MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 2000–2014 doi: 10.1111/mec.13584

Population and phylogenomic decomposition via genotyping-by-sequencing in Australian *Pelargonium*

ADRIENNE B. NICOTRA,* CAROLINE CHONG,*† JASON G. BRAGG,* CHONG REN ONG,* NICOLA C. AITKEN,* AARON CHUAH,‡ BRENDAN LEPSCHI§ and JUSTIN O. BOREVITZ*¶

Research School of Biology, Australian National University, Canberra, ACT 2601, Australia, †Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA, ‡Genome Discovery Unit, Australian National University, Canberra, ACT 0200, Australia, §Australian National Herbarium, Centre for Australian National Biodiversity Research, GPO Box 1600, Canberra, ACT 2601, Australia, ¶Centre of Excellence in Plant Energy Biology, Australian National University, ACT 2601, Australia

Abstract

Species delimitation has seen a paradigm shift as increasing accessibility of genomic-scale data enables separation of lineages with convergent morphological traits and the merging of recently diverged ecotypes that have distinguishing characteristics. We inferred the process of lineage formation among Australian species in the widespread and highly variable genus *Pelargonium* by combining phylogenomic and population genomic analyses along with breeding system studies and character analysis. Phylogenomic analysis and population genetic clustering supported seven of the eight currently described species but provided little evidence for differences in genetic structure within the most widely distributed group that containing *P. australe*. In contrast, morphometric analysis detected three deep lineages within Australian *Pelargonium*; with *P. australe* consisting of five previously unrecognized entities occupying separate geographic ranges. The genomic approach enabled elucidation of parallel evolution in some traits formerly used to delineate species, as well as identification of ecotypic morphological differentiation within recognized species. Highly variable morphology and trait convergence each contribute to the discordance between phylogenomic relationships and morphological taxonomy. Data suggest that genetic divergence among species within the Australian *Pelargonium* may result from allopatric speciation while morphological differentiation within and among species may be more strongly driven by environmental differences.

Keywords: genotyping-by-sequencing, morphometrics, *Pelargonium*, phylogenomic, population genomics, reproductive isolation, speciation

Received 26 February 2015; revision received 21 January 2016; accepted 26 January 2016

Introduction

Species are the fundamental units in the fields of biogeography, ecology and conservation biology (Ross & Shoemaker 2005; Hey 2006; Wiens 2007; Light *et al.* 2008; Renema *et al.* 2008; Ross *et al.* 2010). The past decade has seen the emergence of genomic data sets for species identification as well as the development of novel morphological diagnostic methods (Minder & Widmer 2008; Morris *et al.* 2011; McCormack *et al.* 2012). Together, these tools provide potential for insight into the molecular and morphological processes underlying lineage diversification and speciation that have significance to both theoretical and applied questions.

Despite the fundamental nature of species concepts in biology, they remain contentious and decisions about recognition of taxonomic status can be highly challenging. Shared evolutionary descent is the one common point to most species definitions. The general lineage concept of species describes species as segments of separately evolving meta-population lineages (de Queiroz

Correspondence: Adrienne B. Nicotra, Fax: +61 2 6125 5573; E-mail: Adrienne.nicotra@anu.edu.au

1998). This approach provides power to infer the evolutionary history of a study group (Leache *et al.* 2009; Bacon *et al.* 2012). When phenotypic characters are similar among entities, but genomic or genetic data indicate divergence, recognition of convergence can split groups that were previously described as a single species. Alternatively, when morphological variation is found within a single evolutionary lineage, what were separate species can be combined and better described as recognizable intraspecific taxa or ecologically distinct ecotypes. These closely related types then provide opportunity for deeper study of evolutionary processes underlying trait differentiation.

Determining whether ecologically or morphologically distinct entities qualify as independently evolving lineages is particularly difficult in recently diverged or incipient taxa where the genomic differences constituting lineage delimitation may not have fully developed (de Queiroz 2007; Ross *et al.* 2010). Yet, such groups provide potential for insight into the processes that lead to diversification [e.g. the hyperdiverse East African Rift Lake cichlids (Koblmuller *et al.* 2011), or the Hawaiian *Pritchardia* palms (Bacon *et al.* 2012)]. Notably, radiating taxa have a large amount of intraspecific morphological variation, often as a result of phenotypic plasticity (e.g. *Hexalectris* orchids, Kennedy & Watson 2010; or *Stylophora* or *Porites* corals, Flot *et al.* 2011 and Prada *et al.* 2014; respectively). Homoplasy, the parallel evolution of similar traits among phylogenetically dissimilar species may also occur (e.g. *Ophrys* orchids, Devey *et al.* 2008; or *Cladonia* fungi, Pino-Bodas *et al.* 2011; Wake *et al.* 2011). In addition, hybridization can occur between different species lacking complete reproductive boundaries or isolation. This gene flow can lead to lineage introgression, resulting in a range of intermediate forms (e.g. *Haumania*, Marantaceae, Ley & Hardy 2010; or *Pocillopora* corals Pinzón & LaJeunesse 2011). Thus, genomic evaluation along with morphological analysis provides multiple lines of evidence as to the cause and consequence of lineage separation as it occurs, whether morphological differentiation is apparent or not (e.g. Leache *et al.* 2009; Ross *et al.* 2010; Bacon *et al.* 2012).

Using both genomic and morphological tools to delimit species boundaries provides the potential to detect divergences that straddle the population-species boundary. A genomics approach relies on identifying and genotyping a large set of mostly neutral single nucleotide polymorphism (SNP) markers across the genome (Morris *et al.* 2011; McCormack *et al.* 2012; Zellmer *et al.* 2012). For our analyses, we chose to use genotyping-by-sequencing (GBS), which involves reduced representation by enzymatic restriction digestion, followed by barcoded adaptor ligation and amplification. DNA samples from multiple individuals are then combined together in a single sequencing run (Elshire *et al.* 2011; Grabowski *et al.* 2014), which simultaneously evaluates tens to hundreds of thousands of different loci across the genome. This makes it possible to infer genetic relationships in a way that reflects a large sample of the variation among loci in coalescent histories (Escobar *et al.* 2011) and mutation rates (Whittall *et al.* 2010). The GBS approach has proven its effectiveness as a tool to resolve either within or between species lineages in a number of recent population genomic studies (e.g. post-Pleistocene splits in switchgrass, Morris *et al.* 2011; or species relationships within North American Junco birds and Jamaican *Trochilus* hummingbirds, McCormack *et al.* 2012). Done in conjunction with morphological analyses, results can indicate both recent divergence and ancestral convergence in traits and could highlight responses to important ecological pressures.

We focussed our study on Australian *Pelargonium*, an assemblage of recently diverged herbaceous annual and short-lived perennial species that poses a significant taxonomic challenge to botanists. Last revised in the 1960s (Carolin 1961), the group consists of eight taxa (seven currently recognized species plus one entity awaiting description, Fig. 1) within the genus *Pelargonium*. The Australian taxa are geographically disjunct from other *Pelargonium*, a large radiation of over 280 species mostly occurring in southern Africa. It has been suggested that the genus *Pelargonium* arrived in Australia only during the late Pliocene by long-range dispersal and that the current diversity seen within the Australian group has developed within the course of the last 4–5 million years (Bakker *et al.* 1998). However, nuclear ITS and chloroplast trnL-F markers have failed to resolve species relationships among Australian *Pelargonium* taxa (Bakker *et al.* 1998). Rather, taxonomy within Australian *Pelargonium* has been based upon morphology (Knuth, 1912; Carolin 1961), and species delimitation within the group has been a subject of some contention among botanists for well over a century. Distinguishing between the recognized species is difficult due to the large extent of phenotypic plasticity coupled with overlapping variation in leaf shape and growth habit within and between species (Nicotra *et al.* 2007); based on morphology alone, it is unclear whether traits may have evolved multiple times in different lineages or may also still be segregating among them.

Carolin's (1961) revision of the group posits three groups of species: Group I containing the three small-flowered, fibrous-rooted species *P. littorale, P. inodorum* and *P. helmsii,* Group II containing the larger flowered species with more or less perennial stems, *P. australe* and *P. drummondii* and Group III containing the tuberous species *P. rodneyanum* and *P. havlasae* and into which the undescribed species *P.* sp. Striatellum would likely fall.

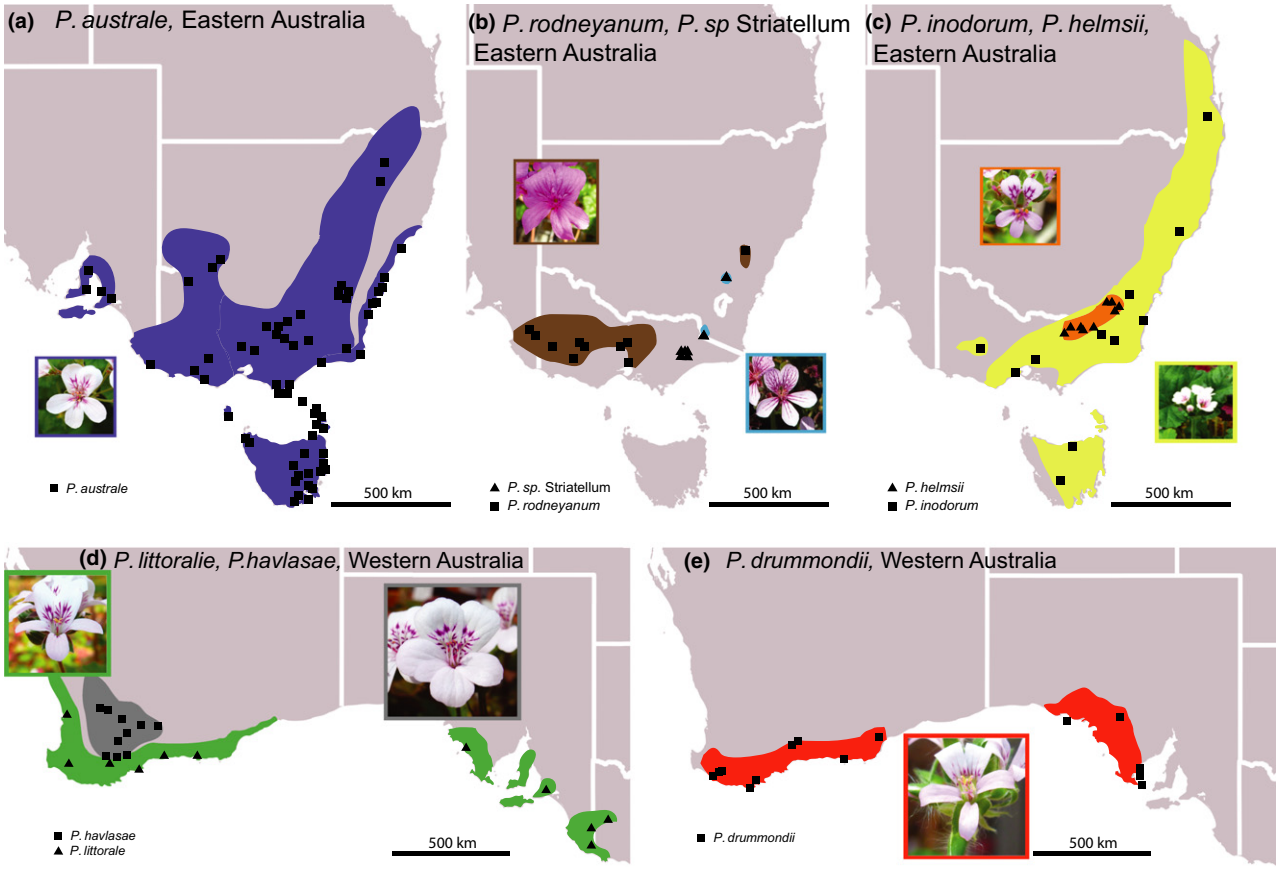

Fig. 1 Distribution maps of Australian *Pelargonium* species inferred from herbarium specimens. Flower sizes are approximately to scale, outline colours match symbols. Relative to Carolin's morphological groupings the species are the small-flowered Group I including (d) *P. littorale* (Panel d, green), *P. inodorum* (Panel c yellow) and *P. helmsii* (Panel c, orange;); the tuberous-rooted and large-flowered Group III including *P. rodneyanum* (Panel b, brown) and *P.* sp Striatellum (Panel b, bright blue), as well as *P. havlasae* (Panel d, grey). The remaining species, *P. drummondii* (Panel e, red) and *P. australe* (Panel a, navy), fall into Carolin's Group II and are characterized by moderate-sized flowers and fibrous roots.

The revision posits that flower size, number of fertile stamens, perenniality of stems and root architecture are diagnostic traits for the group. We assessed these traits and a broad range of others we thought could be of importance.

Here, we describe application of a population and phylogenomic GBS analysis pipeline to span the contentious population/species divide in previously uncharacterized lineages within Australian *Pelargonium*. Species boundaries and affiliations based on Carolin's (1961) revision of the group were used as initial hypotheses and that were tested in two ways: first using our own morphological and reproductive evidence and then using the genomic data. We use our assays of morphological variation and reproductive isolation to observe convergent evolution between lineages as well as character polymorphism within lineages. The demonstration of nonoverlapping patterns in morphological traits is one criterion for delimiting species boundaries (the 'Genotypic Cluster Species Concept', Mallet 2008; Zapata & Jimenez 2012). Morphological discontinuities that reflect the action of evolutionary forces on the phenotype that prevent divergent lineages from homogenizing can lead to the existence of separately evolving taxa (Wiens 2007; Mallet 2008). Traditional inter-cross-compatibility and breeding system studies provide a further complementary approach by which species (particularly of plants) can be distinguished (the Biological Species Concept, Dohzhansky 1935; Mayr 1942; Will & Rubinoff 2004; Schluter 2009). Shifts in breeding systems have occurred thousands of times in angiosperms, often over very short timescales (Foxe *et al.* 2009), and changes in pollination mechanisms are a common means by which reproductive isolation evolves in plants (e.g. Eckardt 2011; Button *et al.* 2012). Nonetheless, many geographically and genetically isolated species, especially plants, remain cross-compatible. Thus, simultaneously assessing information

from morphometric and breeding system studies using the lens of modern genomic tools provides a powerful approach to both reconstruct phylogenetic history and shed insight into selective drivers of diversification.

Materials and methods

Seven putative described species within the Australian *Pelargonium*, and the undescribed *P.* sp. Striatellum (G.W. Carr 10345) (hereafter *P.* sp. Striatellum; see CHAH, 2014) were collected from field sites across their distributions (Interim Biogeographical Regionalisation for Australia framework (IBRA 7), Environment Australia 2000); morphologically distinct types identified from expert knowledge and herbarium samples were also sampled in the field where possible (Fig. 1). A total of 68 populations (29 *P. australe*, 5 *P. drummondii*, 4 *P. inodorum*, 3 *P. helmsii*, 7 *P. rodneyanum*, 3 *P.* sp. Striatellum, 5 *P. havlasae*, 12 *P. littorale*) were collected, which were then used in phylogenomic and morphometric analyses as well as studies of reproductive biology as indicated in Supporting information. Populations were geographically restricted groups of plants. From most populations, 20 individuals (or cuttings thereof) were collected in the field at least 2 m apart and returned to the Australian National University where they were propagated in the plant culture facility under common conditions. Collection sites were a minimum 50 km apart except where two species co-occurred. A subset of these plants representing the breadth of the distributions was used for genetic samples, crossing and autogamy experiments.

Traditional morphometrics and reproductive limits

Morphometric analyses were conducted to test current species boundaries against the criterion of morphological disjunction between lineages. Approximately 5000 herbarium specimens from Australian and New Zealand herbaria were examined prior to the analyses. Complete specimens suited to morphometric analyses were sorted into 12 recognizable entities based on morphology and geography. From these, 151 specimens (approximately 12 per entity) were selected for morphological measurements. Twelve samples were adequate to describe the variation within entity and logistically feasible for measurement of traits. In total, we assessed 38 morphological traits (Tables S1 and S2, Supporting information) on each sample, 14 of which were analysed using principal component analysis (PCA) and principal coordinate analysis (PCoA, for ordinal variables) and cluster techniques using PAST3 (Hammer *et al.* 2001). For more detail on morphometric methods, see Supporting information.

To test the current species delimitation boundaries relative to a hypothesis of reproductive isolation, reciprocal crosses were made between a subset of four species (*P. australe, P. drummondii, P. inodorum* and *P. littorale*). Further, to determine whether a transition from insect pollination to self-pollination has occurred and is associated with flower sizes within Australian *Pelargonium*, patterns of autogamous seed set (i.e. seed set without manual pollination) were assessed (see Supporting information and Table S3 for more detail).

Sequencing and bioinformatic analysis

In total, 192 individuals from 58 collection sites (usually three individuals per site) were selected for genetic analysis (Table S1, Supporting information). Vouchers have been deposited at the Australian National Herbarium in Canberra (CANB, see Supporting information). Approximately 50 mg of leaf tissue was ground using a Qiagen TissueLyser and total genomic DNA was isolated with the Qiagen DNeasy Plant 96 Kit. Two separate GBS libraries, each consisting of 96 samples, were prepared. Briefly, genomic DNA was digested with PstI for genome complexity reduction and DNA was ligated with one of 96 uniquely barcoded sequencing adaptor pairs (Elshire *et al.* 2011; Morris *et al.* 2011). Samples were then individually PCR amplified to avoid sample bias and pooled in an equimolar manner. Library amplicons between 250 and 600 bp were extracted from an agarose gel and sequenced in a HiSeq2000 using a 100-bp paired-end protocol at the Biomolecular Resource Facility at the Australian National University.

Custom perl scripts were used to assign barcode sequences to paired-end reads so that all reads could be used by TASSEL 3.0 UNEAK (Lu *et al.* 2013). SNP calling was performed using the default settings including 5% minor allele frequency and 1 or more reads per sample per locus to make a genotype call. In total, processing of over 142 million useable reads identified 29 531 SNPs. Fifteen samples were omitted because they had insufficient coverage. SNPs were then filtered. Singletons and those SNPs with low call rates (typed as missing in >90 of 177 samples) were removed. Finally, repeats and potential paralogs, where both variants were found too frequently (>100 samples), were removed. The remaining 1463 common SNPs in 177 samples (totalling 192 800 genotypes; 26% missing data, 14% heterozygosity) were used in subsequent analyses.

Genotyping-by-sequencing data were used in three complementary analyses to determine population structure and phylogeny. These are discussed below.

Population structure. The distribution of population genetic groups within *Pelargonium* was evaluated using

model-based individual clustering analysis (STRUCTURE 2.3.2., Pritchard *et al.* 2000). The SNPs were coded using presence of minor alleles (0, 1 or 2) and run using an admixture model with correlated allele frequencies among collection localities. The number of genetic clusters, *K*, was tested from one to 20, and 20 independent runs were computed for each *K* value. The burn-in period was set to 6×10^4 iterations followed by 1×10^5 Markov chain Monte Carlo (MCMC) iterations. Run convergence was assessed visually via parameter distribution (alpha) scores. The most likely number of genetic clusters was determined using the ΔK statistic of Evanno *et al.* (2005) as implemented in Structure Harvester (Earl & Vonholdt 2012) combined with assessment of run consistency using the *clumpp* program version 1.1.2 (Jakobsson & Rosenberg 2007). Output of genetic clustering was subsequently visualized using *distruct* version 1.1 (Rosenberg 2004). Analysis was performed at several different levels of differentiation initially using all samples and then hierarchically delving into smaller structure groups consisting of subgroups of species.

Discriminate analysis of principal components. The genetic relationships among Australian *Pelargonium* were also evaluated and visualized using discriminate analysis (DA) of principal components (DAPC, Jombart 2008; Jombart *et al.* 2010). DAPC begins by performing a PCA to summarize genotypic variation among samples in a relatively small number of uncorrelated variables (PCs). Discriminate analysis is a multivariate procedure that transforms the PCA results to maximize variation among a predefined set of groups and minimize variation within those groups. As with STRUCTURE, for each case, our nominated groups corresponded to species, except for *P. australe*, which was subdivided into five groups, based on location: Tasmanian inland and coastal, mainland coastal, and the small leaf and large leaf mainland groups both of which are found inland (Fig. 1). After each DA, we calculated an '*a*-score', which is a measure of the discriminating power and stability of DAPC, and is obtained via random permutation of group memberships, and comparison of the estimated probabilities of group assignment for the permuted and unpermuted data. Another way to test the robustness of DAPC analyses is with cross-validation, which involves fitting a DAPC to a subset of the data (a 'training' set), and assessing how accurately the model predicts population membership for the samples that were not in the training set. It was not practical to implement a cross-validation of our analyses, because a number of taxa were represented by only a small number of samples. Therefore, we preferred the a-score approach for assessing the stability of our DAPC analyses. In each case, we observed *a*-scores of >0.5 (based on 100 permutations), indicating that our DAPC analyses were not substantially 'overfit'. All DAPC analyses were performed in R (version 3.1.1, R Core Team 2014) using the library adegenet 1.4-2 (Jombart 2008).

Inferring species trees. To identify and understand relationships among the major clades within Australian *Pelargonium*, we built on our Structure and DAPC analyses by estimating a species tree using SNAPP (Bryant *et al.* 2012). SNAPP estimates species tree likelihoods from a set of biallelic genetic markers (assumed to be unlinked) while performing MCMC sampling. Species tree topology, divergence dates and population sizes can then be inferred from the posterior probabilities. SNAPP is computationally intensive, so was performed on a subset of 23 samples. We chose samples with unequivocal group membership and that had relatively few missing data to represent the entities revealed in the STRUCTURE and DAPC analyses (samples are listed in Table S1, Supporting information). Additionally, to minimize missing data for SNAPP analyses (Leache *et al.* 2009), we applied a threshold to identify 463 loci that each had genotype data in at least 22 of the 23 samples. SNAPP was run with parameters $u = 3.0$ and $v = 0.6$, a gamma prior for θ ($\alpha = 11.75$ and $\beta = 200$) and a uniform hyper prior for λ. Model specification follows an analysis of *Ourisia* presented by Bryant *et al.* (2012), with minor modifications based on allele frequencies and the results of trial SNAPP runs. We ran three chains, each of more than 1 000 000 iterations. Each chain was checked for convergence using TRACER v1.5 (Rambaut and Drummond 2009). In general, we observed satisfactory convergence in each chain, but note that several θ parameters exhibited poor convergence (Effective Sample Sizes, or ESS, values of < 100). We proceeded with the analysis of these chains, because the poorly sampled θ parameters were unlikely to affect our estimates of tree topology. We combined the three chains: using logcombiner (BEAST v2.1.3), discarding the first 20% of each chain (burn-in) and yielding a total of 2918 samples. We then used treeannotator (BEAST v2.1.3) to produce a maximum clade credibility tree based on the combined posterior samples.

Results

Traditional morphometrics and reproductive biology

Australian *Pelargonium* species and populations differ in several floral, leaf and stem traits; however, all are diploid ($n = 11$). The three major evolutionary groups defined by Carolin based on architecture, root structure and floral traits (1961) were supported by our

morphometric measures both when we used PCA of 14 continuous morphological traits on herbarium samples (not shown) and when analysed using PCoA cluster analyses (Fig. 2a). The samples from each species were clustered but with considerable overlap in morphospace (Fig. 2b). The analysis also indicated extensive morphological variation within *P. australe* such that several subgroups distinguished both geographically and phenotypically could be identified (Figs 2c, S1, Supporting information).

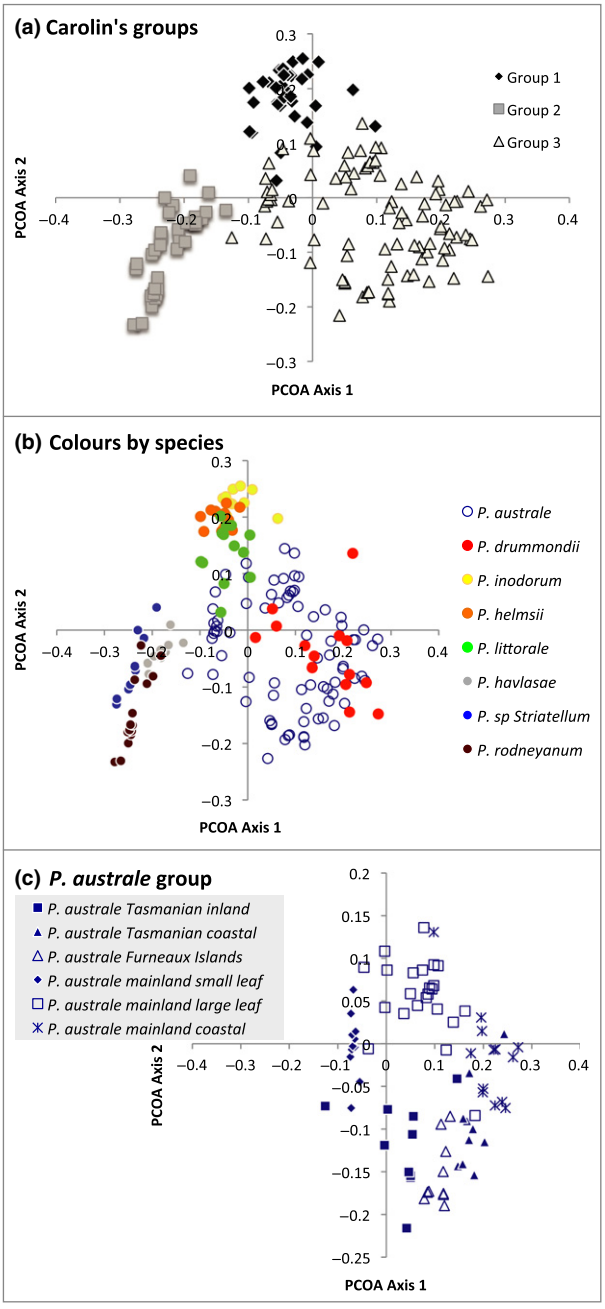

Fig. 2 Distribution of data points in morphospace based on principal coordinates. Axes 1 and 2 explained 22% and 16% of the variance, respectively. (a) Shows that the points can be distinguished according to Carolin's original grouping of the taxa into three main groups (see Fig. 1). (b) Shows the points coloured according to recognized species and demonstrates that species form clusters, but with considerable overlap in morphospace for several. (c) *P. australe* entities alone, showing some separation along axis 1 of leaf size (small to left) and separation along axis two of mainland and Tasmanian/coastal forms.

We also examined self-compatibility and its association with flower size. Flower size measured from herbarium specimens differed significantly between species and again reflected Carolin's groupings with *P. helmsii, inodorum* and *littorale* having significantly smaller flowers than the other species and *P. rodneyanum* being significantly larger (post hoc test based on one-way ANOVA: $F = 58.67$. $P < 0.01$, See Fig. S2a, Supporting information). The percentage of seed set autogamously by flowers differed significantly among species (one-way ANOVA: $F = 87.81$, $P < 0.01$), and species with small flowers (*P. inodorum* and *P. littorale*) were more autogamous than those with larger flowers (note that we did not have sufficient *P. helmsii* flowers for tests of autogamy, see Fig. S2b, Supporting information, for more detail).

In contrast, crossing experiments between *P. australe*, *drummondii*, *inodorum* and *littorale* revealed a lack of barriers to seed production among the Australian species tested with on the order of 20% seed set in between species crosses and no cases of complete cross-incompatibility detected (see Fig. S3, Supporting information). Even crosses between *P. australe* and one of the African species (*P. reniforme*) resulted in apparently viable seed. As such, the morphological results suggest considerable morphological divergence among (and potentially within) taxa in the absence of direct reproductive barriers aside from geographic distance itself. Our analysis does not preclude the potential for other prezygotic barriers (pollinators, phenology) or more subtle post-zygotic issues.

Population and phylogenomic analysis

Population structure. Assignment of all individuals from all geographic localities to genetic clusters using STRUCTURE revealed a maximum modal value of ΔK for $K = 3$ genetic clusters (Fig. 3a) when all samples were included. The proportion of membership of individuals to the three genetic clusters corresponded broadly to the geographic distribution of the samples. Samples from southeastern Australia showed membership predominantly in the two genetic clusters: a major cluster representing *P. australe* samples from coastal and highland regions together with *P. drummondii*, *P. helmsii* and *P. inodorum*, and an additional cluster representing predominantly *P. rodneyanum* and *P.* sp. Striatellum samples. *Pelargonium littorale* and *P. havlasae*

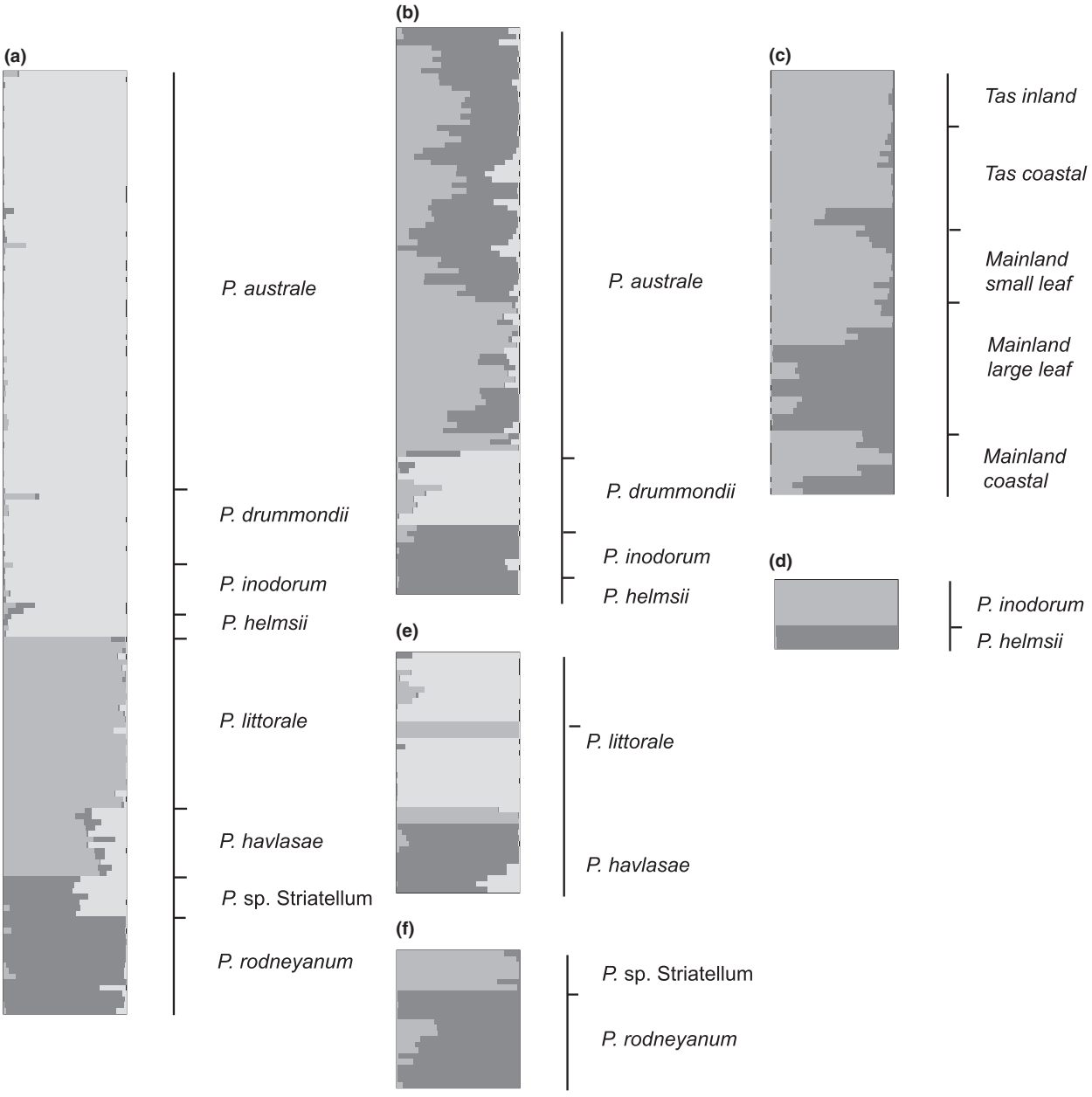

Fig. 3 Hierarchical population structure of *Pelargonium* species and ecotypes. (a) clustering of all samples into $K = 3$ deep groups, (b) focussing on the first group alone to the exclusion of *P. littorale, havlasae,* sp. Striatellum and *rodneyanum*, c) the australe groups alone. The final 3 panels distinguish among species pairs (d) *P. inodorum* from *P. helmsii*, (e) *P. littorale* from *P. havlasae* and (f) *P. rodneyanum* from *P.* sp. Striatellum. Note that shading applies only within a given subfigure and cannot be compared across analyses.

samples constituted a primarily Western Australian genetic group.

Samples within the uppermost cluster in Fig. 3a (i.e. excluding *P. littorale, havlasae,* sp. Striatellum and *rodneyanum*) were next analysed separately and further differentiated into $K = 3$ genetic clusters (Fig. 3b). In this analysis, *P. helmsii* and *P. inodorum* together were assigned to a single ancestral group with strong support (>0.90 posterior probability of assignment). *P. drummondii* individuals were largely assigned to another ancestral group. In contrast, *P. australe* had mixed ancestry, seemingly largely from the *P. helmsii* and *P. inodorum* group. Structure analysis of the *P. australe* samples alone (excluding *P. helmsii*, *P. inodorum* and *P. drummondii*) revealed an optimal ΔK for $K = 2$ genetic clusters (Fig. 3c) with little genomic support for the morphological groupings identified previously. Analysis of *P. inodorum* and *P. helmsii* independently

provides only modest support for their separation into two groups (Fig. 3d).

When the pairs of *P. littorale* and *P. havlasae* (Fig. 3e) and *P. rodneyanum* and *P.* sp. Striatellum (Fig. 3f) were assessed, there was good support for separation within the cluster. In the case of *P. havlasae*, there is some indication that the species may have arisen as a result of a hybridization event between *P. littorale* and a member of the *P. australe* group (presumably *P. drummondii*). Likewise, *P.* sp Striatellum appears likely to be an admixture arising from a hybridization event between *P. rodneyanum* and a member of the *P. australe* group again. However, we note that we have never observed apparent hybrids of these species in the field and that the data cannot be used to infer the timing of such events.

Discriminate analysis of principal components. To complement the STRUCTURE analyses, we performed discriminant analysis of principal components (DAPC) on samples from across the range. At the broadest level of analysis (Fig. 4a), DAPC supported the separation of *P. rodneyanum* and *P.* sp. Striatellum from the remaining taxonomic groups as indicated by STRUCTURE. *P. littorale* was also separated from other *Pelargonium* species, whereas *P. havlasae* exhibited considerable overlap with the remaining species. At this level of analysis, the other groups were not strongly distinguished; however, when we performed DAPC without *P. rodneyanum*, *P.* sp. Striatellum, *P. havlasae* and *P. littorale*, we found strong support for separation of *P. drummondii* from other taxa and modest support for separation among the remaining groups (Fig. 4b). Finally, DAPC restricted to samples of *P. australe* alone was able to detect only a modest level of separation among the five morphologically distinct *P. australe* groups (Fig. 4c).

Phylogenomics. STRUCTURE and DAPC provide a strong indication that the samples form identifiable genetic groups. However, as these methods are not designed to infer the relationships among the groups, we also estimated a phylogenetic tree using SNAPP. Our SNAPP phylogeny (Fig. 5) provided strong support for a clade consisting of *P. littorale* and *P. havlasae* to the exclusion of the other species. Among the other *Pelargonium* species, *P. rodneyanum* and *P.* sp. Striatellum were also placed in a well-supported clade. As with the STRUCTURE analysis, the relationships among the remaining species were not clearly resolved. Although a clade containing *P. helmsii*, *P. inodorum*, *P. australe* and *P. drummondii* could be distinguished, the relationship between *P. drummondii* and the other three species was less clear than in the STRUCTURE analysis.

In summary, the combined results of morphological and genomic analyses supported the initial hypothesis of eight groups, but with a very different evolutionary history from that which was previously described. Incongruence between genetics and morphology was observed in two particularly significant traits that had been used to circumscribe higher species groupings in Carolin's revision (Carolin 1961). Small flowers, characterizing Carolin's Group I, and changes in root architecture, characterizing Carolin's Group III, were each found to have evolved at least twice (Fig. 5). Finally, considering both morphology and genomic results, the *P. australe* clade shows evidence of incipient divergences, particularly in leaf traits and architecture, in most cases coinciding with geographic boundaries, but with little genetic differentiation.

Discussion

Previous phylogenetic studies of the genus *Pelargonium* using limited molecular markers have been unsuccessful in resolving species relationships among Australian members of the genus (Bakker *et al.* 1998, 1999, 2000, 2004). Our results illustrate how a broad bioinformatic approach for both population and phylogenomic data, combined with traditional morphometric data and breeding system studies, can be used to delimit species boundaries and determine character convergence and polymorphism (Emerson *et al.* 2010; Morris *et al.* 2011; Zellmer *et al.* 2012).

Our initial hypotheses of lineage boundaries were based on Carolin's revision of the group (Carolin 1961). And indeed, our morphometric analyses alone would result in recognition of the same three morphologically based groupings proposed by Carolin (1961). However, our genomic analysis demonstrates a history of parallel evolution of key traits among species that does not reflect the previous taxonomy, particularly in that the small-flowered *P. littorale* and *P. inodorum* are not supported as sister species and nor are the tuberous *P. havlasae* and *P. rodneyanum* (Carolin 1961). Reproductive analyses alone would suggest there are no solid barriers to gene flow. Integrating these perspectives, we conclude that the species currently recognized as well as the undescribed *P.* sp. Striatellum warrant recognition as distinct evolutionary lineages, although the small-flowered, high-elevation *P. helmsii* may not warrant species status (see below). In addition, the evidence in this study suggests several ecotypes within *P. australe* that occupy unique geographic ranges and are morphologically easy to diagnose (e.g. the Tasmanian inland small leafed group).

We discuss these results from an integrative perspective considering the multiple lines of evidence we have compiled to assess delineation of taxa. We suggest possible mechanisms driving divergence within the

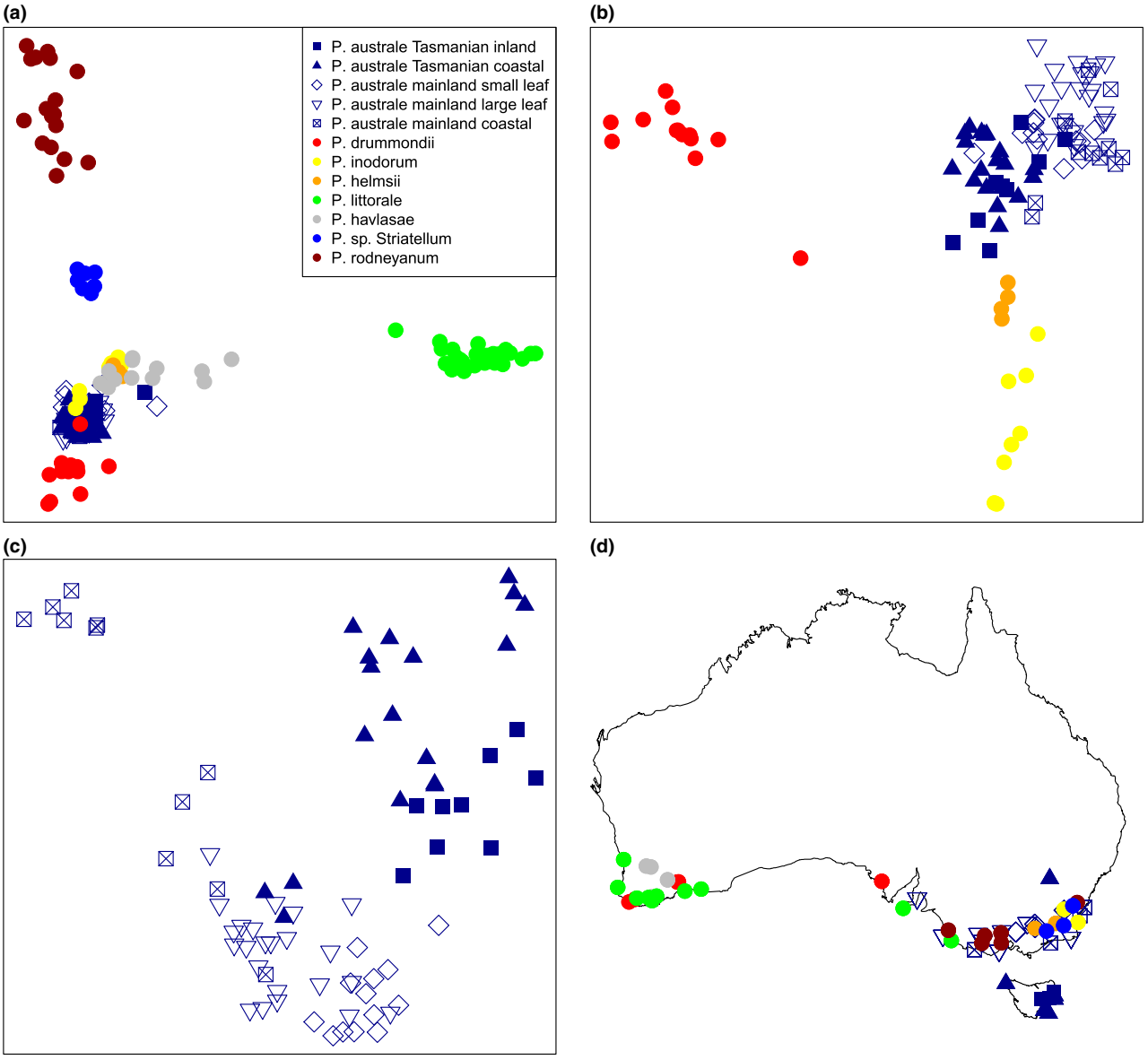

Fig. 4 Discriminant analysis of principal component (DAPC) analyses of genotype data. DAPC was performed using (a) the entire set of samples (as in Fig. 2a; *a*-score = 0.59), (b) a subset of samples excluding *P. littorale, P. havlasae, P. rodneyanum and P.* sp. Striatellum (as in Fig. 2b; *a*-score = 0.67), (c) subset including *P. australe* groups only (as in Fig. 2c; *a*-score = 0.54). Each scatterplot illustrates values of DAPC coordinate 1 (horizontal axis) versus DAPC coordinate 2 (vertical axis). In the three DAPC analyses (panels a, b and c), the first two coordinates account for 79%, 83% and 92% of variance explained (respectively). (d) Map showing locations of all samples with symbols by group. Colours are consistent across plots, see panel (a).

Australian *Pelargonium* that led to currently observed patterns of diversity. This system and analytical approach is an exemplar of how syntheses across multiple lines of evidence can enhance our understanding of evolutionary process in recent speciation events.

Integrating lines of evidence

Ours was a hierarchical approach to applying emerging genomic technologies to investigation of patterns of diversification. At the deepest level, SNAPP allowed us to make inference about the relationships among the major clades. We used STRUCTURE and DAPC to identify clusters across the sampled entities. Together, these approaches provide a robust structure to asses patterns at scales ranging from population to species levels. The SNAPP analysis estimated (with considerable posterior support) that the (*P. rodneyanum*/*P.* sp Striatellum) clade is sister to the (*P. australe*/*P. drummondii*/*P. inodorum*/*P. helmsii*) clade, to the exclusion of the (*P. littorale*/

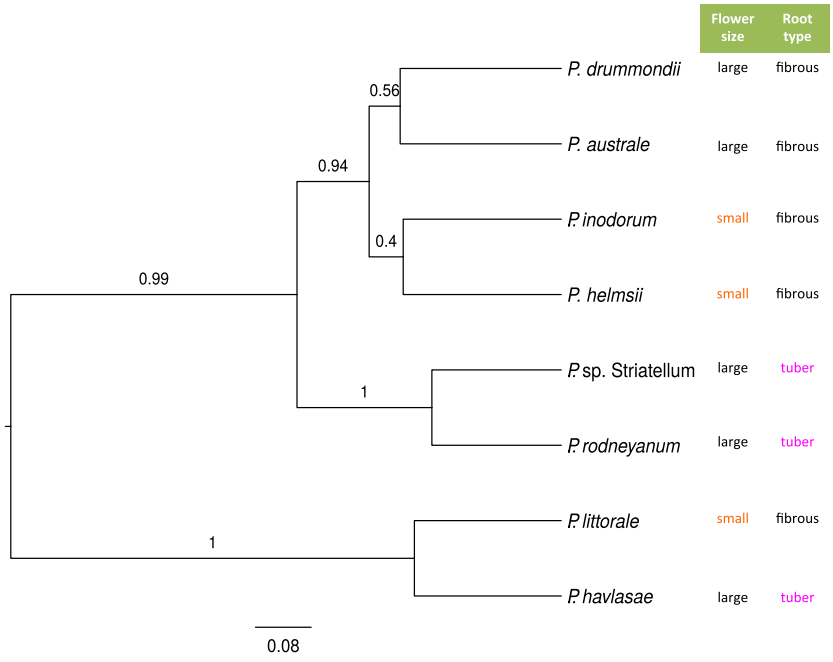

Fig. 5 SNAPP species tree illustrating relationships among the *Pelargonium* taxonomic groups that were identified by STRUCTURE and DAPC. Support values on branches indicate SNAPP posterior probabilities.

P. havlasae) clade. We note that the SNAPP analysis did not resolve relationships among *P. australe, P. drummondii, P. inodorum* and *P. helmsii* (low posterior probabilities, 0.4 and 0.56), but this is not surprising for several reasons. First, the data set used in the SNAPP analyses was a subset of data with few samples per entity thus little power to resolve patterns within groups. These samples were carefully chosen to minimize incidence of missing data while maximizing available loci because previous authors have expressed concern about effects of missing data in SNAPP (Leache *et al.* 2009). Second, the STRUCTURE analysis is more appropriate for assessment of patterns within closely related clades. For example, STRUCTURE identifies clusters within each of the major clades, largely corresponding to the named taxa; however, it also leaves open the possibility of admixture among members of these clusters. If there is contemporary admixture, this would violate an assumption of SNAPP, which attempts to reconstruct the evolutionary history among lineages, based on patterns of incomplete sorting. We therefore were not surprised SNAPP did not resolve relationships among these groups and recommend the combined approach illustrated here as a powerful method for maximizing potential inference from GBS data sets like this one.

Notably, in Australian *Pelargonium*, crossing experiments lend little support for either the morphological or genomic analyses. Increased propensity for self-compatibility or autogamy also appears to have arisen multiple times within the Australian *Pelargonium*. Crossing experiments among a subset of recognized species within the Australian *Pelargonium* displayed a lack of incompatibility between species. The ability to interbreed is commonly observed in recently diverged plant taxa (e.g. Devey *et al.* 2008; McKinnon *et al.* 2010), and our results suggest that cross-incompatibility is not a useful trait by which to distinguish different evolutionary lineages within Australian *Pelargonium*. Indeed, the huge success of the genus as a horticultural enterprise and hobbyist's delight relies on the ease with which distant species can be crossed with simple hand pollination.

Breeding system studies, on the other hand, clearly suggested a distinct difference in pollination mechanisms between the small and large-flowered species on both sides of the continental divide. These may serve as strong forces for prezygotic isolation between species (Widmer *et al.* 2009). The smaller flowered species *P. littorale* and *P. inodorum* were able to autogamously produce large quantities of seed whereas larger flowered species within the group produced little to no seed and thus are likely to be dependent on insects for pollination. Previous work on the group suggests that the section *Peristera*, of which the Australian species are most likely members are commonly self-compatible or autogamous and generally have an ecologically opportunistic habit (e.g. disperse over long distance and recruit after disturbance, Struck 1997). Thus, while full reproductive isolation may not exist within Australian *Pelargonium*, the development of self-pollination within the small-flowered lineages may reduce pollen export from these flowers and encourage selfing within them, effectively achieving reproductive isolation from their closest

large-flowered relatives (*P. littorale* vs. *P. havlasae* and *P. inodorum* vs. *P. australe* (Widmer *et al.* 2009). Shifts in mating systems from outcrossing to self-pollination are very common in many angiosperms (Foxe *et al.* 2009; Goodwillie *et al.* 2010) and are in most cases associated with a vestigialization of traits that attract pollinators, such as conspicuous petals, in other words, a reduction in flower size (Lloyd 1987). Such evolutionary shifts can occur rapidly (Foxe *et al.* 2009) and appear to have done so at least twice in Australian *Pelargonium*.

Demonstration that the tuberous root architecture occurs in two of the major clades suggests that a shift in root form occurred at least twice in Australian *Pelargonium*. The tuberous species in both east and Western Australia are found inland in habitats where water availability is limited and highly seasonal. The Western Australian *P. havlasae* is a spring/summer ephemeral. The eastern species maintain perennial structures aboveground but are highly seasonal in flowering and growth phenology (personal observation). Thus, the tuberous form does potentially reflect an adaption to arid conditions typical of these habitats as postulated by Carolin (1961), but not one that was conserved across the genus.

Finally, our results demonstrated considerable morphological variation, particularly in leaf traits, within the *australe* group that result in several visually distinctive forms that nonetheless cannot be distinguished using genomic data (Figs 2c, and 3c). In the case of the *P. australe* group, leaf morphological and physiological characteristics are also highly plastic and the plasticity itself has been implicated as being adaptive in some traits (Nicotra *et al.* 2007). The differences in mean leaf traits among these entities may likewise have environmental correlates or show signals of local adaptation. Thus, there may be value in recognizing these distinct entities at least as ecotypes (e.g. Moritz 1994), and subsequent investigation of the relationships between their distinctive traits and their environment may reveal correlations between climatic or microenvironmental factors and the morphological differentiation among forms. Similar situations have been observed in the cuticular colour of *Solenopsis* ants (Ross *et al.* 2010) and in body size of tiger snakes (Keogh *et al.* 2005), where the most obvious of morphological traits appears to reflect little of the underlying phylogenetic relationships of the study group.

Drivers of speciation

Homoplasy and morphological variation, both common characteristics encountered in recently diverged taxa, have made Australian *Pelargonium* a taxonomically changing radiation. But the identification of morphologically distinctive types in the absence of significant genomic differentiation at the SNP level actually provides an exciting opportunity to eventually dissect the genetic basis of adaptive evolution at the level of key loci and investigate the selective process driving it. Such morhpo-physiological differentiation may be lead to significant ecological divergence indicative of distinct evolutionary entities, whether called incipient species or ecotypes within species, in the absence of a strong genetic signal in (largely neutral) SNP loci. Discords between morphology and genetics have been observed in many other recently diverged groups (e.g. cichlids, Rüber & Adams 2001; or *Stylophora* and *Pocillorpora* corals, Flot *et al.* 2011 and Pinzón & LaJeunesse 2011; respectively) and are often attributed to differences in mechanisms that drive divergence at each level. In Australian *Pelargonium*, phylogenomic divergence displays clear geographic patterns. Morphological divergence, in contrast, appears to be largely under the influence of environmental factors and may in part also reflect plasticity.

At a broad geographic scale, there appears to be a strong southeast–southwest split between two main clades in the Australian *Pelargonium* (Figs 1 and 5). This deep southeast–southwest phylogenomic split observed in the Australian *Pelargonium* is a common trend in many phylogeographic studies of southern Australian plant and animal taxa (Salinas *et al.* 2009; Ladiges *et al.* 2012) and is highly suggestive of allopatric speciation driving genetic divergence. In plant taxa, such as *Banksia* (Mast & Givnish 2002), *Allocasuarina* (Steane *et al.* 2003) and the pea-flowered Mirbeliae (Orthia *et al.* 2005), the Nullarbor Plain represents a vast, arid calcareous platform serving as an edaphic barrier to gene flow between southwest and southeast Australia, resulting in a phylogeographic pattern consistent to that observed in the Australian *Pelargonium* (Crisp & Cook 2007).

In *P. australe*, the presence of multiple geographically structured lineages (Figs 2 and 3c) is consistent with the distributions of many plant and animal taxa that occur in southeastern Australia (e.g. *Eucalyptus*, Jones *et al.* 2006; and garden skinks Chapple *et al.* 2011). Phylogeographic studies of these taxa suggest that the high-elevation Great Dividing Range as well as the Mt Lofty Ranges in South Australia may provide a range of suitable conditions for mesic-adapted species during arid glacial periods (Byrne 2008; Chapple *et al.* 2011). The current seemingly continuous distribution of *P. australe* may therefore reflect a combination of secondary contact between the clades during warmer periods and specialization along geographic boundaries. Similar patterns of vicariance have been observed within other species that occur in both Tasmania and on the mainland (e.g.

Hardenbergia violacea, Larcombe *et al.* 2011; or *Atherosperma moschatum*, Worth *et al.* 2011).

Taxonomic implications

Taxonomic groupings should reflect evolutionary history (de Queiroz & Gauthier 1992; Cellinese *et al.* 2012). Our phylogenomic analysis has shown that species relationships within the group closely reflect geographic distribution, rather than the current morphology-based groupings of Carolin (1961). We propose that Australian *Pelargonium* should be divided into a Western Australian clade containing *P. littorale* and *P. havlasae* and a southeastern Australian clade containing the remaining species as currently recognized as well as *P.* sp. Striatellum. Furthermore, results indicate that *P. australe* consists of potentially distinct evolutionary lineages. Of particular note are the Tasmanian inland form, which differs in leaf morphology but show little genomic differentiation, and the mainland large leaf form which shows variable morphology but some degree of genetic structure. The Tasmanian inland clade has been recognized at various points in the past, first by Hooker (1834), who described specimens of *P. australe* originating from Tasmania as *P. erodioides*. These populations were subsequently treated at varietal rank within *P. australe* by Bentham (1863) and proposed (but not formally published) at subspecific rank within *P. australe* by Hellbrugge (1997). In addition, we found little evidence of either morphological or genetic differentiation between *P. inodorum* and *P. helmsii*, the latter seemingly being a high-elevation form of the former. Thus, a taxonomic revision of the group with further consideration of structure within the *australe* clade is warranted.

Synthesis and future directions

Australian *Pelargonium* provide a case study of a recent divergence within which a discord between morphology and genetics has resulted in the circumscription of taxonomic groupings that fail to reflect actual evolutionary relationships. The results of this study therefore demonstrate the importance of considering multiple lines of evidence when delimiting species, and stand as a warning against the over reliance on any single line of evidence. By combining genomic, morphological and reproductive studies, we gain power to elucidate both deeper evolutionary splits and recent events, to identify cases of parallel evolution and to develop hypotheses about the selective drivers of diversification within a lineage. The Australian *Pelargonium* remains a fascinating group on which to gain a deeper insight into Australian phylogeography in particular to probe the links between morphological differentiation and fine scale environmental filtering, to explore the adaptive significance of plasticity in those traits and to assess the potential contribution of that plasticity to the diversification process. The results also demonstrate the power of emerging reduced representation libraries as used in GBS approaches to explore patterns of diversification at scales ranging from population to species level and highlight at the same time some of the current limitation in the analytical methods. As these methods develop and their application expands, they will contribute to a greater understanding of diversification and speciation processes, well beyond the Australian *Pelargonium* studied here.

Acknowledgements

We thank Kate Brown, Geoff Carr, Jackie Miles and staff from the many herbaria whose help was absolutely integral to plant hunting. Wei Cheng Tan provided field assistance. Linda Broadhurst and Alexander Schmidt-Lebuhn offered insightful comments on an earlier version of the manuscript. Funding to CRO was provided by the Australian National University and National Parks Board, Singapore. This work was supported by NSF Dimensions of Biodiversity grant 1046328 (JOB & CC) and an ARC Future Fellowship to ABN.

References

Bacon CD, McKenna MJ, Simmons MP, Wagner WL (2012) Evaluating multiple criteria for species delimitation: an empirical example using Hawaiian palms (Arecaceae: Pritchardia). *BMC Evolutionary Biology*, **12**, 23.

Bakker FT, Hellbrugge D, Culham A, Gibby M (1998) Phylogenetic relationships within *Pelargonium* sect. Peristera (Geraniaceae) inferred from nrDNA and cpDNA sequence comparisons. *Plant Systematics and Evolution*, **211**, 273–287.

Bakker FT, Culham A, Daugherty LC, Gibby M (1999) A trnL-F based phylogeny for species of *Pelargonium* (Geraniaceae) with small chromosomes. *Plant Systematics and Evolution*, **216**, 309–324.

Bakker FT, Culham A, Gomez-Martinez R *et al.* (2000) Patterns of nucleotide substitution in angiosperm cpDNA trnL (UAA)-trnF (GAA) regions. *Molecular Biology and Evolution*, **17**, 1146–1155.

Bakker FT, Culham A, Hettiarachi P, Touloumenidou T, Gibby M (2004) Phylogeny of *Pelargonium* (Geraniaceae) based on DNA sequences from three genomes. *Taxon*, **53**, 17–28.

Bentham G (1863) Flora Australiensis.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917–1932.

Button L, Villalobos AL, Dart SR, Eckert CG (2012) Reduced petal size and color associated with transitions from outcrossing to selfing in *Camissoniopsis cheiranthifolia* (Onagraceae). *International Journal of Plant Sciences*, **173**, 251–260.

Byrne M (2008) Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern

Australia inferred from phylogeography. *Quaternary Science Reviews*, **27**, 2576–2585.
Carolin RC (1961) The genus *Pelargonum* L'Hér. ex Ait. in Australia. *Proceedings of the Linnean Society of New South Wales*, **86**, 280–294.
Cellinese N, Baum DA, Mishler BD (2012) Species and phylogenetic nomenclature. *Systematic Biology*, **61**, 885–891.
CHAH (2014) *Australian Plant Census*. Council of Heads of Australasian Herbaria, Australian National Herbarium, Australian National Botanic Gardens and Australian Biological Resources Study. http://www.chah.gov.au/chah/apc/index.html.
Chapple DG, Chapple SNJ, Thompson MB (2011) Biogeographic barriers in south-eastern Australia drive phylogeographic divergence in the garden skink, *Lampropholis guichenoti*. *Journal of Biogeography*, **38**, 1761–1775.
Crisp MD, Cook LG (2007) A congruent molecular signature of vicariance across multiple plant lineages. *Molecular Phylogenetics and Evolution*, **43**, 1106–1117.
Devey DS, Bateman RM, Fay MF, Hawkins JA (2008) Friends or relatives? Phylogenetics and species delimitation in the controversial European orchid Genus *ophrys*. *Annals of Botany*, **101**, 385–402.
Dohzhansky T (1935) A critique of the species concept in biology. *Philosophy of Science*, **2**, 344–355.
Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
Eckardt NA (2011) A sense of self: exploring the selfing syndrome in *Capsella*. *Plant Cell*, **23**, 3086–3086.
Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.
Emerson KJ, Bradshaw WE, Holzapfel CM (2010) Microarrays reveal early transcriptional events during the termination of larval diapause in natural populations of the mosquito, *Wyeomyia smithii*. *PLoS One*, **5**, e9574.
Escobar JS, Scornavacca C, Cenci A *et al.* (2011) Multigenic phylogeny and analysis of tree incongruences in Triticeae (Poaceae). *BMC Evolutionary Biology*, **11**, 181.
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
Flot JF, Blanchot J, Charpy L *et al.* (2011) Incongruence between morphotypes and genetically delimited species in the coral genus *Stylophora*: phenotypic plasticity, morphological convergence, morphological stasis or interspecific hybridization? *BMC Ecology*, **11**, 22.
Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, Wright SI (2009) Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5241–5245.
Goodwillie C, Sargent RD, Eckert CG *et al.* (2010) Correlated evolution of mating system and floral display traits in flowering plants and its implications for the distribution of mating system variation. *New Phytologist*, **185**, 311–321.
Grabowski PP, Morris GP, Casler MD, Borevitz JO (2014) Population genomic variation reveals roles of history, adaptation and ploidy in switchgrass. *Molecular Ecology*, **23**, 4059–4073.
Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, **4**, 9 pp.
Hellbrugge D (1997) *Rievision der Pelargonium Sektion Peristera (Geraniaceae)*. PhD, Westfalische Wilhelms-University Münster, Münster.
Hey J (2006) On the failure of modern species concepts. *Trends in Ecology & Evolution*, **21**, 447–450.
Hooker WJ (1834) Contributions towards a flora of Van Diemen's Land. *The Journal of Botany*, **1**, 252.
Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.
Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.
Jones ME, Shepherd M, Henry RJ, Delves A (2006) Chloroplast DNA variation and population structure in the widespread forest tree, *Eucalyptus grandis*. *Conservation Genetics*, **7**, 691–703.
Kennedy AH, Watson LE (2010) Species delimitations and phylogenetic relationships within the fully myco-heterotrophic Hexalectris (Orchidaceae). *Systematic Botany*, **35**, 64–76.
Keogh JS, Scott IAW, Hayes C (2005) Rapid and repeated origin of insular gigantism and dwarfism in Australian tiger snakes. *Evolution*, **59**, 226–233.
Knuth (1912) Geraniaceae. In: *Das Pflanzenreich* (ed. Engler A), Vol. **IV**(129), pp. 1–640. Engelmann, Leipzig.
Koblmuller S, Salzburger W, Obermuller B, Eigner E, Sturmbauer C, Sefc KM (2011) Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. *Molecular Ecology*, **20**, 2272–2290.
Ladiges PY, Bayly MJ, Nelson G (2012) Searching for ancestral areas and Artifactual Centers of Origin in biogeography: with comment on east-west patterns across southern Australia. *Systematic Biology*, **61**, 703–708.
Larcombe MJ, McKinnon GE, Vaillancourt RE (2011) Genetic evidence for the origins of range disjunctions in the Australian dry sclerophyll plant *Hardenbergia violacea*. *Journal of Biogeography*, **38**, 125–136.
Leache AD, Koo MS, Spencer CL, Papenfuss TJ, Fisher RN, McGuire JA (2009) Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12418–12423.
Ley AC, Hardy OJ (2010) Species delimitation in the Central African herbs *Haumania* (Marantaceae) using georeferenced nuclear and chloroplastic DNA sequences. *Molecular Phylogenetics and Evolution*, **57**, 859–867.
Light JE, Toups MA, Reed DL (2008) What's in a name: the taxonomic status of human head and body lice. *Molecular Phylogenetics and Evolution*, **47**, 1203–1216.
Lloyd DG (1987) Allocations to pollen, seeds and pollination mechanisms in self-fertilizing plants. *Functional Ecology*, **1**, 83–89.

Lu F, Lipka AE, Glaubitz J *et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, **9**, e1003215.

Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2971–2986.

Mast AR, Givnish TJ (2002) Historical biogeography and the origin of stomatal distributions in *Banksia* and *Dryandra* (Proteaceae) based on their cpDNA phylogeny. *American Journal of Botany*, **89**, 1311–1323.

Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.

McCormack JE, Maley JM, Hird SM, Derryberry EP, Graves GR, Brumfield RT (2012) Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular Phylogenetics and Evolution*, **62**, 397–406.

McKinnon GE, Smith JJ, Potts BM (2010) Recurrent nuclear DNA introgression accompanies chloroplast DNA exchange between two eucalypt species. *Molecular Ecology*, **19**, 1367–1380.

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Molecular Ecology*, **17**, 1552–1563.

Moritz C (1994) Defining evolutionarily-significant-units for conservation. *Trends in Ecology & Evolution*, **9**, 373–375.

Morris GP, Grabowski PP, Borevitz JO (2011) Genomic diversity in switchgrass (*Panicum virgatum*): from the continental scale to a dune landscape. *Molecular Ecology*, **20**, 4938–4952.

Nicotra AB, Hermes JP, Jones CS, Schlichting CD (2007) Geographic variation and plasticity to water and nutrients in *Pelargonium australe*. *New Phytologist*, **176**, 136–149.

Orthia LA, Cook LG, Crisp MD (2005) Generic delimitation and phylogenetic uncertainty: an example from a group that has undergone an explosive radiation. *Australian Systematic Botany*, **18**, 41–47.

Pino-Bodas R, Burgaz AR, Martin MP, Lumbsch HT (2011) Phenotypical plasticity and homoplasy complicate species delimitation in the Cladonia gracilis group (Cladoniaceae, *Ascomycota*). *Organisms, Diversity, and Evolution*, **11**, 343–355.

Pinzón JH, LaJeunesse TC (2011) Species delimitation of common reef corals in the genus *Pocillopora* using nucleotide sequence phylogenies, population genetics and symbiosis ecology. *Molecular Ecology*, **20**, 311–325.

Prada C, DeBiasse MB, Neigel JE *et al.* (2014) Genetic species delineation among branching Caribbean Porites corals. *Coral Reefs*, **33**, 1019–1030.

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

de Queiroz K (1998) *Endless Forms: Species and Speciation*. Oxford University Press, Oxford.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.

de Queiroz K, Gauthier J (1992) Phylogenetic taxonomy. *Annual Review of Ecology and Systematics*, **23**, 449–480.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A, Drummond AJ (2009) Tracer v1.5, Available from http://beast.bio.ed.ac.uk/Tracer.

Renema W, Bellwood DR, Braga JC *et al.* (2008) Hopping hotspots: global shifts in marine biodiversity. *Science*, **321**, 654–657.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Ross KG, Shoemaker DD (2005) Species delimitation in native South American fire ants. *Molecular Ecology*, **14**, 3419–3438.

Ross KG, Gotzek D, Ascunce MS, Shoemaker DD (2010) Species delimitation: a case study in a problematic ant taxon. *Systematic Biology*, **59**, 162–184.

Rüber L, Adams DC (2001) Evolutionary convergence of body shape and trophic morphology in cichlids from Lake Tanganyika. *Journal of Evolutionary Biology*, **14**, 325–332.

Salinas M, Bunce M, Cancilla D, Alpers DL, Spencer PBS (2009) Divergent lineages in the heath mouse (*Pseudomys shortridgei*) are indicative of major contraction to geographically isolated refugia on the eastern and western sides of Australia during the early Pleistocene. *Australian Journal of Zoology*, **57**, 41–47.

Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. *Science*, **323**, 737–741.

Steane DA, Wilson KL, Hill RS (2003) Using matK sequence data to unravel the phylogeny of Casuarinaceae. *Molecular Phylogenetics and Evolution*, **28**, 47–59.

Struck M (1997) Floral divergence and convergence in the genus *Pelargonium* (Geraniaceae) in southern Africa: ecological and evolutionary considerations. *Plant Systematics and Evolution*, **208**, 71–97.

Wake DB, Wake MH, Specht CD (2011) Homoplasy: from detecting pattern to determining process and mechanism of evolution. *Science*, **331**, 1032–1035.

Whittall JB, Syring J, Parks M *et al.* (2010) Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology*, **19**, 100–114.

Widmer A, Lexer C, Cozzolino S (2009) Evolution of reproductive isolation in plants. *Heredity*, **102**, 31–38.

Wiens JJ (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, **56**, 875–878.

Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics – The International Journal of the Willi Hennig Society*, **20**, 47–55.

Worth JRP, Marthick JR, Jordan GJ, Vaillancourt RE (2011) Low but structured chloroplast diversity in *Atherosperma moschatum* (Atherospermataceae) suggests bottlenecks in response to the Pleistocene glacials. *Annals of Botany*, **108**, 1247–1256.

Zapata F, Jimenez I (2012) Species delimitation: inferring gaps in morphology across geography. *Systematic Biology*, **61**, 179–194.

Zellmer AJ, Hanes MM, Hird SM, Carstens BC (2012) Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. *Systematic Biology*, **61**, 763–777.

A.B.N., C.R.O., B.L. and J.O.B. designed the study; C.R.O., N.C.A. and A.B.N. collected data; all authors contributed to analysis and writing of the manuscript.

Data accessibility

Data from phenotype analyses, crossing and autogamy experiments as well as accession data have been lodged on DRYAD, doi:10.5061/dryad.31g1b. Genomic data are archived on NCBI (https://www.ncbi.nlm.nih.gov/biosample/4419361 through https://www.ncbi.nlm.nih.gov/biosample/4419542) and SRA (BioProject ID PRJNA309089).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1. Accession details for herbarium samples measured for traditional morphometric.

Table S2. List of continuous and ordinal characters measured for morphometric analysis.

Table S3. Additional detail on methods.

Fig. S1 Results of cluster analyses of morphological data.

Fig. S2 Reproductive traits.

Fig. S3 Percentage seed set in reciprocal crosses.