0.127 to 0.719 species Myr<sup>-1</sup>). Our analyses indicated that this was primarily a result of increased speciation rates, with extinction rates remaining relatively constant. Although our analysis indicated a diversification rate increase in the early Miocene, more recent phenomena might also influence the



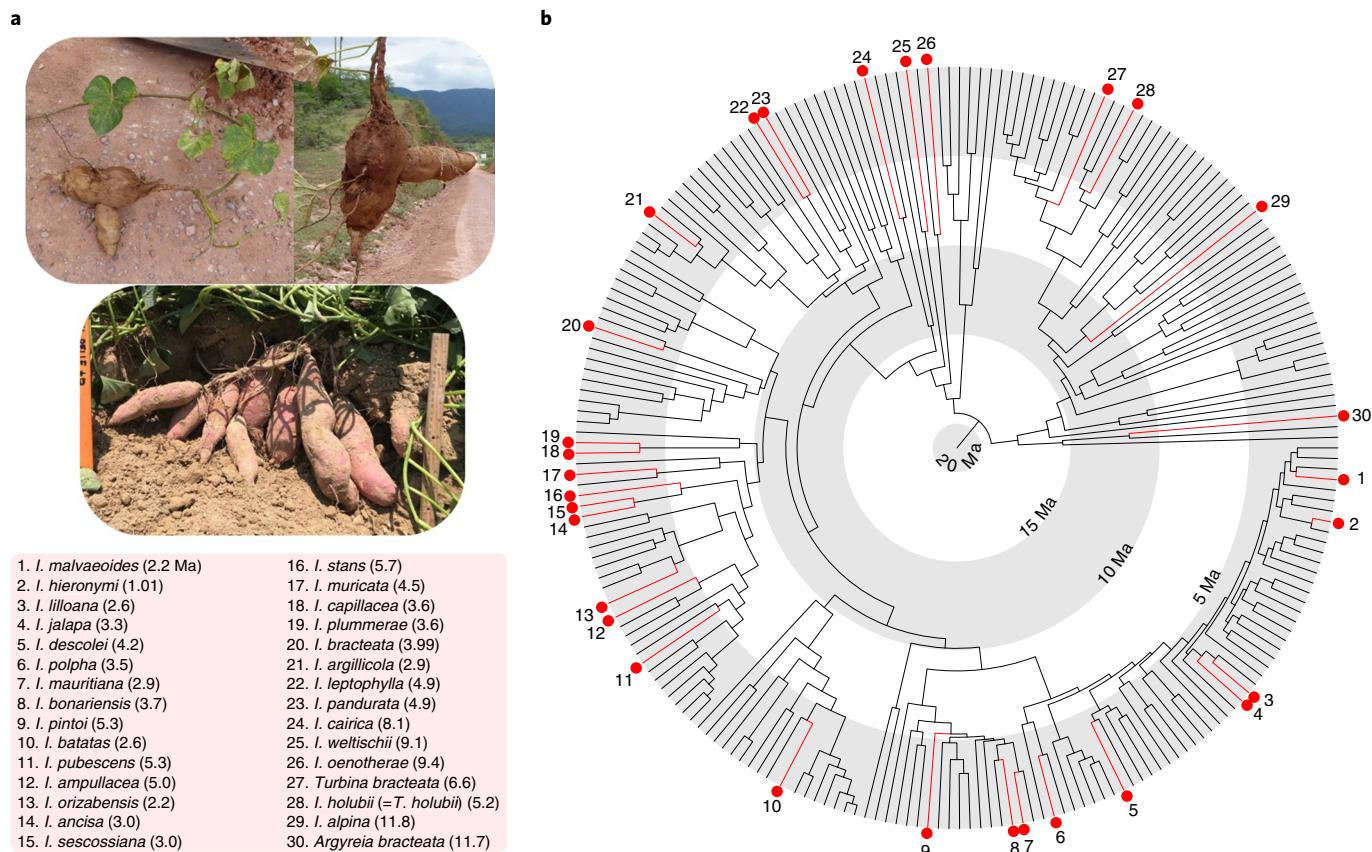
**Fig. 4 | Rapid radiations in *Ipomoea*.** A time-calibrated phylogeny of *Ipomoea*, with branches coloured according to the inferred speciation rate. The map indicates the geographic distribution of two species-rich clades, the species within which exhibit highly overlapping morphologies. Both of these two diverse clades (and a small number of other species) are part of a larger clade in which speciation rates are significantly higher than for the rest of *Ipomoea*. Ma, million years ago.

distinctive diversification dynamics in this part of the phylogeny (for example, many species in this part of the phylogeny occur exclusively in the Cerrado—a biome that probably became established only within the last 10 Myr<sup>48,49</sup>—and there are likely to have been numerous shifts into and out of this biome (Extended Data Fig. 7)). Further, numerous shifts between different growth habits are also likely to have occurred between comparatively recently diverged lineages (Extended Data Fig. 7). A more densely sampled phylogeny is required to determine the nature of the relationship between biome occupancy and growth habit, and whether either of these two factors is likely to have promoted multiple nested diversification rate shifts, rather than the single rate increase reported here. Regardless, our results highlight an increase in net diversification rates in *Ipomoea* that is likely to be of a similar scale to some of the most iconic evolutionary radiations in the plant kingdom<sup>50–53</sup>. Further, unlike many plant radiations, which are strongly associated with a transition into a particular biome, the radiation in *Ipomoea* occurs across a range of biomes, and in some cases, in areas that have been greatly disturbed by human actions. Further study of diversification rate variation in *Ipomoea*, therefore, represents a promising avenue that could lead to fundamental insights into the effects of biome shifts and human disturbance on evolutionary diversification and the assembly of the Neotropical flora.

### Evolution of the sweet potato

Most recent studies on the origin of the sweet potato (*I. batatas* (L.) Lam.) focus on the genetic variation contained within the crop<sup>54,55</sup> or on the sequencing of whole genomes of the crop and one or two related species<sup>56,57</sup>. Meanwhile, the origin and evolution of the sweet potato and its relationship with its crop wild relatives (CWR) has only recently been clarified<sup>27</sup>. The global study of the genus allowed us to identify all sweet potato CWR—two of them new species, *I. lactifera* J. R. I. Wood & Scotland<sup>36</sup> and *I. australis* (O'Donell) J. R. I. Wood & P. Muñoz<sup>38</sup>—and revealed the dual role of *I. trifida* (Kunth) G. Don, the closest wild relative, in the origin of the crop species<sup>27</sup>.

Previous studies have shown that sweet potato CWR do not produce storage roots<sup>58</sup>, so it has been assumed that the transition from non-storage root to storage root was mediated by human domestication<sup>33</sup>, although direct evidence for this claim remains elusive. However, our broad comparative study of the genus offers a novel perspective on the evolution of storage roots in *Ipomoea* and a very different narrative for the evolution of the sweet potato. At least 63 species of *Ipomoea* have been recorded in previously published literature and through our own observations as having storage roots, several of them edible and some bigger than the roots of *I. batatas* (Fig. 5a and Supplementary Information, and see Supplementary



**Fig. 5 | Storage roots evolved multiple times independently in *Ipomoea*.** **a**, Storage roots in *I. lilloana* (top) are as big as those in the sweet potato (bottom).

**b**, Time-calibrated nuclear maximum-likelihood phylogeny highlighting the position of 30 species with storage roots, indicated by red branches and dots.

All of these species originated at least 1 million years ago. We have recorded an additional 33 species with storage roots for which we do not have genomic data. Photographs courtesy of J.R.I.W. (**a**, top) and A. Villordon, Louisiana State Univ. (**a**, bottom).

Table 1 for a list of all species with storage roots). Mapping species with storage roots onto a phylogeny shows that storage roots evolved multiple times independently from species that do not have storage roots (or these have never been recorded) (Fig. 5b).

We wanted to explore this question further and used our time-calibrated phylogenies to investigate the temporal dynamics of sweet potato. We set out to determine whether our data were consistent with sweet potato originating within the time frame of human agriculture (roughly the last 10,000 years) or if it was older. Our results indicated that the sweet potato was likely to have diverged from its closest wild relative, *I. trifida*, over 1 million years ago<sup>27</sup> (Fig. 5b) and that part of the diversity existing within the crop largely predated the origin of agriculture (Fig. 6). This time frame is consistent with the idea that the sweet potato evolved long before the onset of human agriculture, and that the storage root was an existing trait that favoured the species being taken into cultivation by humans. Further, all other species with storage roots also evolved over 1 million years ago (Fig. 5b), many within the time frame associated with the expansion of C4 grasses and the evolution of fire-adapted vegetation types<sup>48,49</sup> in which underground storage organs would be advantageous. In summary, the evidence presented here suggests that the storage root in cultivated sweet potato is not a product of human domestication but rather an existing trait that predisposed the plant for cultivation. To the best of our knowledge, this possibility has not been previously considered.

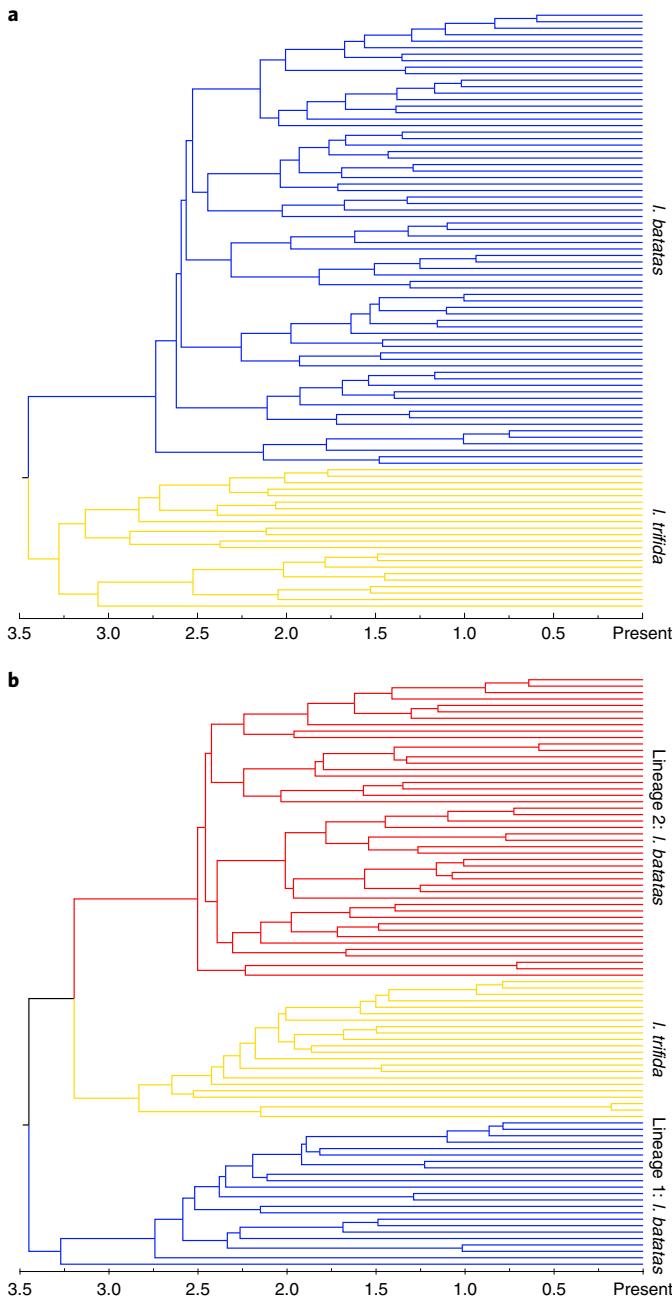
### The importance and potential of taxonomic monography

Taxonomic studies based on the massive number of natural history collections held worldwide highlight the awesome complexity and

wonder of the natural world. They merit a more important role in the task of addressing a range of environmental issues from food security, conservation and biodiversity inventories to ecology in general. The taxonomic community itself needs to embrace and rediscover the value of taxonomic monographs<sup>25,59</sup> within the context of what constitutes world-class science<sup>60</sup>. The full integration of two distinct skill sets, DNA sequencing and morphological studies, is necessary to achieve this. Although other scientific subjects bring a unique perspective to environmental science, including evolution, ecology and population genetics, monographic taxonomy undertaken with modern methods at the global scale has the potential to play a vital role in the contemporary research agenda.

Taxonomy is often seen as a redundant science because of the mistaken idea that biodiversity is as well known overall as it is in a few well-studied, high-profile groups or countries. It is also undervalued by the inaccurate view that taxonomic knowledge steadily accumulates until all species of a particular group are discovered, whereas in reality names, synonyms, mistaken identifications and errors accumulate alongside accepted names and reliable information. This accretion needs to be sifted and new species identified to provide an accurate taxonomy, something that is lacking for the vast majority of tropical flowering plant genera of any reasonable size. With the rapid increase in the number of unstudied collections in the last 50 years, there is now a unique opportunity to embrace the challenges and opportunities that these specimens provide to produce taxonomically sound monographs of the plant diversity these natural history collections represent.

To fully exploit the opportunity and potential of global natural history collections, as undertaken in this study, demands the



**Fig. 6 | Diversity within sweet potato predates agriculture.** Time-calibrated phylogenies for sampled specimens of *I. batatas* and its closest relative *I. trifida*. The divergence times indicate when lineages represented by different specimens are likely to have diverged. **a,b**, Divergence times inferred using nuclear genomic data (**a**) and whole chloroplast genome data (**b**). The two *I. batatas* clades in **b** correspond to the two chloroplast lineages hypothesized in ref. <sup>27</sup>.

integration of different scientific expertise including specimen-based taxonomy, genomics and phylogenetics. This has implications for the type of training that the next generation of biodiversity scientists receive. It seems unrealistic to expect an individual scientist to be expert in all three disciplines but assembling small teams of people with such expertise to tackle the world's major taxonomic problems at a global scale is surely possible given existing resources and expertise. The skills and resources currently exist for the study of many taxonomically diverse groups (and as long as taxonomic

training continues or is increased) and we hope that this study acts as a catalyst in demonstrating the scale of progress that can be achieved in a realistic time frame.

## Methods

In this section, we provide a summary of the methodology underlying our studies of *Ipomoea*. We provide a detailed description of every step in the Supplementary Methods. Although we report the morphology and molecular methods separately, they were, in fact, conducted in parallel and integrated throughout the process.

**Herbarium and field work.** We assembled a preliminary checklist from existing literature for all species of *Ipomoea* (Supplementary Methods, Section 1) and identified herbaria that house significant collections that we would visit or from which we could obtain online images (Supplementary Methods, Sections 2 and 3). Simultaneously, we surveyed morphological variation across the genus—with reference to existing literature as well as specimens—to identify taxonomically useful characters for species delimitation (Supplementary Methods, Sections 4 and 5). We subsequently visited, received loans of material from or studied photographs from the following herbaria (acronyms according to ref. <sup>61</sup>) in Europe (AAU, B, BM, C, CGE, E, G, GOET, K, L, LE, M, MA, OXF, P, PC, RBGE, S, TO and W), the USA (A, ARIZ, BISH, F, FTG, GA, GH, MICH, MO, NY, RSA, SELU, TEX, US and USDA), Latin America (Argentina: CTES and LIL; Bolivia: BOLV, HSB, LPB and USZ; Brazil: CEN, CPAP, CRIA, HEPH, HUEFS, IPA, JPB, MBM, PEUFR, R, RB, SP and UB; Colombia: COL; Cuba: HACB and HAJB; Mexico: IEB and MEXU; Panama: PAM; Paraguay: FCQ, PY and SCP; Peru: CIP, CUZ and USM), China (ISBC and KUN), South East Asia (Malaysia: KEP and SAN; Singapore: SING) and Australia (FRI). We studied the variation in all herbarium material seen and photographed and databased specimens (Supplementary Methods, Sections 2–5). We carried out fieldwork in Bolivia, Paraguay, Argentina and Brazil (Supplementary Methods, Section 6). We also developed a network of contacts with people interested in *Ipomoea* with whom we corresponded over a range of related issues (Supplementary Methods, Section 7).

**Analysis of DNA barcodes.** The analyses using barcodes were based on 3,035 ITS, *matK* and *trnH* sequences from 1,560 specimens (Passport data in Extended Data File 1) (Extended Data Fig. 1). We aligned all sequences using MAFFT v.7.2.1 <sup>62,63</sup> and ran maximum-likelihood phylogenetic analyses in RAxML v.8.64, approximate maximum likelihood in FastTree 2<sup>65</sup> and Bayesian inference in MrBayes<sup>66</sup> (Supplementary Methods, Section 1).

**Analysis of genomic data.** We obtained the whole chloroplast genome and 605 putative single-copy nuclear coding regions from 385 specimens representing 211 species using Hyb-Seq<sup>23</sup> (Supplementary Methods, Section 8). These specimens were selected on the basis of the quality and quantity of the available DNA with the aim of covering as much phylogenetic breadth as possible. We ran phylogenetic analyses on both sets of genomic data. For the nuclear data, we ran additional analyses using only the subset of 434 regions that passed the PHI recombination test<sup>67</sup>. In addition, mapping our data to the recently published *I. trifolia* genome<sup>57</sup> warned us that some of our regions may not be single copy; hence, we ran further analyses using only the subset of 421 regions that we were confident are single copy (Supplementary Methods, Section 8). We used maximum likelihood, approximate maximum likelihood and Bayesian inference to analyse the chloroplast data. Regarding the nuclear coding regions, we used maximum likelihood and approximate maximum likelihood for the analysis of concatenated alignments as well as inferred species trees from gene trees using coalescence methods. All methods and datasets recovered the same major clades within *Ipomoea* and the relationship between taxa within those clades was mostly congruent across phylogenies (Supplementary Discussion, Phylogeny of *Ipomoea*).

**Divergence time estimates.** We estimated divergence times within *Ipomoea* in treePL<sup>58,69</sup>. We used the nuclear genomic phylogeny inferred in FastTree 2<sup>65</sup> as the input tree. We used a smoothing value of 0.01 following extensive cross-validation analyses (Supplementary Methods, Section 9), but also experimented with different smoothing values (0.01, 1, 100 and 10,000) to determine the sensitivity of divergence time estimates to different assumptions about among-branch rate variation. We also inferred time-calibrated phylogenies with the chloroplast phylogeny as the input tree. In this case, we also experimented with different smoothing values (0.01, 1, 10 and 10,000). For these phylogenies, we used a point calibration for the root node of 34.0 Myr. We consider this the most realistic age estimate for *Ipomoea*, following a series of analyses in which we experimented with different methods for calibrating a phylogeny for Convolvulaceae and Solanaceae. The analyses for Convolvulaceae and Solanaceae were performed in RevBayes<sup>70</sup> (Supplementary Methods, Section 9).

We used BAMM<sup>71</sup> to infer diversification rates. The time-calibrated phylogeny inferred from nuclear genomic data in treePL<sup>69</sup> was used as the input phylogeny. When performing this analysis, we specified clade-specific sampling fractions. These were taken into account when estimating diversification rates. We performed several supplementary diversification rate analyses. These used the different time-calibrated phylogenies outlined above as input phylogenies (Supplementary Methods, Section 9).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Passport data for all specimens included in the molecular studies presented in this paper are available in Supplementary Data File 2. Additional records and information for the collections included in this study and for specimens added subsequently are available through the project website (<https://herbaria.plants.ox.ac.uk/bol/ipomoea>). DNA barcode sequences are available through GenBank and genome assemblies are available through the Oxford Repository Archive (<https://doi.org/10.5287/bodleian:kepgnxzeK>). Illumina raw reads are available through the Sequence Read Archive (BioProject PRJNA453382). Alignment files and other materials are available from the corresponding author upon request.

Received: 9 July 2019; Accepted: 23 September 2019;

Published online: 11 November 2019

## References

- Brownsey, P. The Banks and Solander collections—a benchmark for understanding the New Zealand flora. *J. R. Soc. N. Z.* **42**, 131–137 (2012).
- Mabberley, D. J. *Jupiter botanicus: Robert Brown of the British Museum* 1st edn (Lubrecht & Cramer Ltd, 1985).
- Goodwin, Z. A., Harris, D. J., Filer, D., Wood, J. R. I. & Scotland, R. W. Widespread mistaken identity in tropical plant collections. *Curr. Biol.* **25**, R1066–R1067 (2015).
- The Science and Technology Committee *Systematics and Taxonomy in Crisis* (Authority of the House of Lords, 2008).
- Riedel, A., Sagata, K., Suhardjono, Y. R., Tänzler, R. & Balke, M. Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. *Front. Zool.* **10**, 15 (2013).
- Bradley, R. D., Bradley, L. C., Garner, H. J. & Baker, R. J. Assessing the value of natural history collections and addressing issues regarding long-term growth and care. *BioScience* **64**, 1150–1158 (2014).
- Plant Conservation Report. A Review of Progress in Implementing the Global Strategy for Plant Conservation (GSPC)* (Convention on Biological Diversity, 2009).
- Feeley, K. J. & Silman, M. R. The data void in modeling current and future distributions of tropical species. *Glob. Change Biol.* **17**, 626–630 (2011).
- Scotland, R. W. & Wood, J. R. I. Accelerating the pace of taxonomy. *Trends Ecol. Evol.* **27**, 415–416 (2012).
- Bisby, F. A., Shimura, J., Ruggiero, M., Edwards, J. & Haeuser, C. Taxonomy, at the click of a mouse. *Nature* **418**, 367–367 (2002).
- Joppa, L. N., Roberts, D. L. & Pimm, S. L. The population ecology and social behaviour of taxonomists. *Trends Ecol. Evol.* **26**, 551–553 (2011).
- Bacher, S. Still not enough taxonomists: reply to Joppa et al. *Trends Ecol. Evol.* **27**, 65–66 (2012).
- Wheeler, Q. D. et al. Mapping the biosphere: exploring species to understand the origin, organization and sustainability of biodiversity. *Syst. Biodivers.* **10**, 1–20 (2012).
- Costello, M. J., May, R. M. & Stork, N. E. Can we name Earth's species before they go extinct? *Science* **339**, 413–416 (2013).
- Bebber, D. P., Wood, J. R. I., Barker, C. & Scotland, R. W. Author inflation masks global capacity for species discovery in flowering plants. *New Phytol.* **201**, 700–706 (2014).
- Wortley, A. H. & Scotland, R. W. Synonymy, sampling and seed plant numbers. *TAXON* **53**, 478–480 (2004).
- Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P. DNA points the way ahead in taxonomy. *Nature* **418**, 479–479 (2002).
- Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P. A plea for DNA taxonomy. *Trends Ecol. Evol.* **18**, 70–74 (2003).
- Chase, M. W. et al. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Ann. Missouri Bot. Gard.* **80**, 528–580 (1993).
- The Angiosperm Phylogeny Group An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* **141**, 399–436 (2003).
- Hollingsworth, P. M., Li, D.-Z., van der Bank, M. & Twyford, A. D. Telling plant species apart with DNA: from barcodes to genomes. *Phil. Trans. R. Soc. B* **371**, 20150338 (2016).
- CBOL Plant Working Group et al. A DNA barcode for land plants. *Proc. Natl Acad. Sci. USA* **106**, 12794–12797 (2009).
- Weitemier, K. et al. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci.* **2**, 1400042 (2014).
- Frodin, D. G. History and concepts of big plant genera. *TAXON* **53**, 753–776 (2004).
- Wood, J. et al. A foundation monograph of *Convolvulus* L. (Convolvulaceae). *PhytoKeys* **51**, 1–282 (2015).
- Doyle, J. J. Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot.* **17**, 144 (1992).
- Muñoz-Rodríguez, P. et al. Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to Polynesia. *Curr. Biol.* **28**, 1246–1256 (2018).
- Baldwin, B. G. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Mol. Phylogenet. Evol.* **1**, 3–16 (1992).
- Álvarez, I. & Wendel, J. F. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* **29**, 417–434 (2003).
- Feliner, G. N. & Rosselló, J. A. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* **44**, 911–919 (2007).
- Miller, R. E., Rausher, M. D. & Manos, P. S. Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and Waxy sequences. *Syst. Bot.* **24**, 209–227 (1999).
- Huang, J., Corke, H. & Sun, M. Highly polymorphic AFLP markers as a complementary tool to ITS sequences in assessing genetic diversity and phylogenetic relationships of sweetpotato (*Ipomoea batatas* (L.) Lam.) and its wild relatives. *Genet. Resour. Crop Evol.* **49**, 541–550 (2002).
- Roullier, C. et al. Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). *PLoS ONE* **8**, e62707 (2013).
- De Queiroz, K. Species concepts and species delimitation. *Syst. Biol.* **56**, 879–886 (2007).
- Wood, J. R. I., Bianchini, R. S. & Fuentes, A. F. in *Catálogo de las Plantas Vasculares de Bolivia* (eds Jorgensen, P. M. et al.) 520–531 (Missouri Botanical Garden Press, 2015).
- Wood, J. R. I. et al. *Ipomoea* (Convolvulaceae) in Bolivia. *Kew Bull.* **70**, 31 (2015).
- Wood, J. R. I., Martínez Ugarteche, M. T., Muñoz-Rodríguez, P. & Scotland, R. W. Additional notes on *Ipomoea* (Convolvulaceae) in Bolivia. *Kew Bull.* **73**, 57 (2018).
- Wood, J. R. I., Muñoz-Rodríguez, P., Williams, B. R. M. & Scotland, R. W. A foundation monograph of *Ipomoea* (Convolvulaceae) in the New World. *Phytokeys* (in the press).
- Wood, J. R. I., de Arrúa, R. D., de Rojas, G. D. & Scotland, R. W. Two overlooked species of *Ipomoea* L. (Convolvulaceae) from Paraguay. *Kew Bull.* **71**, 25 (2016).
- Wood, J. R. I., Urbanetz, C. & Scotland, R. W. *Ipomoea pantanalensis*, a new species of *Ipomoea* L. (Convolvulaceae) from the Pantanal, Brazil. *Kew Bull.* **71**, 6 (2016).
- Wood, J. R. I. & Scotland, R. W. Notes on *Ipomoea* L. (Convolvulaceae) in Cuba and neighbouring islands with a checklist of species found in Cuba. *Kew Bull.* **72**, 45 (2017).
- Wood, J. R. I. & Scotland, R. W. Misapplied names, synonyms and new species of *Ipomoea* (Convolvulaceae) from South America. *Kew Bull.* **72**, 9 (2017).
- Wood, J. R. I. & Scotland, R. W. Notes on *Ipomoea* (Convolvulaceae) from the Amazonian periphery. *Kew Bull.* **72**, 10 (2017).
- Wood, J. R. I., Muñoz-Rodríguez, P., Degen, R. & Scotland, R. W. New species of *Ipomoea* (Convolvulaceae) from South America. *PhytoKeys* **88**, 1–38 (2017).
- Wood, J. R. I., Buril, M. T. & Scotland, R. W. Remarkable disjunctions in *Ipomoea* species (Convolvulaceae) from NE Brazil and Central America and their taxonomic implications. *Kew Bull.* **72**, 44 (2017).
- Wood, J. R. I., Vasconcelos, L. V., Simão-Bianchini, R. & Scotland, R. W. New species of *Ipomoea* (Convolvulaceae) from Bahia. *Kew Bull.* **72**, 8 (2017).
- Wilkin, P. A morphological cladistic analysis of the Ipomoeae (Convolvulaceae). *Kew Bull.* **54**, 853–876 (1999).
- Beerling, D. J. & Osborne, C. P. The origin of the savanna biome. *Glob. Change Biol.* **12**, 2023–2031 (2006).
- Scheiter, S. et al. Fire and fire-adapted vegetation promoted C4 expansion in the late Miocene. *New Phytol.* **195**, 653–666 (2012).
- Baldwin, B. G. & Sanderson, M. J. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl Acad. Sci. USA* **95**, 9402–9406 (1998).
- Hughes, C. & Eastwood, R. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc. Natl Acad. Sci. USA* **103**, 10334–10339 (2006).
- Givnish, T. J. et al. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proc. R. Soc. B* **276**, 407–416 (2009).
- Koenen, E. J. M. et al. Exploring the tempo of species diversification in legumes. *South Afr. J. Bot.* **89**, 19–30 (2013).
- Zhang, D., Cervantes, J., Huamán, Z., Carey, E. & Ghislain, M. Assessing genetic diversity of sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars from tropical America using AFLP. *Genet. Resour. Crop Evol.* **47**, 659–665 (2000).
- Roullier, C., Kambouo, R., Paofa, J., McKey, D. & Lebot, V. On the origin of sweet potato (*Ipomoea batatas* (L.) Lam.) genetic diversity in New Guinea, a secondary centre of diversity. *Heredity* **110**, 594–604 (2013).

56. Yang, J. et al. Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nat. Plants* **3**, 696–703 (2017).
57. Wu, S. et al. Genome sequences of two diploid wild relatives of cultivated sweet potato reveal targets for genetic improvement. *Nat. Commun.* **9**, 4580 (2018).
58. Austin, D. F. The *Ipomoea batatas* complex-I. Taxonomy. *Bull. Torrey Bot. Club* **105**, 114–129 (1978).
59. Harris, D. J. & Wortley, A. H. *Monograph of Aframomum (Zingiberaceae)* (The American Society of Plant Taxonomists, 2018).
60. Drew, L. W. Are we losing the science of taxonomy? As need grows, numbers and training are failing to keep up. *BioScience* **61**, 942–946 (2011).
61. Thiers, B. *Index Herbariorum: A Global Directory of Public Herbaria and Associated Staff* (New York Botanical Garden's Virtual Herbarium, 2018).
62. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
63. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
64. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
65. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
66. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
67. Bruen, T. C. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681 (2005).
68. Sanderson, M. J. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* **19**, 101–109 (2002).
69. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
70. Hohna, S. et al. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).
71. Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* **9**, e89543 (2014).

## Acknowledgements

We acknowledge the financial support of The Leverhulme Trust for our *Ipomoea* Foundation Monograph project and the University of Oxford through The John Fell

Fund for travel and sequencing costs. P.M.-R. was funded by a BBSRC scholarship granted through the Interdisciplinary Bioscience DTP Programme and by the University of Oxford Global Challenges Research Fund; he also received additional funding from a Santander Travel Award and from the Synthesys project (FR-TAF-6575). J.R.I.W. received travel awards from the Synthesis project to visit Paris (FR-TAF), Madrid (ES-TAF) and Stockholm (SE-TAF) and B.R.M.W. received a Synthesis travel award to visit Leiden (NL-TAF). R.W.S. and P.M.-R. acknowledge funding from the BBSRC GCRF-IAA fund (BB/GCRF-IAA/16 and BB/GCRF-IAA/17/16). T.C. was funded by a NERC scholarship granted through the Environmental Research DTP Programme. We thank all herbarium curators for granting access to their collections. We thank T. Wells for his comments on the genomic analyses. We also thank all colleagues who contributed to this project through fieldwork and continuous discussion (see the list in Supplementary Information, Section 7).

## Author contributions

Conceptualization, supervision and project administration was performed by R.W.S. Funding acquisition was performed by R.W.S., J.R.I.W., P.M.-R. and T.C. Methodology was defined by R.W.S., J.R.I.W., A.L., S.K., K.W., B.K., D.H., D.F., P.M.-R. and T.C. Resources were obtained by J.R.I.W., B.R.M.W., P.M.-R., A.S., Z.G., N.L.A. and M.D.R. Formal analysis and investigation were performed by P.M.-R., T.C. and J.R.I.W. Writing of the original draft was performed by P.M.-R., R.W.S., T.C. and J.R.I.W., and writing and review of the final draft was performed by all authors. Visualization and image design was performed by P.M.-R.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41477-019-0535-4>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41477-019-0535-4>.

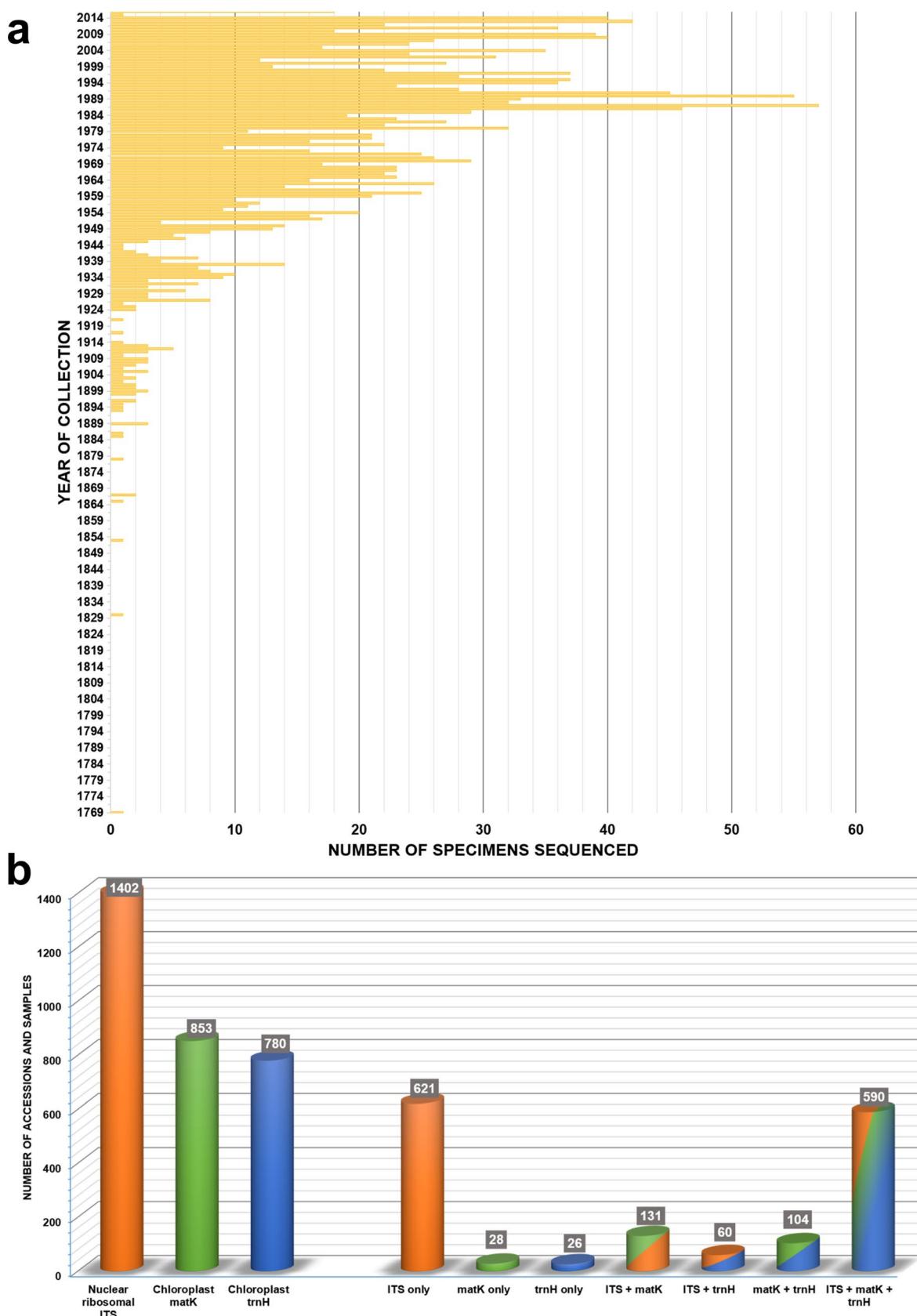
**Correspondence and requests for materials** should be addressed to R.W.S.

**Peer review information** *Nature Plants* thanks Matt Lavin, Quentin Wheeler and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

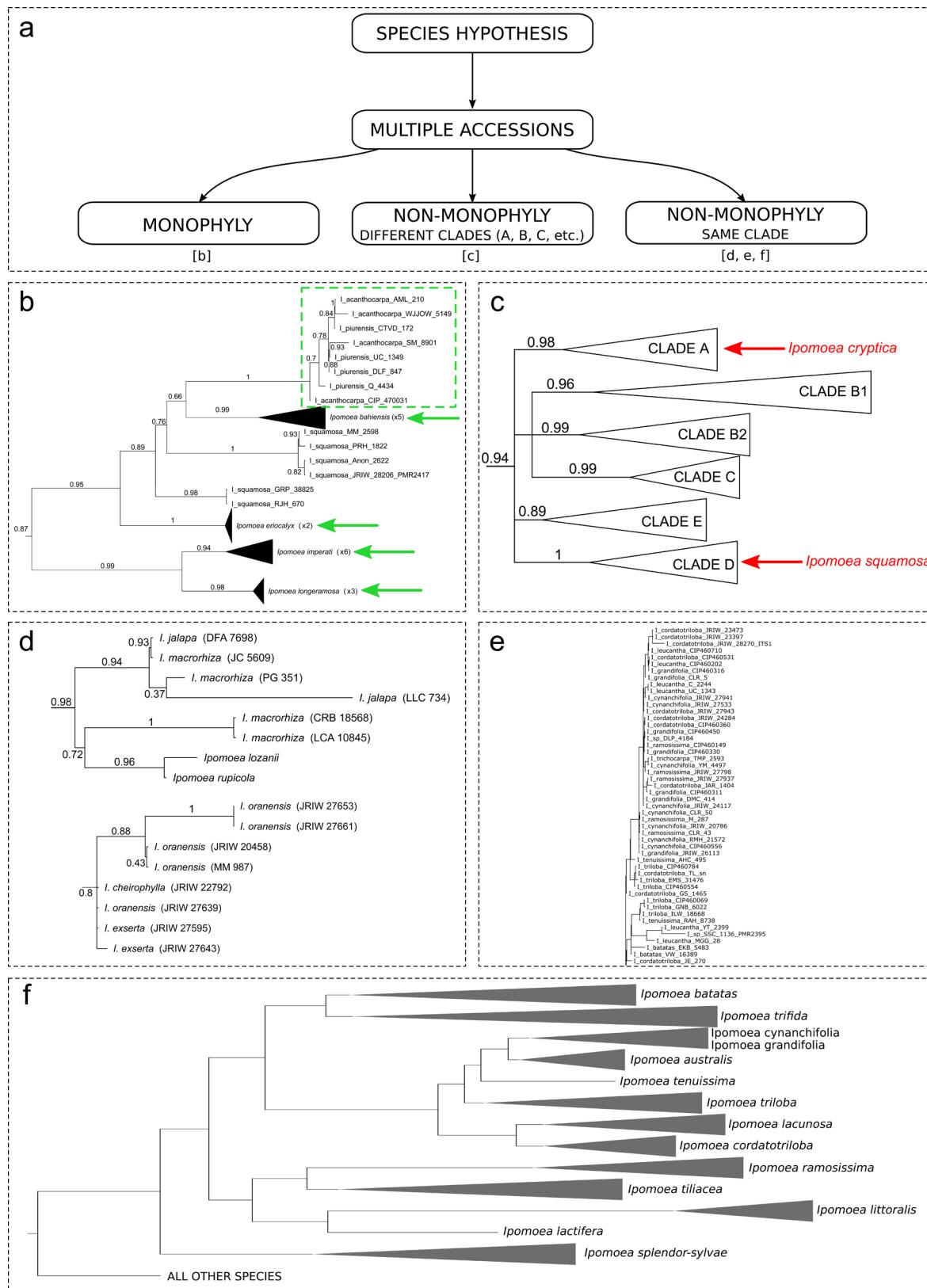
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

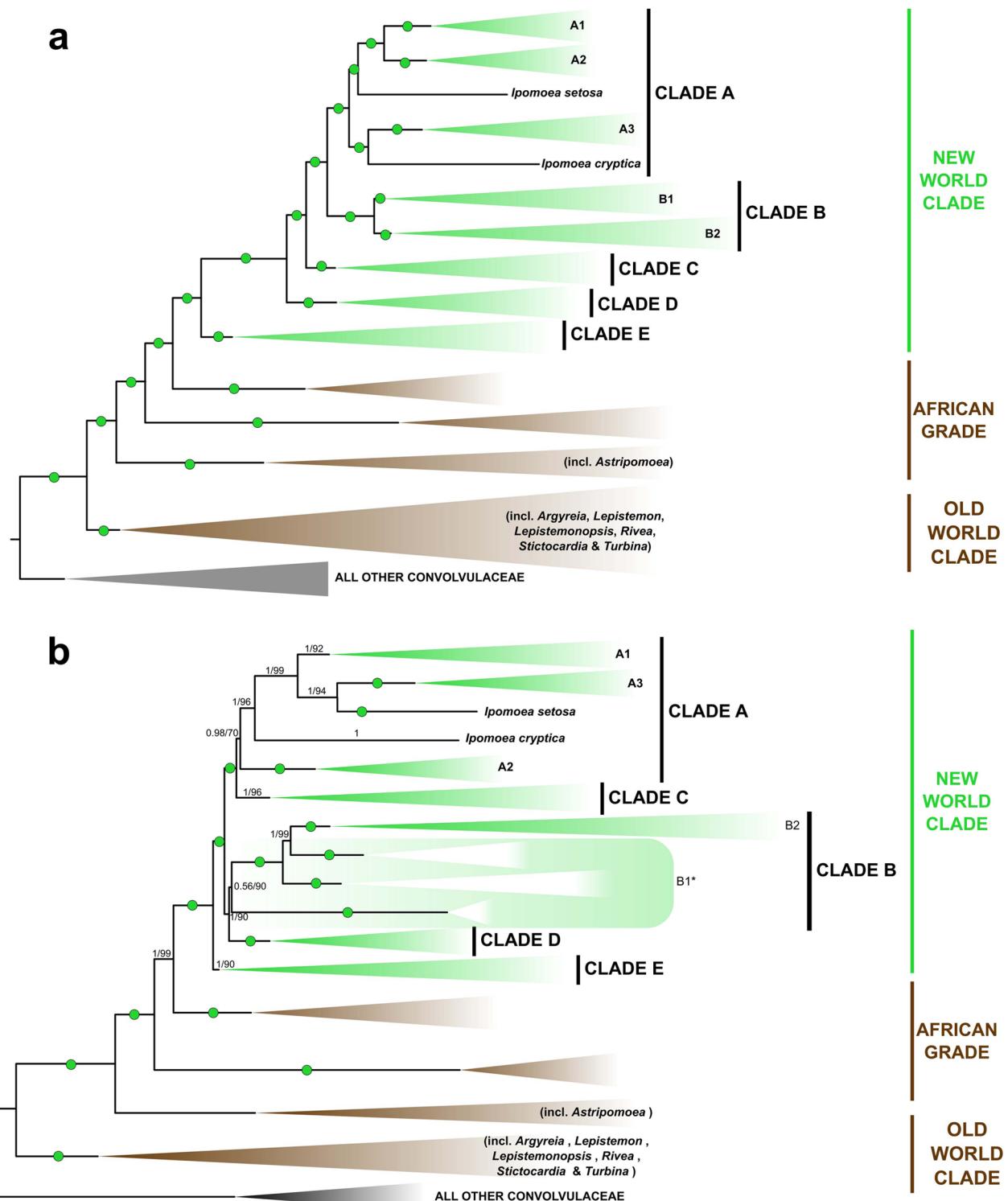


**Extended Data Fig. 1 | DNA barcode sequences used and statistics.** 3,035 DNA barcode sequences were obtained from herbarium specimens spanning the last two and a half centuries. 88.5% specimens were collected in the last 50 years and 22 samples came from pre-20<sup>th</sup> century collections.

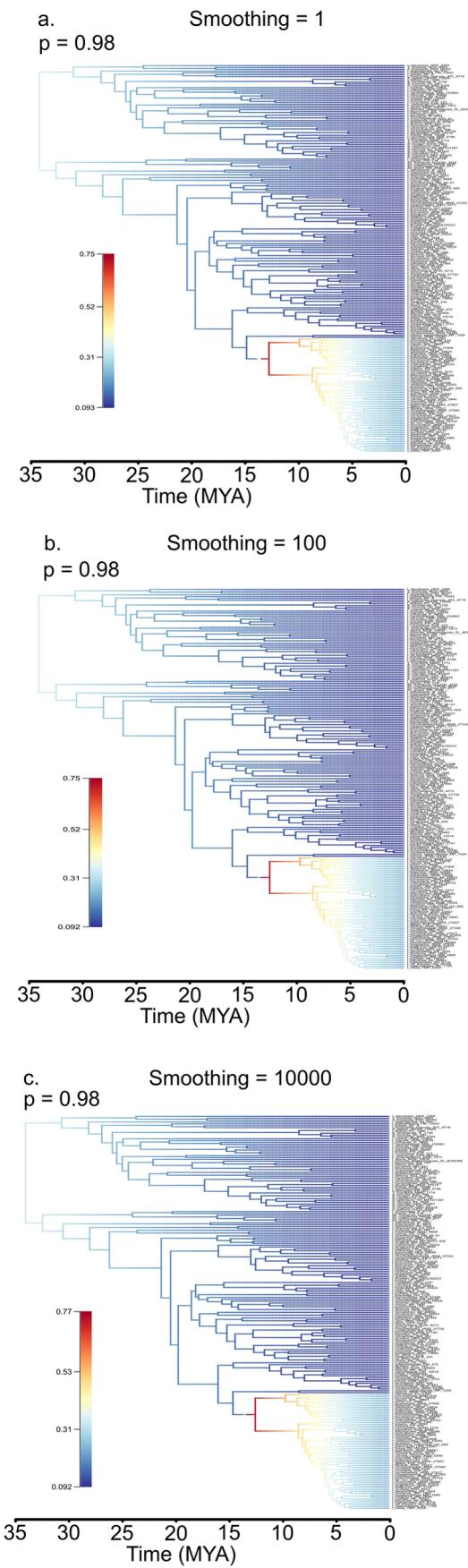


Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Use of ITS phylogeny in the taxonomic process.** We used the *ITS* phylogeny of *Ipomoea* as a single taxonomic character. **a**, We inferred a ML tree that included, when possible, multiple accessions of each putative species and interpreted it in three distinct ways, monophlyy of accessions, non-monophlyy and distantly related, non-monophlyy but closely related. **b**, accessions of a given species are monophyletic (black triangles). Also, two existing species (*I. piurensis* and *I. acanthocarpa*) that we considered conspecific formed a clade of accessions that are intermingled (dashed-line green box), confirming our hypothesis. **c**, accessions of a given species are non-monophyletic and appear in very different parts of the *Ipomoea* phylogeny. In the example shown, our hypothesis based on morphology was a monophyletic *I. squamosa* but the *ITS* tree split the specimens, with some forming a clade in clade D and others forming a clade in clade A which is very distantly related to clade D. In such cases, we re-examined the morphology and usually found a mis-identified specimen(s). In other cases, specimens were similar on herbarium sheets but could be distinguished on closer inspection. In this case, the *ITS* tree alerted us to re-examine and have a closer look at specimens and subsequently describe the new species *I. cryptica*. **d–e**, in some parts of the *ITS* phylogeny there was a lack of resolution, with only a small number of subclades recognised and generally lacking support. Here we show an example with the clade including sweet potato. In such cases, we did not follow the *ITS* phylogeny but relied on morphology and genomic data when available. This decision was based on the fact that **f**, genomic phylogenies for the sweet potato clade demonstrate that nearly all species are monophyletic.

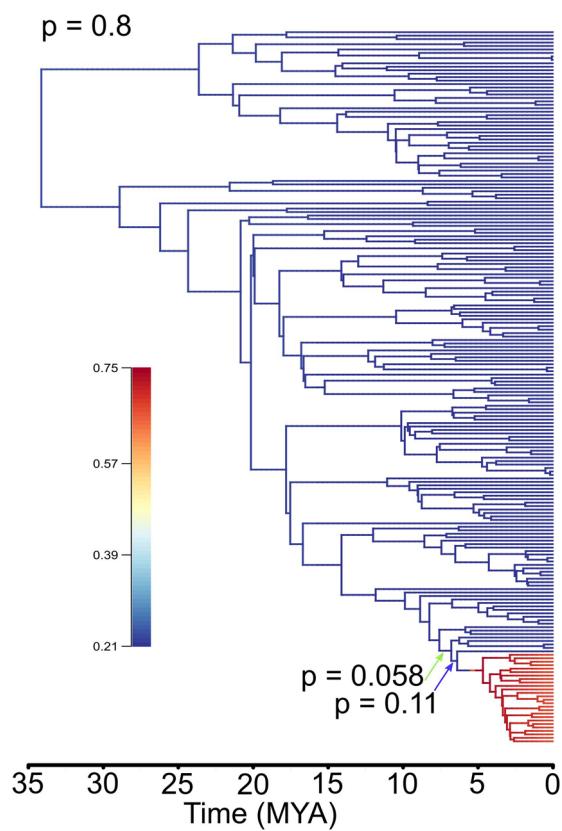


**Extended Data Fig. 3 | Summary genomic phylogenies of *Ipomoea*.** Summary phylogeny of *Ipomoea* inferred from A) 605 putative single-copy nuclear coding regions using Astral-II and B) whole chloroplast genomes using Maximum Likelihood showing the main clades identified in the genus. See Supplementary Discussion, Phylogeny of *Ipomoea*. Complete phylogenies in Supplementary data files 3–7.

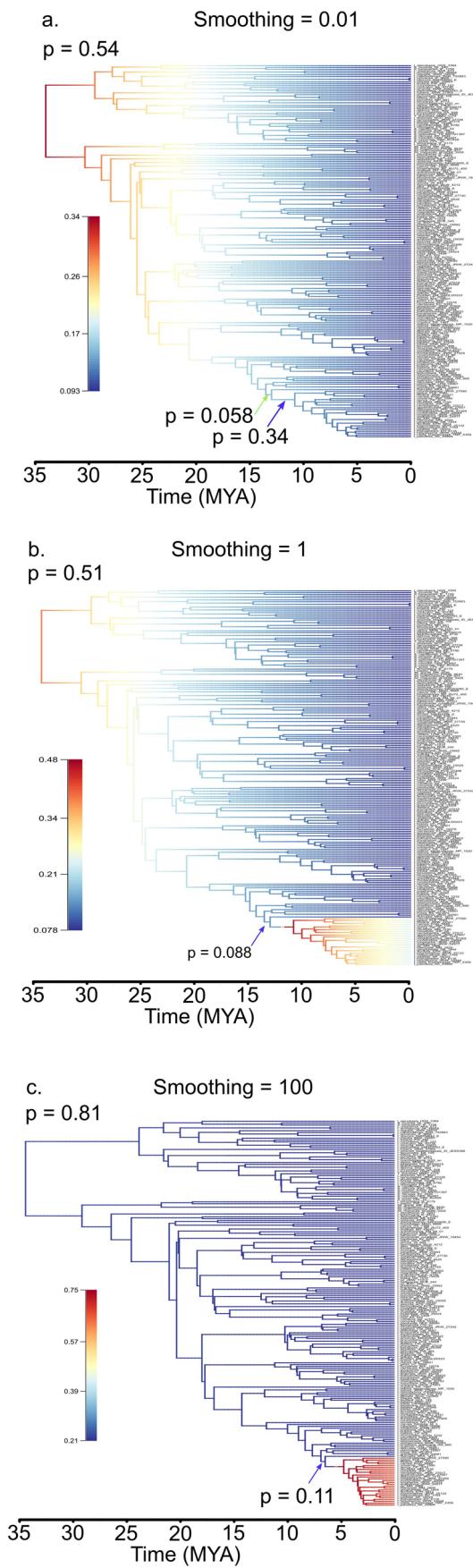


Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | Patterns of diversification-rate-variation in nuclear time-calibrated phylogenies inferred with smoothing values that differ from the optimum.** The coloured tree in each sub-figure represents the pattern of diversification-rate-variation with the highest posterior probability and the posterior probability is shown in the top left corner of each sub-figure. If present, alternative patterns of diversification-rate-variation that have a posterior probability of greater than 0.05 are indicated with coloured arrows, with the position of the arrow indicating the position of the diversification rate shift. Arrows with the same colour indicate a single set of diversification rate shifts that have a specific posterior probability (indicated next to one arrow for each colour). a) A single rate shift at the origin of the diverse South America and Central American clades has a posterior probability of 0.98. b) A single rate shift at the origin of the diverse South America and Central American clades has a posterior probability of 0.98. c) A single rate shift at the origin of the diverse South America and Central American clades has a posterior probability of 0.98.

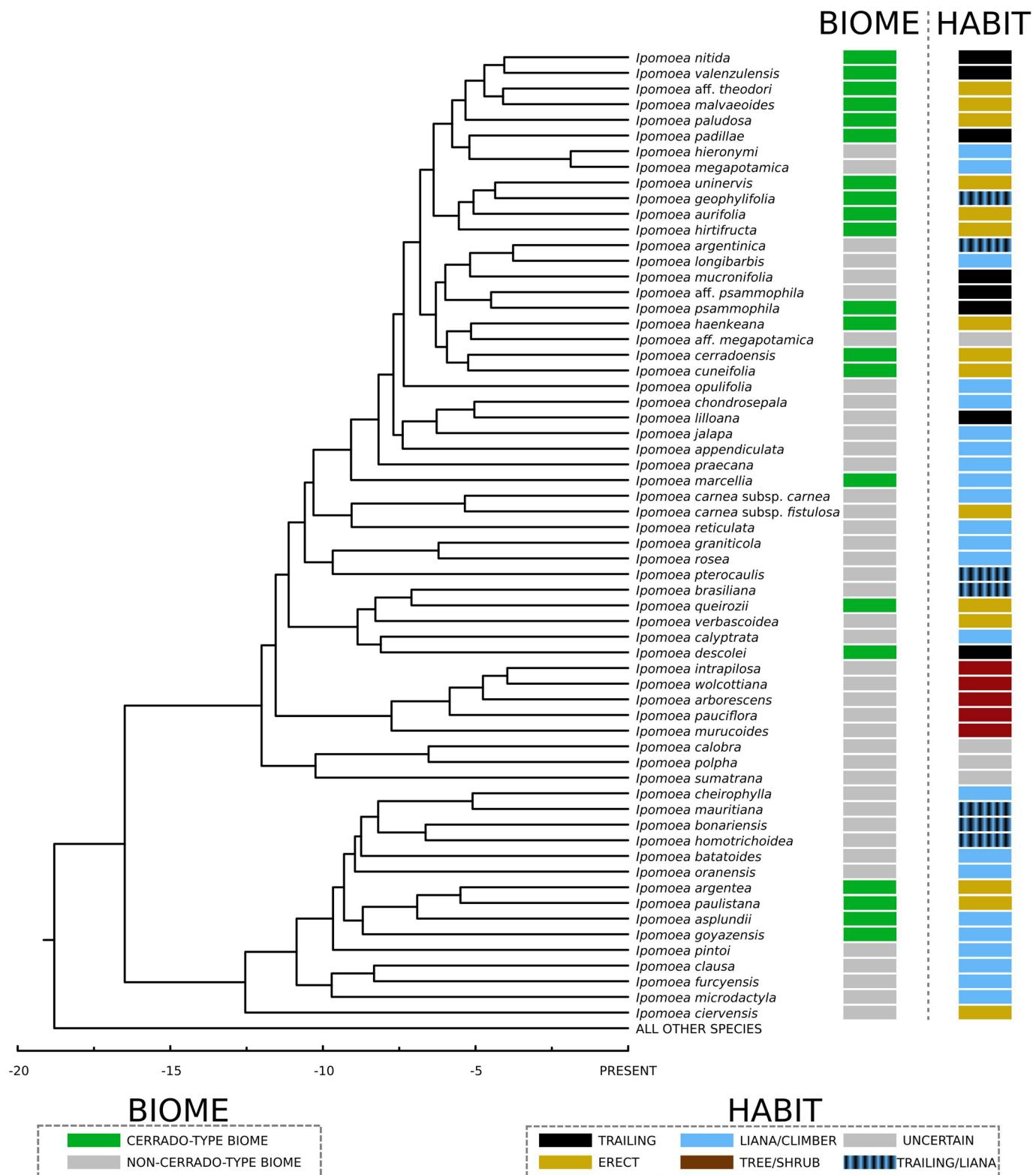


**Extended Data Fig. 5 | Patterns of diversification-rate-variation in the chloroplast time-calibrated phylogeny inferred with the optimum smoothing value.** The coloured tree, posterior probabilities, and coloured arrows are indicated according to the same conventions as Extended Data Fig. 4. A single rate increase near the origin of the diverse South American clade has a posterior probability of 0.8. Two alternative patterns of diversification-rate-variation, where there are rate increases on immediately ancestral branches, have posterior probabilities of 0.11 and 0.058.



Extended Data Fig. 6 | See next page for caption.

**Extended Data Fig. 6 | Patterns of diversification-rate-variation in chloroplast time-calibrated phylogenies inferred with smoothing values that differ from the optimum.** The coloured trees, posterior probabilities, and coloured arrows are indicated according to the same conventions as Extended Data Fig. 4. **a)** A model with no discrete rate shifts has a posterior probability of 0.54. Two alternative patterns of diversification-rate-variation, where there are rate increases near the origin of the diverse South American clade, have posterior probabilities of 0.34 and 0.058 respectively. **b)** A single rate shift near the origin of the diverse South American clade has a posterior probability of 0.51. An alternative pattern of diversification-rate-variation, where there is a rate increase on immediately ancestral branch, has a posterior probability of 0.088. **c)** A single rate shift at the origin of the diverse South American clade has a posterior probability of 0.81. An alternative pattern of diversification-rate-variation, where there is a rate increase on an immediately ancestral branch, has a posterior probability of 0.11.



**Extended Data Fig. 7 | Biome occupancy and growth habit of sampled taxa for the clades with elevated speciation rates.** A section of the time-calibrated phylogeny shown in Fig. 4, for which elevated speciation rates were inferred. Biome occupancy and growth habit of sampled taxa are indicated. This indicates that there are likely to have been multiple shifts into and out of fire-adapted cerrado-type habitats, and multiple shifts between different growth forms amongst recently diverged taxa.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used to collect data.

Data analysis

DNA barcodes obtained using Sanger sequencing at SourceBioscience. High-throughput data obtained using Illumina HiSeq-3000 at the Center for Genome Research and Biocomputing, Oregon State University. All software used for phylogenetic analysis is available online and includes:  
 - DNA barcodes (assembly): Mega v.5.0, Mega v.6.0, Geneious v.9.1.2  
 - DNA barcodes (alignment and phylogenetic inference): MAFFT v.7.2.1, jModelTest 2, RAxML v.8, FastTree 2, MrBayes v.3.2.  
 - High-throughput data (assembly): YASRA, PRICE, SSPACE, BLASTN, MUSCLE, PHYLIP, SPades.  
 - High-throughput data (alignment and phylogenetic inference): MAFFT, PRANK, GBlocks, jModelTest 2, RaxML, FastTree 2, Astral-II, MrBayes v.3.2.  
 - Time-calibrated phylogenies: MAFFT, Gblocks, RevBayes v.1.0.4, Tracer v.1.6.0, treePL, BAMM (including set\_priors.R script), BAMMtools

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Passport data of all herbarium specimens included in the molecular studies presented in this paper is available as supplementary file. Additional records and

information of the collections included in this study and of specimens added subsequently are available through the website of the project (<https://herbaria.plants.ox.ac.uk/bol/ipomoea>). All data used for phylogenetic analysis is available in public repositories: DNA barcode sequences are available through GenBank and genome assemblies are available through the Oxford Repository Archive. Illumina raw reads are available through the Sequence Read Archive.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Monographic study of the genus <i>Ipomoea</i> (morning glories) integrating the morphological study of herbarium specimens with the use of high-throughput DNA sequencing (Hyb-Seq) and DNA barcodes for phylogenetic analysis.
Research sample	All <i>Ipomoea</i> specimens from 47 herbaria worldwide used for morphological study. 385 samples representing c. 210 species sequenced to obtain their whole chloroplast genome and 605 putative single copy nuclear coding DNA regions using Hyb-Seq. 1,560 specimens sequenced for at least one DNA barcode using Sanger sequencing.
Sampling strategy	Sampling strategy aimed at obtaining, when possible, multiple specimens from each putative species studied. When available, material collected in the last 50 years was preferred as this is more suitable to produce good sequencing results.
Data collection	Samples for DNA analysis collected from herbarium material by P.M.R., J.R.I.W., B.R.M.W. and A.S. Morphological study of the specimens conducted by J.R.I.W.
Timing and spatial scale	Project developed between 2012 and 2019 with continuous sampling of herbarium material.
Data exclusions	No data was excluded from the study.
Reproducibility	Methodology is extensively described and all data files, including raw sequence data, are made available through publicly accessible repositories.
Randomization	Does not apply in a study of this type.
Blinding	Does not apply in a study of this type.

Did the study involve field work?     Yes     No

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Involved in the study                                |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines       |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data               |

### Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involved in the study                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |