Shaik, Z., Bergh, N. G., Oxelman, B., & Verboom, G. A. (2023). Genomic, phenotypic, and ecological data provide decisive support for multiple species in the *Seriphium plumosum* complex (Gnaphalieae: asteraceae). *Bulletin of the Society of Systematic Biology*. https://doi.org/10.18061/bssb.v2i2.9000

Investigations

Genomic, Phenotypic, and Ecological Data Provide Decisive Support for Multiple Species in the *Seriphium plumosum* Complex (Gnaphalieae: Asteraceae)

Zaynab Shaik[1,2,3,4,5], Nicola G Bergh[1,2], Bengt Oxelman[4,5], G Anthony Verboom[1]

[1] Department of Biological Sciences, University of Cape Town, [2] Foundational Biodiversity Sciences Division, South African National Biodiversity Institute, [3] Department of Botany and Zoology, Stellenbosch University, [4] Department of Biological and Environmental Sciences, University of Gothenburg, [5] Gothenburg Global Biodiversity Centre

Keywords: Asteraceae, Cape Floristic Region, Genotyping-by-Sequencing, STACEY, Seriphium plumosum.

https://doi.org/10.18061/bssb.v2i2.9000

Bulletin of the Society of Systematic Biology

Abstract

Parallel advances in next-generation sequencing and the development of analytical tools that use sequence data from multiple, unlinked loci have, over the past decade, shifted species discovery and delimitation firmly into the genomic sphere. While most recent analytical applications are underpinned by the Multi-Species Coalescent (MSC), the assumption that species conform to Wright-Fisher populations compromises the ability of this model to resolve species which are loosely knit, metapopulation-like entities that display significant spatial genetic structure. In this paper, we therefore employ a suite of analytical methods (MSC-based species discovery, DISSECT/STACEY; sparse Non-negative Matrix Factorisation, sNMF; discriminant analysis of principal components, DAPC) to thousands of massively parallel-sequenced loci, generated via Genotyping-by-Sequencing, for the purpose of resolving metapopulation-like species in the taxonomically recalcitrant *Seriphium plumosum* L. (Asteraceae; Gnaphalieae) species complex. These methods are complementary in illuminating different characteristics of the entities being resolved (e.g., monophyly, gene pool exclusivity) and so testing different aspects of a species hypothesis. In our study, these methods display remarkable congruence in terms of the groups they reveal, together yielding a set of 12 consensus population clusters which correspond to putative taxonomic species. Using linear discriminant analyses and tests of isolation-by-distance, we further assess whether these entities are phenotypically and ecologically distinct and, where possible, whether they describe a common spatial field of gene exchange. Based on our results, we find evidence of at least eight independent species in the *S. plumosum* complex, most of which are not truly cryptic, and at least five of which maintain their genetic distinctness in sympatry and so qualify as biological species. Our findings corroborate previous work suggesting a need for diverse tools to properly circumscribe the species-level diversity of South Africa's Cape flora.

INTRODUCTION

Building on a long history of conceptual development (e.g., Darwin, 1859; de Queiroz, 1998, 2007; Dobzhansky, 1937; Freudenstein et al., 2017; Wiley, 1978), species delimitation has recently moved into the genomic sphere with the development of methods which use sequence data, commonly from multiple putatively unlinked DNA loci, to infer species boundaries. Most recent applications are underpinned by the Multi-Species Coalescent model (MSC; Grummer et al., 2014; Jones, 2016; Jones et al., 2015; Leaché et al., 2014; O'Meara et al., 2006; Rannala & Yang, 2017; Z. Yang & Rannala, 2014; Ziheng Yang, 2015; Ziheng Yang & Rannala, 2010), but this approach has been criticized for resolving populations rather than species (Sukumaran & Knowles, 2017), a consequence of its assumption that species conform to the Wright-Fisher model (Fisher, 1930; Wright, 1931). This is a particular problem in species which are widespread or whose spatial scale of gene flow is insufficient to fulfil the assumption of panmixia (Chambers & Hillis, 2020; Mason et al., 2020). While such species are expected to exhibit significant spatial genetic structure, the frequent movement of genes between their component populations should nonetheless render them genomically discernible, albeit using a different suite of analytical tools.

Certainly, such metapopulation-like entities qualify as species when viewed through the lens of the evolutionary/unified species concept (de Queiroz, 1998, 2007; Wiley, 1978).

The resolution of genetically less-cohesive, metapopulation-like entities requires an approach whose assumptions are less restrictive and/or which allows for the discovery of entities that are more polythetic in nature. Population structure analysis and ordination are two such approaches. In the case of population structure analysis, as implemented in STRUCTURE (Pritchard et al., 2000) and analogous software (ADMIXTURE: Alexander et al., 2009; Tang et al., 2005; fastSTRUCTURE: Raj et al., 2014; Snapclust; Beugin et al., 2018; Sparse Non-negative Matrix Factorisation: Frichot et al., 2014), the discovery of spatially structured, polythetic population clusters is enabled by its resolution of gene pools which can span multiple geographical populations and, at least in the context of an admixture model, be multiply represented within a single individual. Thus, although the individual gene pools are constructed to maximise fit to the assumptions of Hardy-Weinberg and linkage equilibrium (e.g., Bourgeois & Warren, 2021), the possibility of multiple gene pools being represented in a single individual allows for geographical populations to be grouped into more inclusive metapopulation entities based on genetic compositional overlap, with the resulting groups often being polythetic in terms of their gene pool composition (e.g., Gratton et al., 2016; Leaché et al., 2019; Medrano et al., 2014). In contrast to population structure analysis, ordination approaches, like that implemented in discriminant analysis of principal components (DAPC), resolve genetic clusters in the absence of an explicit biological model (Jombart et al., 2010). As such, they are able to resolve somewhat diffuse entities while at the same time being computationally highly efficient. Although there is no guarantee that the population groups retrieved using ordination methods and population structure analyses will correspond either to each other or to clades in a population tree (Carstens et al., 2013), this may commonly be the case. Thus, in concert with MSC-based delimitation approaches, these methods may be able to resolve putatively monophyletic metapopulation-like entities. In this paper we use this approach to identify such entities and argue for their taxonomic recognition at the species rank. In the rest of this paper, we use the term "species" with this meaning.

Another way in which genomic data can be used to test species limits involves assessing patterns of isolation-by-distance (IBD; Wright, 1943, 1946) which have predictable patterns within and across species (Edwards & Knowles, 2014; Gratton et al., 2016; Hausdorf & Hennig, 2020; Medrano et al., 2014). This approach is premised on the logic that ongoing, spatially restricted gene flow should produce a signature of IBD between conspecific populations but that, owing to the existence of isolating mechanisms, such a signature should not be detectable across species boundaries. Accordingly, several studies have tested species limits using signatures of IBD (e.g., Bamberger et al., 2021; Gratton et al., 2016; Medrano et al., 2014; Spriggs et al., 2019). With the statistical validity of permutation-based Mantel and partial Mantel tests of IBD under debate (Frantz et al., 2009; Guillot & Rousset, 2013), Hausdorf and Hennig (2020) developed an explicit, regression-based operational framework for species delimitation. As their regression-based tests require that candidate species be defined *a priori*, they are invariably applied in conjunction with other analytical methods, including genetic clustering approaches. Crucially, since IBD patterns are explicitly spatial, this test is logically decoupled from other methods.

Species delimitation using genomic data has the potential to resolve species which are phenotypically discernible or which show substantial overlap in their morphological character variation. While some workers have vigorously defended the evolutionary reality of such "cryptic species" (e.g., Cotterill et al., 2014), noting that convergent evolution can produce indistinguishable phenotypes in independent lineages via different pathways (e.g., Lim et al., 2019), others have argued for the importance of distinct biological role and/or phenotype as a crucial and necessary component of the species concept (e.g., Freudenstein et al., 2017). Under the latter view, truly cryptic entities, even those that are evolutionarily distinct, do not warrant formal taxonomic status on the basis that "biodiversity is inherently a phenotypic concept" whose "critical value... lies in the myriad roles that organisms exhibit" (Freudenstein et al., 2017). Even allowing that this position is extreme, there is little doubt that phenotypic diagnosability yields taxonomies which are both more practicable and less prone to "taxonomic inflation" (Fujita et al., 2012). In addition, phenotypic and ecological differentiation provide valuable corroboration for the evolutionary distinctness of genomic entities, an idea central to integrative taxonomy (Dayrat, 2005; Fujita et al., 2012; Noguerales et al., 2018; Padial et al., 2009; Solís-Lemus et al., 2014). In any event, the question of whether to recognize cryptic species may be irrelevant since closer inspection may reveal some degree of phenotypic difference in many supposedly cryptic species (e.g., Korshunova et al., 2019; Mann & Evans, 2008). As noted by Struck et al. (2018), the designation of crypsis in species generally lacks rigour, with almost half of recent claims of species crypsis being unsupported by phenotypic evidence.

The Cape Floristic Region (CFR; Goldblatt, 1978) of South Africa displays an exceptional prevalence of range-restricted vascular plant species (Cowling et al., 1992; Cowling & McDonald, 1998). This is probably a consequence of the strong spatial heterogeneity in topography, climate, and soils of the region, as well as its long-term climatic stability (Goldblatt, 1978, 1997; Goldblatt & Manning, 2002; Schnitzler et al., 2011) and these factors acting in concert to favour the evolution of relatively short-distance seed dispersal syndromes (Goldblatt, 1997; Latimer et al., 2005, 2009; Linder, 1985; Slingsby & Bond, 1985). Widespread species are thus comparatively unusual in the Cape flora, and it has been suggested that some widespread plant species comprise multiple cryptic, semi-cryptic, or pseudo-cryptic species (e.g., Britton et al., 2014; Prunier & Holsinger, 2010). The Composite shrub *Seriphium plumosum* L. (Asteraceae, Gnaphalieae; *Seriphium* hereafter abbreviated as "*S.*") is one of the most ubiquitous and wide-

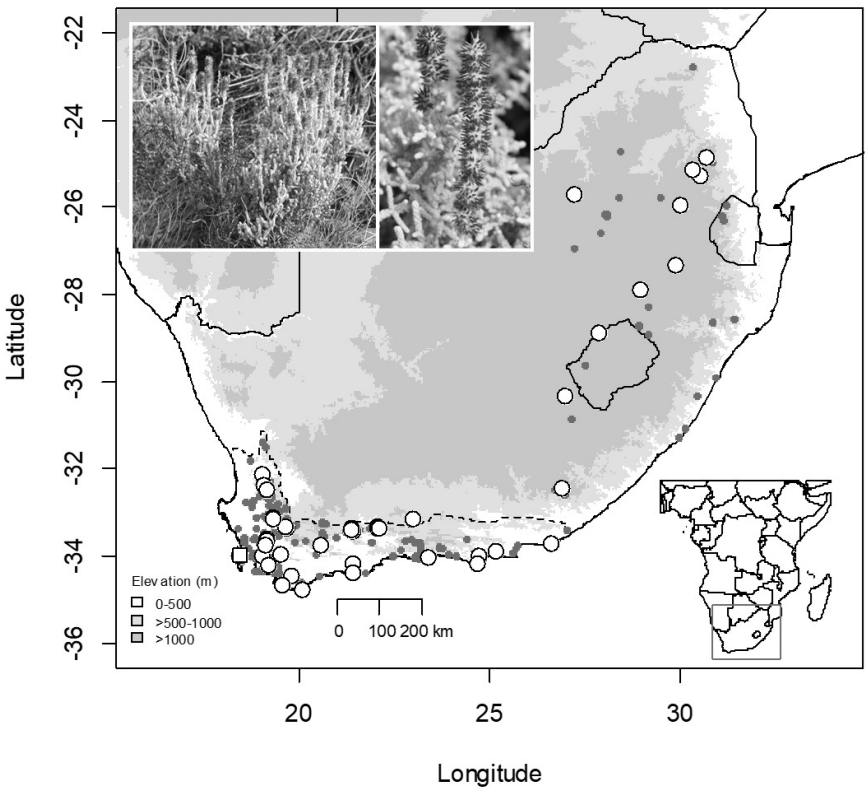

Figure 1. Map showing *Seriphium cinereum* (open square) and *Seriphium plumosum* sampling localities (open circles) with the background distribution (filled circles) of *S. plumosum* in South Africa, Lesotho, and Eswatini based on accessions in the Bolus (BOL) and Compton (NBG) herbaria.

Inset photographs show the whole plant and synflorescence of *S. plumosum,* in this case a high-elevation form sampled from Conical Peak in the southwestern Cape at 1900 m elevation.

spread plant species in the CFR, being distributed across the region and beyond to the north-eastern corner of South Africa (Fig. 1) with seldom-documented outlying populations in the highlands of the Northern Cape, Namibia, Angola, and Zimbabwe. Across this large distribution, the species occupies a range of elevations (sea level to ca. 2000 m) and spans an ecologically important winter-summer rainfall boundary. It also displays considerable, albeit subtle, variation in growth form, structure, and arrangement of leaves, synflorescence structure, and flowering time. Consequently, there is reason to suspect that *S. plumosum* comprises multiple species, a suspicion which is reinforced by its historical segregation into three species (Levyns, 1937) under the generic name *Stoebe* (hereafter abbreviated as "*St.*"), namely *St. plumosa* (L.) Thunb., *St. burchellii* Levyns, and *St. vulgaris* Levyns. Koekemoer (2016), however, merged these species under *S. plumosum* based on the difficulty of distinguishing "well-defined forms" using morphology; although, she stated that the full variation of the complex remained unexplored.

The central aim of this study is to use a combination of genomic, morphological, and ecological data to assess the hypothesis that *S. plumosum* comprises multiple species, to identify and characterise any such species, and to assess morphological crypsis in such entities. Our protocol involves four steps which, together, constitute an integrative approach to species delimitation. These are:

1. To delimit population-like entities and to infer their interrelationships in an MSC framework using the Bayesian DISSECT discovery method (Jones et al., 2015) with improvements as implemented in STACEY (Jones, 2016).
2. To identify more inclusive, metapopulation-like entities that potentially correspond to species using DAPC (Jombart et al., 2010; Jombart & Ahmed, 2011) and the STRUCTURE-like method sparse Non-negative Matrix Factorisation (sNMF; Frichot et al., 2014).
3. To assess patterns of IBD within and across closely related putative species following the regression-based tests of Hausdorf and Hennig (2020).
4. To assess whether the putative species identified using the above methods are phenotypically and/or ecologically differentiated using phenotypic and environmental data from the same specimens/localities used to obtain the genomic data.

Our overall objective is to resolve evolutionarily meaningful species-level entities within the *S. plumosum* complex, which will be taxonomically formalized in a forthcoming sister paper.

MATERIALS & METHODS

Sampling

Populations representing all previously recognised species and the full range of geographical and morphological variation in the *S. plumosum* complex were sampled in South Africa between April 2017 and April 2018 from 47 localities (Fig. 1; Supplementary data). A single population of *S. cinereum* L., the putative sister species of *S. plumosum* based on phylogenetic analysis of ITS and plastid Sanger sequences (N.G. Bergh, personal communication), was also sampled near Cape Town (Fig. 1). A single 20 cm branch bearing leaves and flowers was sampled from six individuals at each locality. Sampled plants were spaced >5 m apart to minimise spatial autocorrelation. Short (ca. 5 cm) leaf cuttings were collected from each branch cutting and stored on silica-gel for DNA isolation. A single flowering branch (where present) was also fixed in FAA (60% ethanol, 25% distilled water, 10% formalin, 5% glacial acetic acid) solution. Population vouchers were deposited at the Compton Herbarium, Kirstenbosch, South Africa (NBG).

Molecular data assembly

DNA isolation, sequencing. – For each of the 288 sampled accessions, 100 µg of silica-dried leaf material was pulverised in liquid nitrogen. Total DNA was isolated from the pulverised material using the DNEasy plant mini kit (*Qiagen,* Venlo, Netherlands) with final elution in 30 µl TE buffer to achieve concentrations of ⩾50 ng/µl, with 260/280 absorbance ratios between 1.8 and 2.0. Isolates were stored at -80 °C before being express-shipped to Novogene Genome Sequencing Company Ltd. (Beijing, China) for Genotyping-by-Sequencing (GBS). Here, isolates were fragmented with restriction enzymes MseI, HaeIII, and MspI, and adaptor-ligated fragment lengths assayed on an Agilent 2100 bioanalyzer (*Agilent, California, United States*). Fragments approximately 350 base-pairs (bp) long were amplified by quantitative real-time PCR. Size-selected fragments were then paired-end sequenced on an Illumina HiSeq PE150 (*Illumina, Inc*.) platform.

Assembly of biallelic SNPs. – Raw sequence data were filtered to exclude paired reads with adaptor contamination, >10% ambiguously called nucleotides, and paired reads where >50% of either read comprised low quality nucleotides (⩽Q5). Filtered reads were used to assemble a *de novo* reference genome based on all 288 accessions using SOAPdenovo 2.04 (Luo et al., 2012) with parameter settings [-K 41 -R -d 2 -p 10] and minimum quality thresholds; (i) contigs (consensus sequences) N50 ⩾30 kb and (ii) scaffolds N50 ⩾1 Mb. Following sequencing, *de novo* reference genome assembly, and preliminary filtering, we generated a single concatenated variant data set across all 288 accessions by quality-trimming and mapping reads to the *de novo* reference, calling variant sites, and filtering the variant sites using Puritz et al.'s (2014) dDocent pipeline. Quality trimming of cleaned reads was performed using Trimmomatic (Bolger et al., 2014), which removed any remaining adaptor contamination and low-quality bases (<Q20, or error rate 1%) from read ends. Trimmed reads were mapped to the *de novo* reference genome using the MEM algorithm of the Burrows-Wheeler Aligner, the preferred algorithm for low-divergent sequences of greater than 70 bp long (Li & Durbin, 2010), with default conservative mapping parameter settings [-A 1 -B 4 -O 6]. Variant calling was performed in parallel across 64 processing cores with 128 GB of memory (University of Cape Town ICTS High Performance Computing Centre) using FreeBayes 1.2.0 (Garrison & Marth, 2012). To minimise linkage among SNPs, one SNP per read was retained. Filtering of called SNPs and export of a complete variant call file (VCF) were performed using VCFtools 3.0 (Danecek et al., 2011).

Assembly of full sequences. – Full sequence alignments for DISSECT/STACEY (STACEY hereafter) were assembled by backreferencing typed biallelic SNPs to their mapped source reads using a custom R 3.5.1 (R Core Development Team, 2018) script available on GitHub (https://github.com/zaynabthebotanist/SNP-to-alignments) which: (i) queries the identity of the unique scaffold to which each short read containing a SNP was aligned; (ii) uses the scaffold identifiers to subset filtered reads in the binary alignment/map (BAM) files; and (iii) uses the filtered BAM reads to assemble two haplotype sequences per individual at each locus, assuming all individuals are diploid (Semple & Watanabe, 2009; Vallès et al., 2013). This assumption is supported by the low level of genome size variation apparent in the complex (P. Travnicek et al., personal communication) and by the closely related *S. kilimandscharicum* being a confirmed diploid (Hedberg & Hedberg, 1977). For reasons of computational tractability (for STACEY analysis), sequence alignments were assembled for a subset of 36 accessions, each representing a single accession per locality across the full distribution of *S. plumosum*. Individual BAM files for the 36 selected accessions were filtered using the R package *ssviz* 1.16.1 (Low, 2019) to retain reads containing biallelic SNPs. Mapped reads were aligned using the R package *muscle* 3.26.0 (Edgar, 2004) and haplotyped using the R package *pegas* 0.12 (Paradis, 2010). Alignments with third and fourth haplotypes of read depth ⩾3 were discarded as putative multicopy loci, while loci with either one or two haplotypes of read depth ⩾3 were retained as orthologous (Andermann et al., 2019). Paired haplotypes were sorted by scaffold identifier, producing alignments containing up to 72 sequences (2 x 36 accessions). As STACEY requires full accession representation across all loci, where both haplotypes for an accession were absent at a locus on account of having >2 haplotypes of depth ⩾3, or where only one haplotype of depth ⩾3 was sequenced (either because of homozygosity or failed sequence capture of a second haplotype), blank sequences matching the alignment length were inserted. Overall, however, mean coverage was high (68.92/72), and there were few or no missing sequences per alignment. To further exclude non-orthologues, the absolute number of parsimony-informative sites across all samples was computed for each locus using *ips* 0.0.11 (Heibl, 2008) as implemented in R, and alignments containing >3 times the median number (n = 6 for our data) of parsimony-informative sites discarded. Despite the conservative settings

used when mapping reads to the *de novo* reference genome, putative non-orthologues (i.e., highly divergent haplotypes within individuals) remained in the resulting alignments. To detect these, we constructed UPGMA trees (Sokal & Michener, 1958) for each locus in the R package *phangorn* 2.5.5 (Schliep, 2011) and discarded loci with ⩾3 individuals whose haplotypes were separated across the root nodes. All file conversions in the R environment were performed using the R package *Biostrings* 2.50.2 (Pagès et al., 2019). Final alignments were converted to FASTA format in the R package *seqinr* 3.6.1 (Charif & Lobry, 2007).

STACEY analysis (sequence data)

Phylogenetic relationships and species limits were jointly estimated on the assembled haplotype alignments using STACEY (Jones, 2016), a method based on a modified birth/death model (Jones et al., 2015) under which divergence times associated with very recent splits are approximated as zero. This allows for the identification of minimal clusters of individuals which correspond to Wright-Fisher species as in the MSC (Heled & Drummond, 2010; Jones, 2016) where gene tree discord is explicitly modelled. While two-step summary tree inference methods (e.g., ASTRAL; Mirarab et al., 2014) may be more computationally efficient than STACEY and other full-likelihood approaches, they ignore error in gene tree estimation, which may be particularly high where divergence events occur in rapid succession and the number of phylogenetically informative sites is consequently low (Cai et al., 2021). Thus we elected not to apply these methods here.

Since the STACEY template used to assemble XML files for BEAST analysis in BEAUTi 2.5.0, Java 1.8.0_91 (Bouckaert et al., 2014) failed when hundreds of loci were imported, XML files were assembled in R using a custom script available on GitHub (https://github.com/zaynabtheb-otanist/make-large-DISSECT-xml). Each accession was defined *a priori* as a minimal cluster, except for population 32 (*S. cinereum*) whose three representative accessions were defined as a single minimal cluster. Since each accession was represented by two haplotypes, the number of sequences per locus and minimal cluster ranged from two to six. STACEY was implemented in BEAST 2 version 2.6.0 with STACEY 1.2.5 (Bouckaert et al., 2014). The substitution model of Jukes and Cantor (1969) and a strict molecular clock were used, with the substitution rate of the first (arbitrarily selected) alignment set to one and those of the remaining alignments estimated relative to this. Collapse height (ε) was set to 10^{-5} (Jones et al., 2015). Priors on the clock rates were all drawn from independent lognormal distributions with log mean and log standard deviation set to 0 and 1.25 respectively; hereafter written as lognormal (0, 1.25). The growth rate prior was drawn from a lognormal (5, 2) distribution. The population size prior was lognormal (-7, 2), and death rate and collapse weights were uniformly distributed using a $\beta(1,1)$ distribution. We ran six parallel MCMC analyses for several billion generations with 32 GB per CPU allocated to run on resources provided by UNINETT Sigma2, the National Infrastructure for High Performance Computing and Data Storage in Norway (Uninett Sigma2, 2020) on the "Saga" cluster, sampling parameters and trees every 500 000th generation.

Because a preliminary analysis (Analysis 1) yielded root node depths for some gene trees that were one to two orders of magnitude greater than that of the species tree, the alignments were inspected manually for sources of error. We identified stretches of poorly aligned sequences in 24 of the 524 assembled alignments, usually as flanking, but strongly conflicting sequences. These sections of sequence were therefore replaced with blank sequences and the MCMC repeated (Analyses 2a-2d). Since neither Analysis 1 nor Analyses 2a-2d converged or showed sufficient mixing, with estimates of effective sample size (ESS) for the posterior, likelihood, smcCoalescent and TreeHeight falling in the range 50-100, we also performed a less parameter-rich analysis on the corrected alignments, in which the clock rates were all set to 1 (Analysis 3). In an exploration of the factors affecting convergence, we ran three replicate runs using a random, smaller subset of 56 alignments (Analyses 0a-c) and gene-specific clock rates. SpeciesDelimitationAnalyser 1.8 (Jones, 2019) was used to summarise the posterior frequencies of clusters based on the posterior sample of trees with burn-in determined by maximising the ESS values, $\varepsilon = 10^{-5}$ and similarity cut-off set to 1. The number of species or minimal clusters (SMCs sensu Jones et al., 2015) relative to the number of accessions, mean posterior and likelihood ESS per unit time, and summed clade support for resolved SMCs were computed and used as a basis for comparison between the various STACEY analyses (Table 1). TreeAnnotator 2.6.0 (Drummond & Rambaut, 2007) was used to identify the topology with the greatest product of posterior branch support (the maximum clade credibility or MCC tree) from the posterior distribution of species trees. MCC trees were visualised in FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

sNMF and DAPC analyses (SNP data)

Ancestry coefficient estimates were computed with SNP data representing all 288 accessions using sNMF (Frichot et al., 2014) as implemented in the R package *LEA* 2.2.0 (Frichot & François, 2015). One hundred ancestry coefficient matrix replicates were computed for each value of k between k = 2 and k = 20 (Supplementary data). The optimal number of gene pools (the value of k that best explains the genotypic data) was selected using entropy criterion estimates for each k-value (Frichot & François, 2015). Ancestry coefficient estimates were summarised using Clumpak (cluster Markov packager across K; Kopelman et al., 2015) and visualised with the R package *conStruct* 1.0.4 (Bradburd, 2019). For identifying putative, metapopulation-like species, the sNMF ancestry coefficients were used to aggregate populations into clusters based on gene pool similarity.

For DAPC, prior k-means clustering and linear discriminant analysis (LDA) of principal components were implemented in the R package *adegenet* 1.3-1 (Jombart & Ahmed, 2011) with a matrix of biallelic SNPs typed for all 288 accessions. Bayesian Information Criterion (BIC) values were

Table 1. Summary of results of STACEY analyses on 36 accessions.

Analysis	Loci	Clock rates	Burn-in fraction	Generations	Clusters/ accessions[a]	ESS per day[b]	No. nodes with support ≥ 0.95[c]	Mean node support[d]
0a	56	Locus-specific	0.10	1×10^9	0.88	584.13	4	0.91
0b			0.10	1×10^9	0.89	471.27	5	0.94
0c			0.10	1×10^9	0.89	496.07	5	0.93
1	524[e]	Locus-specific	0.66	8.91×10^9	0.95	1.04	8	1.00
2a	524[f]	Locus-specific	0.71	5.54×10^9	0.88	2.25	8	1.00
2b			0.22	5.72×10^9	0.87	2.91	8	1.00
2c			0.90	4.76×10^9	0.96	1.49	8	1.00
2d			0.30	6.07×10^9	0.94	2.49	8	1.00
3	524[f]	Equal	0.38	5.69×10^9	0.95	5.49	8	1.00

[a] Number of Species or Minimal Clusters (SMCs *sensu* Jones et al., 2015) divided by the number of accessions.
[b] Mean effective sample size (ESS) per day computed as $[0.5 \times (posterior\ ESS + likelihood\ ESS)]/analysis\ runtime\ in\ days$
[c] Number of clades recovered in common by all STACEY analyses with posterior probability ⩾ 0.95.
[d] Mean posterior probability for clades recovered in common by all analyses.
[e] 524-locus data set with conflicting sequences in 24 loci uncorrected.
[f] 524-locus data set with conflicting sequences in 24 loci corrected.

Table 2. Clusters of populations recovered by consensus across alternative genomic methods.

Consensus clusters[a]	STACEY clades	sNMF	DAPC	Populations
I	a	A	4	2
II	l	B	10	1, 3, 12
III	i	C	11	9, 10, 11, 13, 28
IV	e	D	5 +12	4, 5, 6, 7, 8, 14, 15, 16
V	k	F	8	17, 18
VI	d	G	2	20, 21, 26, 27
VII	f	H	14	22, 30, 44, 48
VIII	g + h	I + J	7	19, 23, 29, 31, 42, 43
IX	j	K	3	33, 35, 36, 37, 38, 39, 40, 41
X	c	L	1	34, 45
XI	b	M	6	32
XII	m[b]	E	9 + 13	24, 25, 46, 47

[a] Consensus defined as groups recovered by all genomic methods without partitioning.
[b] Grade recovered with PP < 1.00.

used to select the optimal number of clusters for values of k between 1 and 20 (Supplementary data).

Population clusters recovered by sNMF and DAPC were compared with the clades resolved by STACEY. Based on this comparison, we resolved a series of monophyletic consensus clusters (Table 2) corresponding to putative species.

IBD (SNP data)

Patterns of IBD within and between putative species (consensus clusters) were explored using Hausdorf and Hennig's (2020) regression-based tests of equality. These analyses are conducted in a pairwise manner, and for this study only compared consensus clusters showing a significant, linear relationship ($p < 0.05$) of genetic to geographic distance and for which at least three populations had been sampled. Genetic differentiation was quantified using Weir and Cockerham's (1984) F_{ST} as implemented in the R package *hierfstat* 0.5-7 (Goudet & Jombart, 2020) using SNPs typed for all 288 accessions. Like other dispersal-limited Cape plants which are disjunctly distributed in the archipelagic high elevation zone of the Cape Fold Belt mountains (Britton et al., 2014; Ellis et al., 2014; Pirie et al., 2017; Verboom et al., 2015), the distributions of the putative species resolved within *S. plumosum* tend to be nonlinear and elevationally-defined (Fig. 5; Supplementary data). As such, distances between populations were computed using least-cost paths with the R package *topoDistance* 1.0.1 (Wang, 2019) using elevation to define a 30 arc-second cost surface (Hijmans et al., 2005) with values 0.95 and 0.05 assigned according to

the minimum and maximum elevations occupied by montane and coastal consensus clusters, respectively (Table 2; Supplementary data). Significance was assessed using jackknife-based tests of equality among IBD regressions, as implemented in the R package *prabclus* 2.3-2 (Hennig & Hausdorf, 2020). These tests assess support for the null hypothesis that genetic distances between units belonging to putatively independent species do not exceed that expected based on within-species regressions. The first test (H_{01}) is used to assess the equality of slope and intercept estimates in separate regressions of two putatively independent species. Where H_{01} is not rejected, a one-sided test (H_{02}) is used to assess whether a regression of within-species distances differs significantly from a regression of between-species distances, with rejection of H_{02} providing evidence for species independence. Where H_{01} is rejected, test H_{03} is used to assess whether the genetic distances predicted (at the mean between-species distance) by regressions of species-specific distances (within-species, one species at a time) plus between-species distances significantly exceeds that predicted by regressions based on species-specific distances alone. Only significance for *both* H_{03} tests provides evidence for species distinctness (Hausdorf & Hennig, 2020).

Phenotypic and environmental data

To assess the phenotypic and ecological distinctness of putative species (consensus clusters), we conducted separate LDAs, using the R package MASS 7.3-54 (Venables & Ripley, 2002), on a set of 12 quantitative phenotypic traits and a set of nine environmental variables. Since the species represented by populations 2 and 32 (*S. cinereum*) are readily distinguished from the remainder, the former being a new species and the latter a long-recognised species outside of the *S. plumosum* complex, they were omitted from these analyses. The phenotypic trait set was assembled from FAA-fixed voucher material and focused on capitulum, floret, and leaf traits, which were measured using a dissecting microscope with an eyepiece graticule precise to 0.2 mm. Traits included were: number of outer involucral bracts ("nbr"), inner involucral bract length ("lib"), outer:inner involucral bract length ratio ("blr"), number of pappus bristles ("npb"), length of pappus bristles ("plb"), corolla length ("cle"), width of corolla at base ("cbw"), terminal tube length:basal corolla width ratio ("cwr"), terminal tube length:lobe length ratio ("clr"), length of leaf subtending leaf fascicle ("lil"), length:width ratio of fascicle-subtending leaf ("lsl"), and length of leaf within leaf fascicle ("lsl"). To assess the environmental distinctness of consensus clusters, we quantified each population locality in terms of four climatic variables (minimum temperature in the coldest month ["mtc"], mean temperature of the wettest quarter ["mtw"], precipitation in the wettest quarter ["pwq"], precipitation in the driest quarter ["pdq"]), four soil variables (soil cation exchange capacity ["sce"], soil clay percentage ["scl"], soil carbon ["sca"], KCl-based soil pH ["spk"]), and elevation ("elv"), using the R package *raster* 3.4.13 (Hijmans, 2021) to extract these values from the relevant WorldClim (Fick & Hijmans, 2017) and SoilGrids (Hengl et al., 2017) layers. Where the climatic variables were selected to reflect variation in minimum temperature and in the amount and seasonality of precipitation, soil variables were selected to reflect variation in soil texture and chemistry.

RESULTS

Molecular data assembly

Following initial quality trimming, GBS yielded approximately 200,000 150 bp paired-end reads per accession with an average read depth of 8, totalling more than 5.7 x 10^8 raw sequence reads across all 288 accessions. Bioinformatic processing in the dDocent pipeline produced 7 610 586 variant sites, including bi- and multi-allelic SNPs, indels, and composite indel and substitution events. After filtering reads using genotype depth, locus quality, minor allele frequency, and genotype call depth, 32 145 variant sites across 288 accessions remained. Filtering to include only biallelic SNPs produced a 28 023-SNP data set with a mean SNP missing data level of 16.83% across all 288 sequenced accessions. Filtering to include only SNPs typed for all 288 accessions (for DAPC and F_{ST} estimates for IBD) yielded 415 SNPs. Mining full sequences from the 28 023-SNP data set by back-referencing the location of SNPs for a 36-accession subset yielded 1,329 sequence alignments. After filtering for the number of parsimony-informative sites, more than two haplotypes of read depth ⩾3 per accession, and haplotype separation across UPGMA root nodes, 524 alignments remained. These had a mean length of 155.43 bp (SD = 16.48) and contained a mean of 6.17 parsimony-informative sites (SD = 4.25), such that the full sequence alignment data set comprised 3 234 parsimony-informative sites across up to 72 haplotypes in 36 accessions.

STACEY (sequences)

In the context of the full 524-alignment data set, applying a common clock to all loci (Analysis 3, Table 1) considerably improved both parameter estimation and convergence, the latter being poor when each locus had its own clock (Analyses 1 and 2, Table 1). With the first 2.16 x 10^9 generations removed as burn-in, Analysis 3 yielded ESSs >177 for all parameters (Supplementary data). Interestingly, using a subset of 56 randomly selected alignments (Analysis 0, Table 1) yielded an ESS increase rate more than two orders of magnitude greater than those shown by the 524-alignment analyses and gave virtually identical and well-supported (posterior probability, PP >0.95; Supplementary data) phylogenetic relationships. Since the minimal clusters and MCC tree resolved by Analyses 1, 2, and 3 are all very similar (Supplementary data), we present only the pairwise conspecificity matrix and MCC tree for Analysis 3, whose ESSs are superior (Fig. 2).

STACEY resolves species or minimal clusters (SMCs *sensu* Jones, 2016) corresponding to Wright-Fisher species. Our analysis consistently identified every accession as representing a distinct minimal cluster (i.e., species), the sole exception being populations 30 and 44 which were identified as conspecific in 75.32 % of the posterior samples

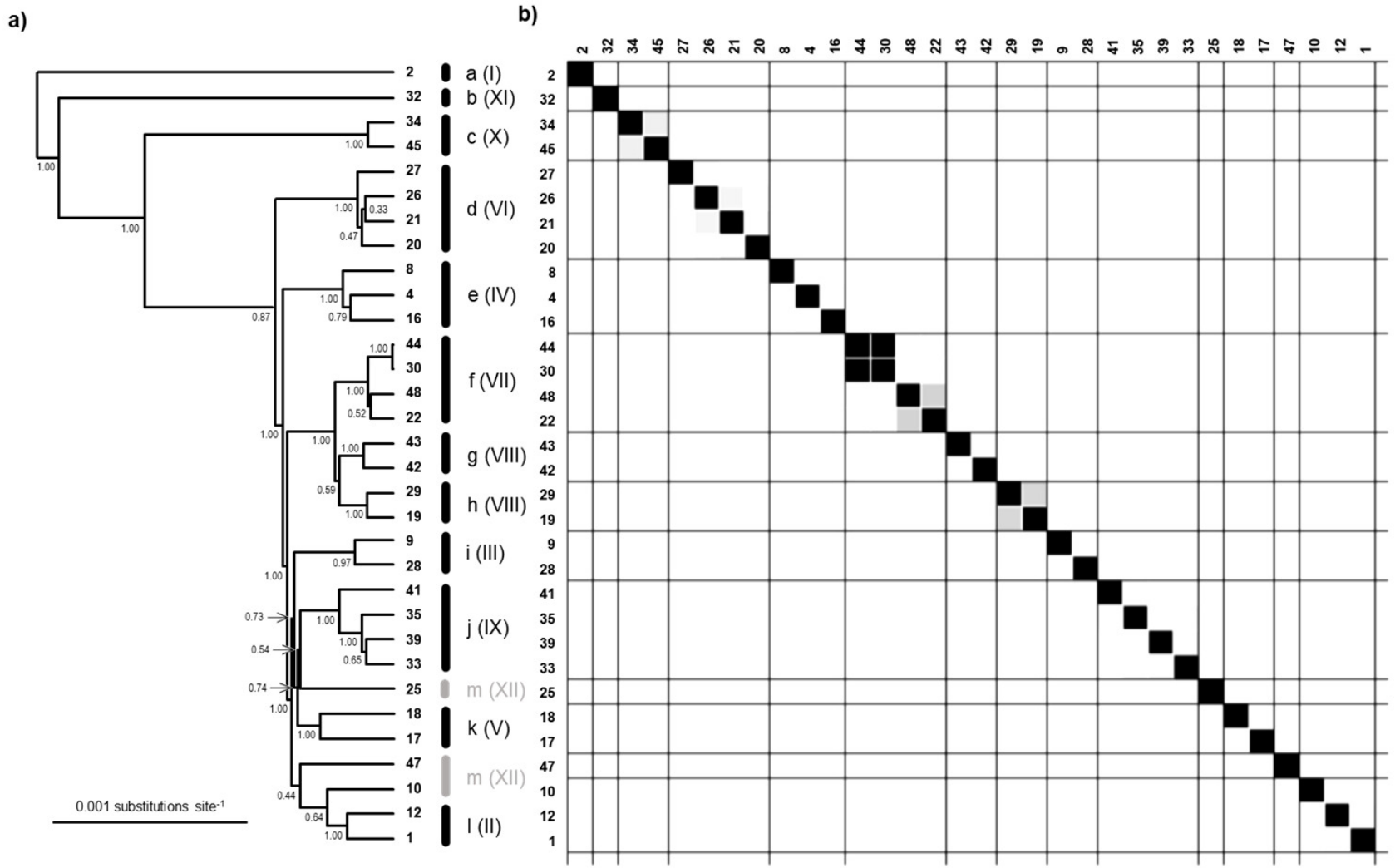

Figure 2. STACEY analysis: (a) MCC tree depicting population relationships for 36 *S. plumosum* and *S. cinereum* accessions.

Branch labels on the tree denote posterior probabilities, and branch lengths are in units of expected substitutions per site. Supported clades (a-l) and a residual grade (m) are indicated to the right, along with the consensus cluster (I-XII; Table 2) to which each belongs (in parentheses). (b) Two-way conspecificity matrix with pairwise posterior probabilities (white = 0, black = 1) that accessions are members of the same species or minimal cluster *sensu* Jones et al. (2015) under $\varepsilon = 10^{-5}$.

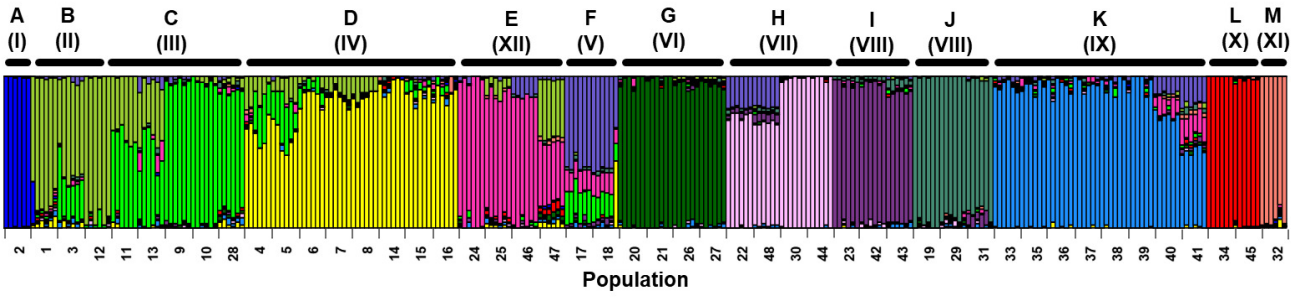

Figure 3. sNMF analysis: gene pools recovered under the major mode for the minimum entropy criterion obtained at k =13.

Vertical bars represent 288 accessions in *S. plumosum* and *S. cinereum*. Bar height represents the probability of accession assignment to a gene pool. Numbering below bars denotes populations, and colours gene pools. Letters above bars (A-M) depict population clusters inferred based on gene pool sharing, along with the consensus cluster (I-XII; Table 2) to which each belongs (in parentheses).

(Fig. 2b). Resolution and node support in the MCC population tree was generally high, especially on the deeper nodes (Fig. 2a), allowing for the identification of nine well-supported, two- to four-population clades (Fig. 2a: a-l) that potentially correspond to metapopulation-like species. Populations 2 (Matroosberg) and 32 occupy isolated positions in the tree, the former representing an apparent single-population species and the latter being the only sampled representative of *S. cinereum*, while populations 10, 25, and 47 form a grade (Fig. 2a: m) that is paraphyletic with respect to clades i, j, k, and l.

sNMF (SNPs)

Cross entropy criterion values were highly similar for values of k between about ten and 17 (Supplementary data) but achieved a minimum of 0.254 for k = 13. The major mode for k = 13 (Fig. 3; supported by 48 replicates), as well as the minor modes (supported by 23, 19, and ten replicates, respectively), are largely consistent in the allocation of accessions to gene pools (Supplementary data). Working with an arbitrarily selected minimum threshold probability/frequency of ca. 0.25, gene pool sharing patterns allow for the recognition of 13 population clusters (Fig. 3: A-M); these clusters also are distinct with regards to gene pool composition. Twelve of these clusters are located within *S. plumosum* (Fig. 3: A-L), and one within *S. cinereum* (Fig. 3: M). Within *S. plumosum* there is evidence of substantial admixture between population clusters B and C (light green and olive gene pools) and between B and E (olive and dark pink gene pools) and more limited admixture between E and F (dark pink and violet gene pools), of C into D (green into yellow gene pool), of C into F (green into violet gene pool), of F into H (violet into pale pink gene pool), and of E/F into the K (violet and dark pink into blue gene pool). Most admixture, then, appears to be between population clusters B, C, E, and F, with clusters A, D, and G-M being more distinct.

DAPC (SNPs)

DAPC employs principal components analysis to reduce the dimensionality of the genetic data before using a simple combination of k-means clustering and discriminant analysis to resolve and assign individuals to genetic clusters. Principal components analysis of the 415 SNPs typed for all 288 accessions captured 97.5 % of the total genetic variability in the first 150 principal components, which formed the input of subsequent steps. The BIC identified 14 clusters as optimal, with k = 14 having BIC = 1000.95 and alternative k values having BIC scores at least five units higher (Supplementary data). Under k = 14, k-means clustering resolved populations 2 (Matroosberg), 32 (*S. cinereum*), and 47 as single-population clusters, the remaining populations being grouped into clusters of two to eight populations (Fig. 4). Linear discriminant analysis reveals clusters 2 (Matroosberg), 7, and 14 to be especially distinct.

Consensus clusters

There is a high degree of correspondence between the clusters resolved by sNMF and DAPC and the supported clades resolved by STACEY (Table 2). All three analyses identified populations 2 (Matroosberg) and 32 (*S. cinereum*) as distinct entities, these occupying isolated, "early-divergent" positions on long branches in the population tree. Based on the results of the three analyses, the remaining populations can be grouped into ten consensus clusters, labelled I to XII (including the entities represented by populations 2 and 32). Areas of disagreement between the three analyses are as follows:

1. Consensus cluster IV – Where STACEY and sNMF retrieve a clade/cluster comprising populations 4-8 and 14-16, this is split into two clusters by DAPC.
2. Consensus cluster VIII – Where DAPC retrieves a cluster comprising populations 19, 23, 29, 31, 42, and 43, both STACEY and sNMF split this into two clusters; the split, however, differs between the two analyses.
3. Consensus cluster XII – Where sNMF resolves a cluster comprising populations 24, 25, 46 and, 47, DAPC splits this into two clusters, and STACEY resolves this assemblage as a paraphyletic grade. Also, where STACEY includes population 10 in this grade, population 10 is placed in a separate cluster (III), also con-

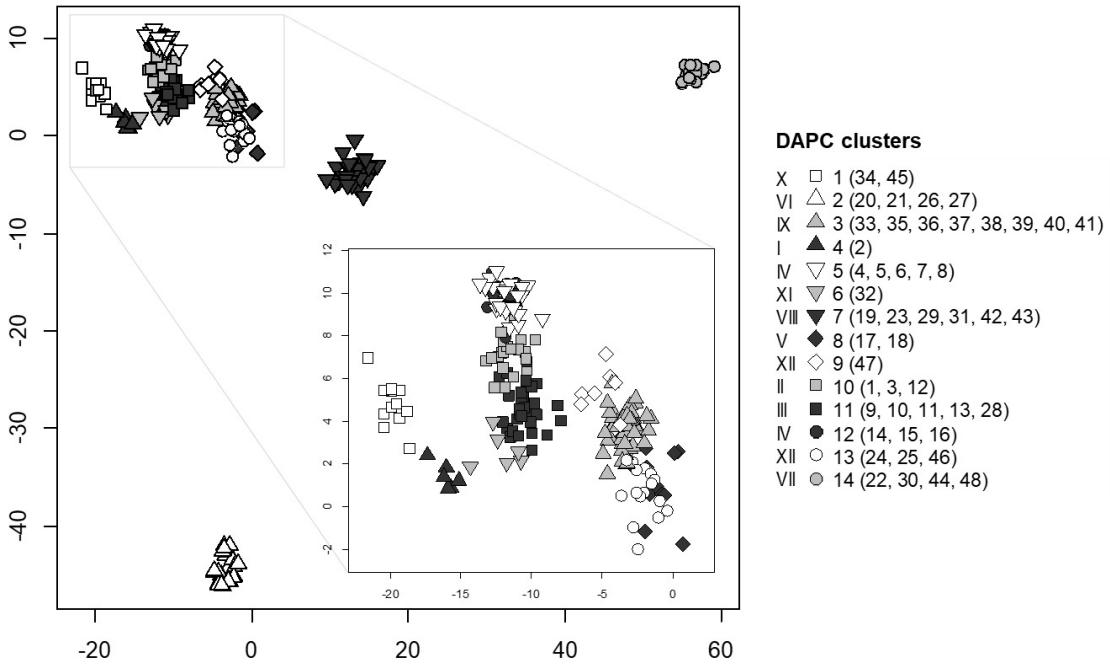

Figure 4. SNP-based discriminant analysis of principal components (DAPC) for the optimal BIC criterion obtained at k = 14.

Points represent 288 *S. plumosum* and *S. cinereum* accessions, and symbols denote clusters resolved by DAPC. The legend indicates which populations are assigned to each DAPC cluster and the consensus cluster (I-XII; Table 2) to which each belongs.

taining populations 8, 11, 13, and 28, by DAPC and sNMF.

Consensus clusters I to XII have distinct geographic distributions (Fig. 5), with clusters I, II, III, and IV associating with the mountains of the western CFR; clusters V, VII, and VIII associating with lower elevations along the CFR south coast; clusters VI and XII associating with mountains along the interior margin of the CFR; and clusters IX and X occurring outside the CFR, in the eastern parts of South Africa. Despite several cluster pairs being broadly sympatric (occurring within a few kilometres of each other) at some localities, sNMF indicates that they maintain their genetic integrity at these sites, showing little evidence of gene pool sharing (Fig. 3). These broadly sympatric clusters include clusters VII (population 44) and XI (population 32) on the Cape Peninsula; VII (population 23) and VIII (population 48) at Oyster Bay; III (population 28) and IV (populations 6, 7 and 8) at Limietberg; I (population 2) and II (population 3) at Matroosberg; VI and XII at Seweweekspoort (VI: population 27; XII: populations 46 and 47) and Swartberg (VI: population 26; and XII: populations 24 and 25). Indeed, only clusters II (population 12) and III (population 13) show evidence of potential gene pool sharing in sympatry at Skurweberg.

IBD (SNPs and geography)

Least-cost path geographic distance is a significant predictor of genetic distance in four of the six pairs of consensus clusters for which regression-based tests of equality were conducted. Among these, the pairwise genetic independence of consensus cluster pairs IV/IX, VIII/IX, IV/VIII, and VIII/XII are well-supported (Fig. 6a, b, d, e). While the tests of equality do not support the genetic independence of cluster pairs IX/XII (Fig. 6c) and IV/XII (Fig. 6f), this may reflect low statistical power associated with the small number of cluster XII populations sampled.

Phenotypic and ecological distinctness

With the 12 consensus clusters treated as *a priori* groups, LDA revealed generally subtle phenotypic differentiation between clusters. Since an initial LDA revealed clusters VI, VII, and VIII to be relatively distinct (Fig. 7a), a second LDA was run with these clusters omitted (Fig. 7c). This shows separation of clusters IX+X from II+III+IV+V+XII along LD axis 1 and separation of IX from X and V from II+III+IV+XII along LD axis 2 (Fig. 7c). The latter separations are subtle, however, reflecting considerable phenotypic overlap. Although trait vectors (Fig. 7a, c) and univariate trait comparisons (Supplementary materials) permit the identification of key traits which allow for some consensus clusters, specifically V, VI, VII, VIII, and X, to be distinguished from some or all remaining species (Table 3), trait overlap means that the diagnostic utility of these traits is imperfect. Certainly, phenotype provides little basis for distinguishing between clusters II, III, IV, IX, and XII, although within this group cluster IX is perhaps most distinctive on account of its relatively longer, narrower leaves.

Linear discriminant analysis based on environmental variables identifies the consensus clusters as aggregating in distinct ecological settings (Fig. 7e; Table 3), with clusters

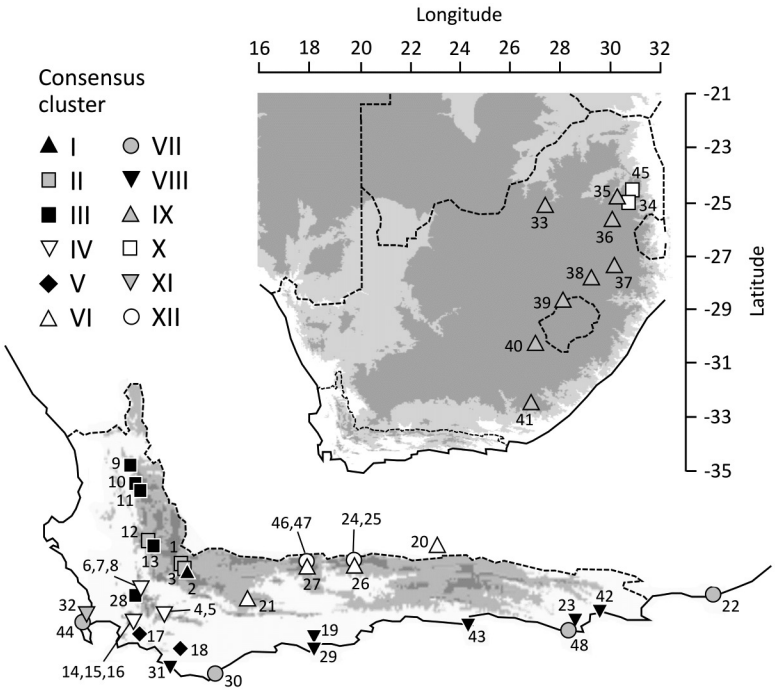

Figure 5. Map of South Africa, with the Cape Floristic Region (dashed line) expanded (below).

Sample populations are mapped, with symbols indicating the consensus cluster (I-XII; Table 2) to which each belongs.

IX and X occupying higher elevation, summer-precipitation sites corresponding to the eastern plateau of South Africa, and the remainder occupying comparatively lower-lying, winter- and all year-precipitation sites in the CFR. Amongst the latter, clusters V, VII, and VIII associate with frost-free, low-elevation sites along on the CFR coastal platform whose soils are devoid of organic material (low soil carbon); while clusters II, III, and IV associate with wetter, higher-elevation sites located in the Cape Fold Mountains whose soils are rich in organic material. Finally, clusters VI and XII associate with high-elevation, all year-precipitation sites located along the interior margin of the CFR.

DISCUSSION

Genetic discovery of metapopulation-like species

Many well-defined species are not panmictic across their ranges on account of habitat discontinuity, mating and dispersal behaviour, and/or population history, instead being comprised of more loosely-knit, metapopulation-like entities that display significant spatial genetic structure (e.g., Huang, 2020; Twyford et al., 2020; Watts et al., 2016). Since MSC-based species delimitation approaches, at least as currently implemented, assume within-species panmixia, they are likely to overlook such metapopulation-like structuring and so resolve as distinct species units that correspond more closely to biological populations (Chambers & Hillis, 2020; Sukumaran & Knowles, 2017). This appears to be true for the *S. plumosum* complex, in which STACEY identifies all populations except 30 and 44 as a separate "species" (Fig. 2). This is not surprising, perhaps, given the complex, heterogeneous character of the Cape physical environment (Cramer & Verboom, 2016; van Mazijk et al., 2021) which promotes range discontinuities and spatial genetic structure in the ranges of many species (e.g., Britton et al., 2014; Prunier & Holsinger, 2010; Tassone et al., 2020).

In this paper, we employ an alternative suite of analytical methods for the genetic discovery of species that have more polythetic and metapopulation-like characteristics. These methods are complementary in yielding different biological insights and testing different aspects of a species hypothesis. Where STACEY provides the genealogical insights that are needed to ensure that the putative species correspond to lineages, an essential property of species under most species concepts (de Queiroz, 1998, 2007; Freudenstein et al., 2017), sNMF and other STRUCTURE-like approaches provide insights into patterns of gene pool sharing, both historical and contemporary (Lawson et al., 2018). Tests of IBD, on the other hand, examine patterns of gene sharing between populations in an explicitly spatial context, thereby allowing for the discovery of population clusters that correspond to spatial fields of gene exchange. Finally, although DAPC clustering lacks a clear biological interpretation, its assumption-free, ordination basis allows for the discovery of polythetic groups that may correspond to species and so accord it corroborative value (cf. Carstens et al., 2013). In our study, these methods show remarkable congruence in terms of the groups they reveal, with most of the population clusters resolved by sNMF-inferred gene pool sharing and DAPC corresponding to clades, and the boundaries between them displaying IBD disjunctions. On this basis we resolve twelve consensus population clusters (I to XII) representing putative species. As in previous work (Britton et al., 2014; Unmack et al., 2021)

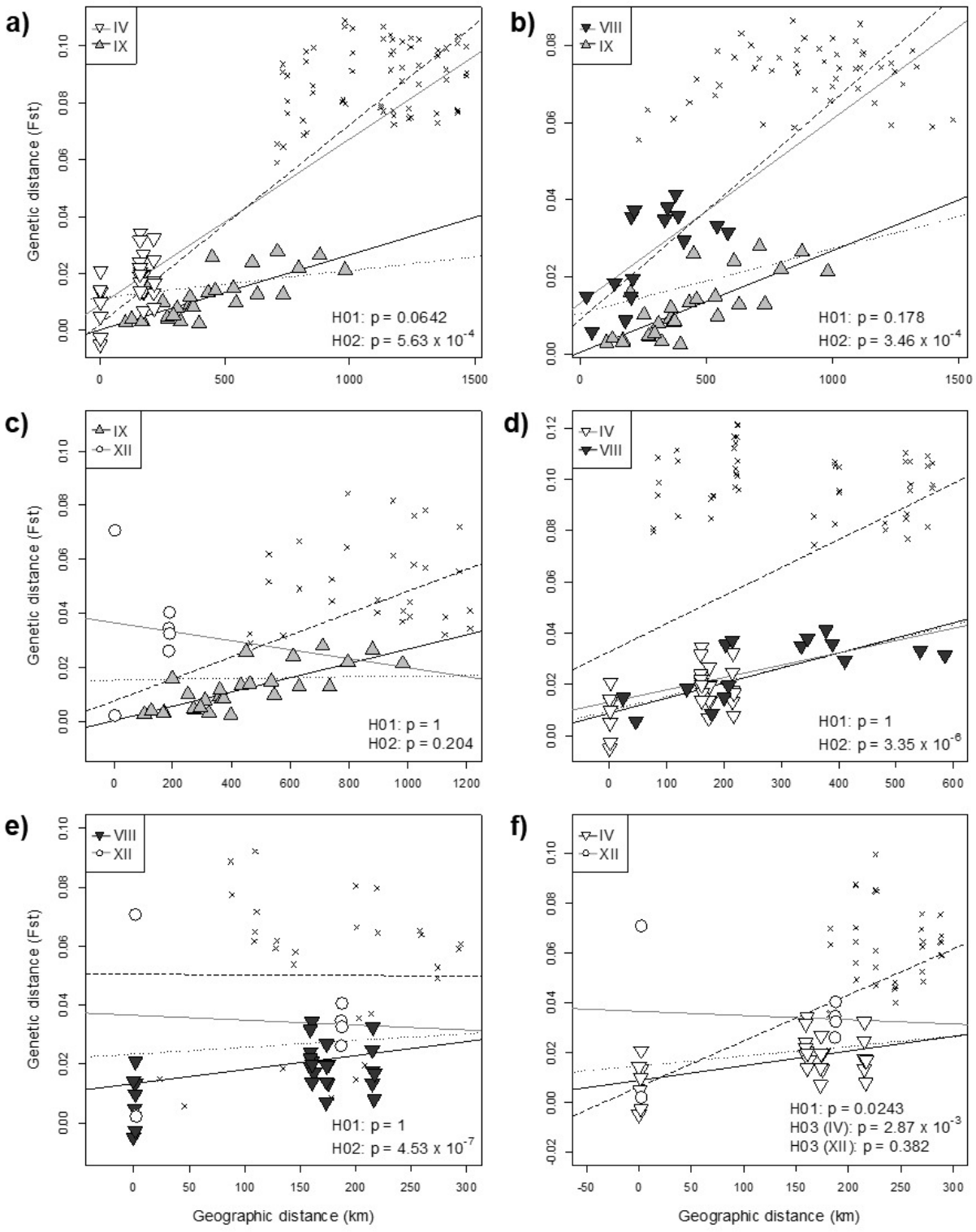

Figure 6. Regression-based tests of equality for all pairwise combinations of four qualifying genomic consensus clusters as candidate species (Supplementary data; Table 2).

Cross symbols denote distances between candidate species; filled symbols denote distances between populations within candidate species. Top left insets show consensus clusters; bottom right insets show Bonferroni aggregated (H_{01}) and approximate (H_{02} and H_{03}) p-values. (a - e) Pairwise comparisons for which distances in two candidate species can be modelled by a single regression (H_{01} is not rejected). Solid lines: regression lines within candidate species. Dashed line: regression line for distances in the two candidate species only. Dotted line: regression line for all distances, including distances between candidate species. H_{02} p-values indicate support for candidate species as independent. (f) Pairwise comparison where regression lines between candidate species differ significantly (H_{01} is rejected). Solid lines: regression lines within candidate species. Dashed line: regression line for candidate species XII and distances between candidate species. Dotted line: regression line for candidate species IX and distances between candidate species. Only significant p-values for both H_{03} tests provide support for candidate species as independent.

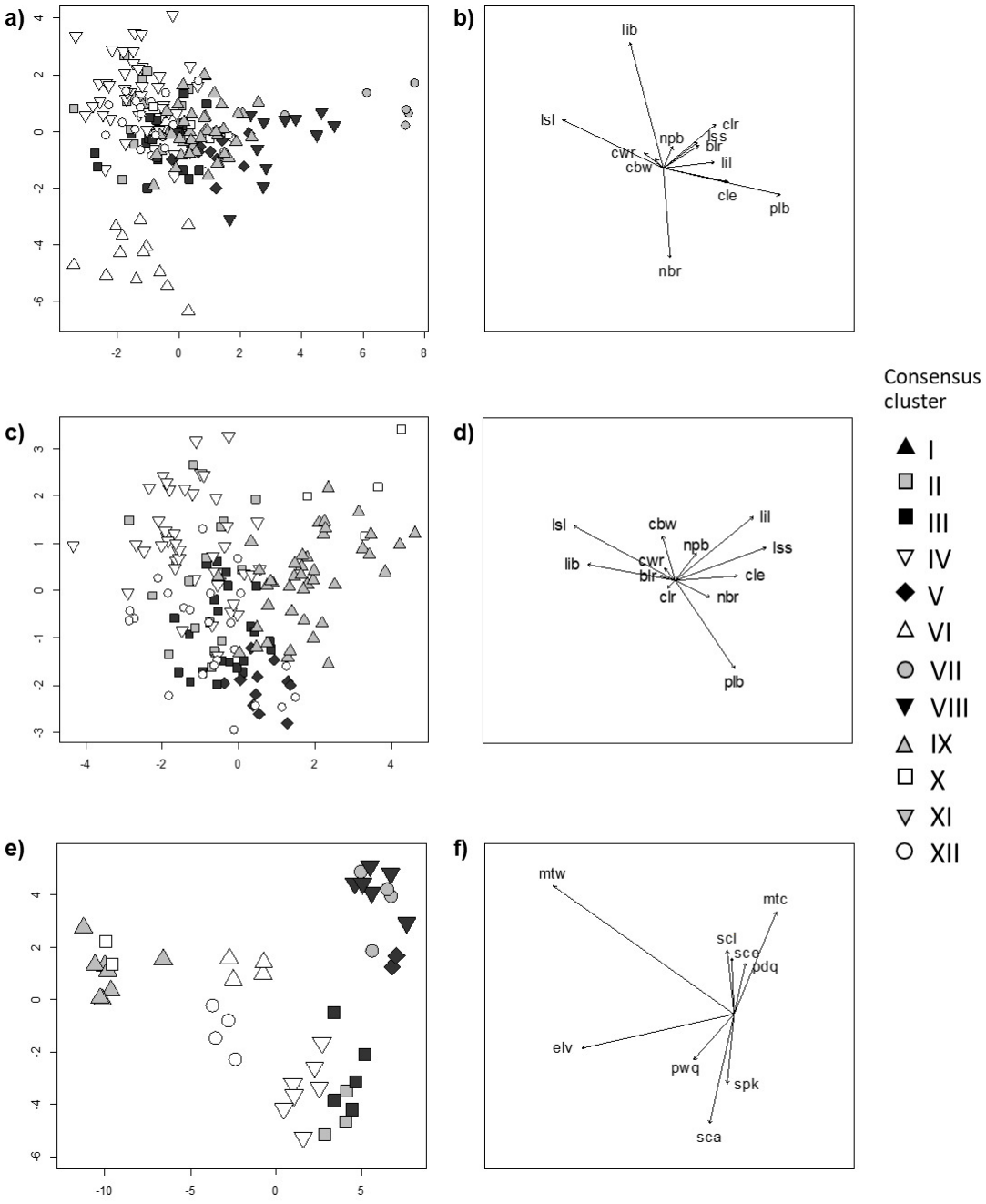

Figure 7. Linear discriminant analysis of phenotypic and environmental data: (a) ordination of 12 phenotypic characters for consensus clusters II-X and XII (Table 2): number of outer involucral bracts ("nbr"), inner involucral bract length ("lib"), outer:inner involucral bract length ratio ("blr"), number of pappus bristles ("npb"), length of pappus bristles ("plb"), corolla length ("cle"), width of corolla at base ("cbw"), terminal tube length:basal corolla width ratio ("cwr"), terminal tube length:lobe length ratio ("clr"), length of leaf subtending leaf fascicle ("lil"), length:width ratio of fascicle-subtending leaf ("lsl"), and length of leaf within leaf fascicle ("lsl") with (b) associated variable vectors; (c) ordination of 12 phenotypic characters after excluding clusters VI-VIII with (d) associated variable vectors; and (e) ordination of nine environmental variables: minimum temperature in the coldest month ("mtc"), mean temperature of the wettest quarter ("mtw"), precipitation in the wettest quarter ("pwq"), precipitation in the driest quarter ("pdq"), soil cation exchange capacity ("sce"), soil clay percentage ("scl"), soil carbon ("sca"), KCl-based soil pH ("spk"), and elevation ("elv") with (f) associated vectors.

Table 3. Taxonomic, morphological, and ecological characteristics of consensus clusters I-X in the *Seriphium plumosum* complex.

Consensus clusters	Interim taxon designation	Morphological characteristics	Ecological characteristics
I	*S. sp. nov.* 1	Compact dwarf shrub, leaves generally not fascicled, densely grey-woolly, relatively large and adpressed to stem	Known from a single population of c. 100 plants on shale near the summit of Conical Peak in the Hex River mountains above 1900 m elevation. Co-occurs with a population that matches with the narrower concept of *S. plumosum* as defined here
II, III, V, XII	*S. plumosum* L. (narrower concept as defined here)	Synflorescence spicate or paniculate, branching angle generally acute, especially prominent pair of oil glands at base of leaves subtending fascicles	Core distribution in the mountains and valleys of the Cape Fold Belt, distribution large but populations often highly localised, on shale, sandstone and quartz between 100 and 2000 m elevation
IV	*S. sp. nov.* 2	Resembles *S. plumosum* but can be distinguished using a combination of characters; leaves generally longer (Fig. 7a, b), always densely tomentose, imbricate or widely spaced and arranged in poorly defined fascicles	On shale and quartz on the slopes and summits of the mountains in the extreme southwestern Cape between 1000 and 1700 m elevation
VI	*St. burchellii* Levyns	Defined by glabrous, minute (<1mm) leaves and short (<3 mm) inner bracts (Supplementary data), synflorescence always paniculate	On shale and sandstone slopes in and around the Little Karoo in the eastern part of the Cape Floristic Region above 500 m elevation, co-occurs with *S. plumosum* as defined here
VII	*S. sp. nov.* 3	Defined by supple, whip-like branches and grey-white succulent leaves, synflorescence always borne below the level of the uppermost leaf shoots. Synflorescence an aggregated paniculate assemblage rather than a tight spike. Corolla and pappus bristles longer than all other species in the complex, up to 3.5 mm and 3.4 mm respectively (Fig. 7a, b; Supplementary data)	Endemic to unconsolidated coastal dunefields and dune slacks below 10 m elevation, forms isolated stands
VIII	*S. sp. nov.* 4	Closely resembles *S. sp. nov*. 3 but leaves non-succulent, branches not whip-like, corolla and pappus bristles similar in length to other species in the complex (Supplementary data)	On consolidated calcareous substrates along the southern coast from Noordhoek near Cape Town to Uitenhage near Gqeberha between 10 and 550 m elevation
IX	*St. vulgaris* Levyns	Resembles *S. plumosum* as defined here but with longer leaves and has a subterranean rootstock from which new stems resprout post-fire. These are absent in *S. plumosum* which is killed by fire. Branching angle always acute	Widespread in grassland habitats of Angola, Namibia, Lesotho, Eswatini, Zimbabwe and the summer-rainfall region of South Africa on a variety of substrates between 1400 and 1800 m elevation
X	*S. sp. nov.* 5	Resembles *S. vulgaris* and *S. cinereum* in growth form and leaf length but always branches at 90° and has non-reflexed leaves in fascicles up to 3 mm long (Supplementary data)	Known from two populations on shale and quartz near the extreme north-eastern edge of the Great Escarpment above 1600 m elevation.
XI	*S. cinereum* L.	Leaves always reflexed and longer than most other species in the complex. Synflorescence always a well-defined, tight spike	On moist mountain slopes in the southwestern Cape mountains on sandstone and shale above 100 m elevation

the maintenance of genetic distinctness of some clusters in zones of sympatry provides valuable insights, suggesting that these entities function as biological species *sensu* Mayr (1969).

Our work highlights the utility of GBS as a cost-effective and highly flexible approach for capturing short reads with moderate read depth (mean = 8 for our data) from which both SNP markers and full sequences can be assembled *de novo* (Cotton, 2016; Leaché & Oaks, 2017; Rubin et al.,

2012). As demonstrated in this paper, a single GBS data set can be used to assess Wright-Fisher species limits, reconstruct phylogenetic relationships at shallow evolutionary timescales, and resolve genetically distinct clusters of populations. In principle, our novel approach for assembling haplotyped sequences from GBS data in R can be extended for RADseq data, although the Stacks 2.0 pipeline developed by Rochette et al. (2019) for short-read data is probably more user-friendly and can be run in parallel across multiple processing cores. A preliminary comparison of our pipeline and Stacks 2.0 (Supplementary data), however, shows that Stacks generates slightly shorter (mean = 151.58 bp) full sequences with only half the number (mean = 3.06) of parsimony-informative sites.

Species delimitation in the *S. plumosum* complex

Our population phylogeny (Fig. 2) identifies consensus clusters I, X, and XI as the most divergent lineages sampled in this study. Cluster XI corresponds to *S. cinereum*, a species that is widespread in the southwestern CFR where it co-occurs with *S. plumosum* and from which it is consistently distinguished by its longer, reflexed leaves. Although our genetic sampling of this species is limited to a single population, our data show no evidence of admixture between this species and any other lineage sampled (Fig. 3), despite the sampled population (32) being broadly sympatric with population 44 (consensus cluster VII). Clusters I and X, on the other hand, represent newly discovered species, the first ("*S. sp. nov.* 1") known only from a single site on the Matroosberg and the second ("*S. sp. nov.* 5") from a few sites in the escarpment grasslands of Mpumalanga. Like *S. cinereum*, both these species are phenotypically distinct, *S. sp. nov.* 1 (cluster I) being distinguishable by its compact growth form and larger, imbricate leaves adpressed to the stem and *S. sp. nov.* 5 (cluster X) by its 90-degree branching angle and longer, non-reflexed leaves. Both are also genetically distinct, maintaining this distinctness even where they co-occur with members of other clusters.

Within the clade comprising clusters II-IX and XII, clusters IV, VI, VII, and VIII are phylogenetically distinct with good support, except for VIII whose two subclades are better supported than the clade as a whole. These clusters are also largely distinct in terms of the gene pool composition, this distinctness being maintained in sites of sympatry, and IBD tests revealing gene flow breaks between them and other clusters. Finally, except for IV, these clusters are all phenotypically distinguishable from the core *plumosum-vulgaris* clade (II+III+V+IX+XII): although, individual trait measurements invariably show some overlap. Where cluster VI is distinguishable primarily by its minute, ovate leaves, cluster VII is principally distinguishable by its growth form, succulent leaves, and longer florets, and cluster VIII by its strong resemblance to cluster VII except for its non-succulent leaves. Since, on phylogenetic, genetic, and phenotypic evidence, clusters IV, VI, VII, and VIII all appear to function as independent evolutionary entities, they qualify as distinct species; a lack of gene pool mixing in sites of sympatry also identifies them as potential biological species. While VI was previously recognized as distinct by Levyns (1937) and corresponding to *St. burchellii*, IV ("*S. sp. nov.* 2"), VII ("*S. sp. nov.* 3"), and VIII ("*S. sp. nov.* 4") are all new and await description.

Within the core *plumosum-vulgaris* clade (Fig 2a: d-l), cluster IX, which corresponds to the widespread, summer-rainfall grassland species *St. vulgaris* (Levyns, 1937), is monophyletic and possesses a largely distinct gene pool. It is also defined phenotypically by its longer leaves and subterranean rootstock from which new stems sprout post-fire, which no other species in the complex possess. On this basis, and in view of its distinct environmental niche (Table 3; and below), we recognize this entity (*St. vulgaris*) as distinct from the remainder of the *plumosum-vulgaris* clade. The case for segregating the remainder of this clade into multiple species is, however, less convincing. Although this residue includes three supported lineages (clusters II, III, and V), these show considerable gene pool admixture and are phenotypically indistinguishable. Moreover, together with cluster XII, they are ecologically similar, all inhabiting quartzitic landscapes of the CFR, albeit with some differences in elevational range. On that basis and given that support for the non-monophyly of this assemblage is weak, we propose to merge these entities into a single species representing a narrower concept of *S. plumosum* than that which has traditionally been employed.

Is species diversity in the S. plumosum complex cryptic?

Recognizing the evolutionary and conservation importance of cryptic diversity (Bernardo, 2011; Fišer et al., 2018), we do not necessarily share the view that distinctness of biological role and/or phenotype is a crucial and necessary component of species (Freudenstein et al., 2017). However, we readily acknowledge the utility of phenotypic diagnosability as a species property and support its use as an auxiliary criterion in species delimitation, particularly where the case for evolutionary distinctness, inferred using other lines of evidence, is equivocal. Indeed, this approach is foundational to our delimitation of the *S. plumosum* complex. Most species resolved in this study are not, however, truly cryptic *sensu* Mann and Evans (2008), instead showing some level of morphological differentiation (see above), range exclusivity (Fig. 5), and, in most cases, habitat difference (Table 3). Where *St. vulgaris* and *S. sp. nov.* 5 associate with grassland habitats in the summer-rainfall eastern part of South Africa, all remaining species associate with shrublands in the winter- and all-year rainfall regions of the CFR. The latter also show ecological differentiation, including species that associate strictly with high-elevation shales (*S. sp. nov.* 1), with mid- to high-elevation quartzites (*S. plumosum*, *St. burchellii*, *S. sp. nov.* 2), with unconsolidated coastal dunes (*S. sp. nov.* 3), and with consolidated calcareous soils on the coastal platform (*S. sp. nov.* 4). On that basis, most species diversity in the *S. plumosum* complex is best considered pseudocryptic or, in some cases, semicryptic *sensu* Mann and Evans (2008). Indeed, perhaps the only true instance of species crypsis in this complex in-

volves *S. sp. nov.* 2 and *S. plumosum*; these entities being virtually indistinguishable phenotypically and ecologically, but representing distinct, non-sister lineages and distinct gene pools.

Genetic species discovery in the Cape

In exploiting the complementary insights provided by different data sets and analytical approaches, our work on the *S. plumosum* complex employs an integrative approach to species discovery and taxonomy (Dayrat, 2005; Fujita et al., 2012; Padial et al., 2009), perhaps most closely resembling the work of Unmack et al. (2021) who similarly stress the value of integrating phylogenetic and population genetic approaches in an explicitly spatial context. Spatial context is important since the maintenance of genetic integrity in sympatry provides powerful evidence of reproductive isolation (Britton et al., 2014; Huang, 2020), while analysis of IBD patterns (e.g., Hausdorf & Hennig, 2020) in conjunction with genetic clustering (e.g., Frichot et al., 2014; Pritchard et al., 2000) reveals entities that share a common spatial field of gene exchange. In this way within-species spatial genetic structure is explicitly accounted for, thereby enabling the discovery of metapopulation-like species.

In the context of phenotypically recalcitrant species complexes, genetic species discovery paves the way for a more focused re-evaluation of morphology, phylogenetic, and population genetic patterns enabling diagnostic and non-diagnostic traits to be distinguished. While attempts to resolve species in such groups using morphology alone may be unfruitful, considering morphology in conjunction with molecular data often helps to make sense of the morphologic variation, even showing species which were initially considered cryptic to be semicryptic or non-cryptic (e.g., Korshunova et al., 2019). This is certainly the case in our study, in which most newly discovered species are ecologically distinct and show some level of phenotypic distinctness.

Recent DNA-based work on Cape plants, particularly on widespread species, commonly finds these to comprise complexes of multiple semi-cryptic to non-cryptic species. Examples include *Tetraria triangularis* (Cyperaceae; Britton et al., 2014) and *Erica abietina/viscaria* (Ericaceae; Pirie et al., 2017). Also, within *Protea mundii* and *P. repens* Prunier and Holsinger (2010) and Prunier et al. (2017) discovered multiple geographically and genetically distinct clusters consistent with evolutionary species. Our study adds to this growing body of work in providing evidence to support both the resurrection of two previously described species and the recognition of a further five new species within the *S. plumosum* complex and its relatives.

CONCLUSION

Stable, robust taxonomies remain an important imperative for downstream applications in research and conservation. In challenging groups where any single source of taxonomic information has limited power to resolve independent species, complementary sources of evidence may be essential for comprehensively cataloguing species-level diversity. Parallel advances in high-throughput sequencing and MSC- and IBD-based approaches for species delimitation over the last decade have further expanded the set of tools available to alpha taxonomists. Here we employ genomic, spatial, phenotypic, and environmental evidence to assess species limits in the taxonomically recalcitrant Cape-centred *Seriphium plumosum* complex. We recover compelling support for the recognition of at least eight independent species, most of which qualify as biological species (*sensu* Mayr, 1969) and show limited crypsis. Our findings, if extensible to other Cape plants, suggest that undescribed vascular plant diversity in the Cape flora considerably exceeds the recent one percent estimate of Treurnicht et al. (2017). Our work showcases the application of short-read massively-parallel sequence data with alternative, logically decoupled methods for species delimitation and, more broadly, highlights the utility of integrative taxonomy.

Acknowledgements

We thank the University of Cape Town's ICTS High Performance Computing team (hpc.uct.ac.za), the CIPRES Science Gateway (phylo.org), UNINETTSigma2, the National Infrastructure for High Performance Computing and Data Storage in Norway (sigma2.no), and Linköping University's National Supercomputer Centre (nsc.liu.se) for access to computational resources on cluster Tetralith. We thank Mr T. Trinder-Smith and Miss C. Christians for their generous assistance at the Bolus Herbarium. We also thank J. Burrows, S. Burrows, and B. Turpin of the Buffelskloof Private Nature Reserve and several postgraduate students of the Biological Scicnes Dept, UCT for assistance with field work. Professor A. Leaché and Dr G. Jones are thanked for helpful advice running MSC models.

Funding

This work was supported by grant 105976 from the South African National Research Foundation, the South African National Research Foundation's Masters Innovation award, the South African National Biodiversity Institute, and the Joan Wrench Scholarship Fund.

Supporting information

Genotyping-by-sequencing data in binary alignment mapping (BAM) format and associated *de novo*-assembled genome have been deposited in the NCBI short-read archive: BioProject PRJNA705849. Data available from the Dryad Digital Repository: https://datadryad.org/stash/dataset/doi:10.5061/dryad.0cfxpnw46.

Submitted: June 14, 2022 EDT, Accepted: October 17, 2023 EDT

References

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Andermann, T., Fernandes, A. M., Olsson, U., Töpel, M., Pfeil, B., Oxelman, B., Aleixo, A., Faircloth, B. C., & Antonelli, A. (2019). Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. *Syst Biol.*, *68*(1), 32–46. https://doi.org/10.1101/255752

Bamberger, S., Xu, J., & Hausdorf, B. (2021). Evaluating species delimitation methods in radiations: the land snail Albinaria cretensis complex on Crete. *Systematic Biology*, *71*(2), 439–460. https://doi.org/10.1093/sysbio/syab050

Bernardo, J. (2011). A critical appraisal of the meaning and diagnosability of cryptic evolutionary diversity, and its implications for conservation in the face of climate change. In T. R. Hodkinson, M. B. Jones, S. Waldren, & J. A. N. Parnell (Eds.), *Climate Change, Ecology and Systematics* (pp. 380–438). Cambridge University Press. https://doi.org/10.1017/cbo9780511974540.019

Beugin, M.-P., Gayet, T., Pontier, D., Devillard, S., & Jombart, T. (2018). A fast likelihood solution to the genetic clustering problem. *Methods in Ecology and Evolution*, *9*(4), 1006–1016. https://doi.org/10.1111/2041-210x.12968

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Bourgeois, Y. X. C., & Warren, B. H. (2021). An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. *Molecular Ecology*, *30*(23), 6036–6071. https://doi.org/10.1111/mec.15989

Bradburd, G. (2019). *conStruct: models spatially continuous and discrete population genetic structure*. R Package Version 1.0.4. https://CRAN.R-project.org/package=conStruct

Britton, M. N., Hedderson, T. A., & Anthony Verboom, G. (2014). Topography as a driver of cryptic speciation in the high-elevation Tetraria triangularis (Boeck.). *Molecular Phylogenetics and Evolution*, *77*, 96–109. https://doi.org/10.1016/j.ympev.2014.03.024

Cai, L., Xi, Z., Lemmon, E. M., Lemmon, A. R., Mast, A., Buddenhagen, C. E., Liu, L., & Davis, C. C. (2021). The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. *Systematic Biology*, *70*(3), 491–507. https://doi.org/10.1093/sysbio/syaa083

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, *22*(17), 4369–4383. https://doi.org/10.1111/mec.12413

Chambers, E. A., & Hillis, D. M. (2020). The multispecies coalescent over-splits species in the case of geographically widespread taxa. *Systematic Biology*, *69*(1), 184–193. https://doi.org/10.1093/sysbio/syz042

Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In U. Bastolla, M. Porto, H. E. Roman, & M. Vendruscolo (Eds.), *Structural approaches to sequence evolution: molecules, networks, populations* (pp. 207–232). Springer. https://doi.org/10.1007/978-3-540-35306-5_10

Cotterill, F. P. D., Taylor, P. J., Gippoliti, S., Bishop, J. M., & Groves, C. P. (2014). Why one century of phenetics is enough: response to “are there really twice as many bovid species as we thought?” *Systematic Biology*, *63*(5), 819–832. https://doi.org/10.1093/sysbio/syu003

Cotton, J. A. (2016). From sequence reads to evolutionary inferences. In P. D. Olson, J. Hughes, & J. A. Cotton (Eds.), *Next generation systematics* (pp. 305–335). Cambridge University Press. https://doi.org/10.1017/cbo9781139236355.016

Cowling, R. M., Holmes, P. M., & Rebelo, A. G. (1992). Plant diversity and endemism. In R. M. Cowling (Ed.), *The Ecology of fynbos: Nutrients, fire and diversity* (pp. 62–112). Oxford University Press.

Cowling, R. M., & McDonald, D. J. (1998). Local endemism and plant conservation in the Cape Floristic Region. In P. W. Rundel, G. Montenegro, & F. Jaksic (Eds.), *Landscape degradation and biodiversity in Mediterranean-type ecosystems* (pp. 171–188). https://doi.org/10.1007/978-3-662-03543-6_10

Cramer, M. D., & Verboom, G. A. (2016). Measures of biologically relevant environmental heterogeneity improve prediction of regional plant species richness. *Journal of Biogeography*, *44*(3), 579–591. https://doi.org/10.1111/jbi.12911

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Darwin, C. (1859). *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life*. John Murray, Albemarle Street. https://doi.org/10.5962/bhl.title.82303

Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, *85*(3), 407–415. https://doi.org/10.1111/j.1095-8312.2005.00503.x

de Queiroz, K. (1998). The general lineage concept of species criteria, and the process of speciation. In *Endless Forms: Species and Speciation* (pp. 57–75). Oxford University Press.

de Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879–886. https://doi.org/10.1080/10635150701701083

Dobzhansky, T. (1937). *Genetics and the origin of species*. Columbia University Press.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*(1), 214. https://doi.org/10.1186/1471-2148-7-214

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Edwards, D. L., & Knowles, L. L. (2014). Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences*, *281*(1777), 20132765. https://doi.org/10.1098/rspb.2013.2765

Ellis, A. G., Anthony Verboom, G., van der Niet, T., Johnson, S. D., & Peter Linder, H. (2014). Speciation and extinction in the Greater Cape Floristic Region. In N. Allsopp, J. F. Colville, & G. A. Verboom (Eds.), *Lynbos: ecology, evolution, and conservation of a megadiverse region* (pp. 119–141). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679584.003.0006

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, *27*(3), 613–635. https://doi.org/10.1111/mec.14486

Fisher, R. A. (1930). *The genetical theory of natural selection* (2nd ed.). Clarendon Press. https://doi.org/10.5962/bhl.title.27468

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population. *Journal of Applied Ecology*, *46*(2), 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Freudenstein, J. V., Broe, M. B., Folk, R. A., & Sinn, B. T. (2017). Biodiversity and the Species Concept—Lineages are not Enough. *Systematic Biology*, *66*(4), syw098. https://doi.org/10.1093/sysbio/syw098

Frichot, E., & François, O. (2015). Lea: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210x.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983. https://doi.org/10.1534/genetics.113.160572

Fujita, M. K., Leaché, A. D., Burbrink, F. T., Mcguire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, *27*(9), 480–488. https://doi.org/10.1016/j.tree.2012.04.012

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*. arXiv preprint arXiv:1207.3907 [q-bio.GN]. https://arxiv.org/abs/1207.3907

Goldblatt, P. (1978). An analysis of the flora of southern Africa: its characteristics, relationships, and origins. *Annals of the Missouri Botanical Garden*, *65*(2), 369–436. https://doi.org/10.2307/2398858

Goldblatt, P. (1997). Floristic diversity in the Cape flora of South Africa. *Biodiversity and Conservation*, *6*(3), 359–377. https://doi.org/10.1023/a:1018360607299

Goldblatt, P., & Manning, J. C. (2002). Plant Diversity of the Cape Region of Southern Africa. *Annals of the Missouri Botanical Garden*, *89*(2), 281. https://doi.org/10.2307/3298566

Goudet, J., & Jombart, T. (2020). *hierfstat: Estimation and Tests of Hierarchical F-Statistics* [R package version 0.5-7].

Gratton, P., Trucchi, E., Trasatti, A., Riccarducci, G., Marta, S., Allegrucci, G., Cesaroni, D., & Sbordoni, V. (2016). Testing classical species properties with contemporary data: how "bad species" in brassy ringlets (Erebia tyndarus complex, Lepidoptera) turned good. *Systematic Biology*, *65*(2), 292–303. https://doi.org/10.1093/sysbio/syv087

Grummer, J. A., Bryson, R. W. Jr., & Reeder, T. W. (2014). Species delimitation using Bayes Factors: simulations and application to the Sceloporus scalaris species group (Squamata: Phrynosomatidae). *Systematic Biology*, *63*(2), 119–133. https://doi.org/10.1093/sysbio/syt069

Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, *4*(4), 336–344. https://doi.org/10.1111/2041-210x.12018

Hausdorf, B., & Hennig, C. (2020). Species delimitation and geography. *Molecular Ecology Resources*, *20*(4), 950–960. https://doi.org/10.1111/1755-0998.13184

Hedberg, I., & Hedberg, O. (1977). Chromosome numbers of afroalpine and afromontane angiosperms. *Botaniska Notiser*, *130*, 1–24. https://agris.fao.org/agris-search/search.do?recordID=US201302970122

Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse software packages* [R package version 0.0.11].

Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*(3), 570–580. https://doi.org/10.1093/molbev/msp274

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hennig, C., & Hausdorf, B. (2020). *prabclus: Functions for Clustering and Testing of Presence-Absence, Abundance and Multilocus Genetic Data* [R package version 2.3-2].

Hijmans, R. J. (2021). *raster: Geographic Data Analysis and Modeling* [R package version 3.4-13].

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Huang, J.-P. (2020). Is population subdivision different from speciation? From phylogeography to species delimitation. *Ecology and Evolution*, *10*(14), 6890–6896. https://doi.org/10.1002/ece3.6524

Jombart, T., & Ahmed, I. (2011). *adegenet 1.3-1*: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

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*(1), 94. https://doi.org/10.1186/1471-2156-11-94

Jones, G. (2016). Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology*, *74*(1–2), 447–467. https://doi.org/10.1007/s00285-016-1034-0

Jones, G. (2019, November 12). *SpeciesDelimitationAnalyser*.

Jones, G., Aydin, Z., & Oxelman, B. (2015). DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics*, *31*(7), 991–998. https://doi.org/10.1093/bioinformatics/btu770

Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. In H. N. Munro (Ed.), *Mammalian Protein Metabolism* (Vol. 3, pp. 21–126). Academic Press. https://doi.org/10.1016/b978-1-4832-3211-9.50009-7

Koekemoer, M. (2016). A monograph of Stoebe and Seriphium (Gnaphalieae, Asteraceae) in southern Africa and neighbouring islands. *Phytotaxa*, *242*(1), 1–84. https://doi.org/10.11646/phytotaxa.242.1.1

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inference across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Korshunova, T., Picton, B., Furfaro, G., Mariottini, P., Pontes, M., Prkić, J., Fletcher, K., Malmberg, K., Lundin, K., & Martynov, A. (2019). Multilevel fine-scale diversity challenges the 'cryptic species' concept. *Scientific Reports*, *9*(1), 1–23. https://doi.org/10.1038/s41598-019-42297-5

Latimer, A. M., Silander, J. A. Jr., & Cowling, R. M. (2005). Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science*, *309*(5741), 1722–1725. https://doi.org/10.1126/science.1115576

Latimer, A. M., Silander, J. A., Jr, Rebelo, A. G., & Midgley, G. F. (2009). Experimental biogeography: the role of environmental gradients in high geographic diversity in Cape Proteaceae. *Oecologia*, *160*(1), 151–162. https://doi.org/10.1007/s00442-009-1275-3

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05257-7

Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. *Systematic Biology*, *63*(4), 534–542. https://doi.org/10.1093/sysbio/syu018

Leaché, A. D., & Oaks, J. R. (2017). The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 69–84. https://doi.org/10.1146/annurev-ecolsys-110316-022645

Leaché, A. D., Zhu, T., Rannala, B., & Yang, Z. (2019). The spectre of too many species. *Systematic Biology*, *68*(1), 168–181. https://doi.org/10.1093/sysbio/syy051

Levyns, M. R. (1937). A revision of Stoebe L. *S Afr J Bot*, *3*, 1–35.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Lim, M. C. W., Witt, C. C., Graham, C. H., & Dávalos, L. M. (2019). Parallel molecular evolution in pathways, genes, and sites in high-elevation hummingbirds revealed by comparative transcriptomics. *Genome Biology and Evolution*, *11*(6), 1573–1585. https://doi.org/10.1093/gbe/evz101

Linder, H. P. (1985). Gene flow, speciation, and species diversity patterns in a species-rich area: the Cape flora. In E. S. Vrba (Ed.), *Species and Speciation* (pp. 53–57). Transvaal Museum.

Low, D. H. P. (2019). *ssviz: A small RNA-seq visualizer and analysis toolkit* [R package version 1.16].

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., … Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, *1*(1), 1–6. https://doi.org/10.1186/2047-217x-1-18

Mann, D. G., & Evans, K. (2008). The species concept and cryptic diversity. In Ø. Moestrup (Ed.), *Proceedings of the 12th international conference on harmful algae* (pp. 262–268). HAB Publications. https://www.researchgate.net/publication/262412308_The_species_concept_and_cryptic_diversity

Mason, N. A., Fletcher, N. K., Gill, B. A., Funk, W. C., & Zamudio, K. R. (2020). Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. *Systematics and Biodiversity*, *18*(3), 269–280. https://doi.org/10.1080/14772000.2020.1730475

Mayr, E. (1969). The biological meaning of species. *Biological Journal of the Linnean Society*, *1*(3), 311–320. https://doi.org/10.1111/j.1095-8312.1969.tb00123.x

Medrano, M., López-Perea, E., & Herrera, C. M. (2014). Population genetic methods applied to a species delimitation problem: endemic trumpet daffodils (Narcissus Section Pseudonarcissi) from the Southern Iberian Peninsula. *International Journal of Plant Sciences*, *175*(5), 501–517. https://doi.org/10.1086/675977

Mirarab, S., Reaz, R., Bayzid, Md. S., Zimmermann, T., Swenson, M. S., & Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, *30*(17), i541–i548. https://doi.org/10.1093/bioinformatics/btu462

Noguerales, V., Cordero, P. J., & Ortego, J. (2018). Integrating genomic and phenotypic data to evaluate alternative phylogenetic and species delimitation hypotheses in a recent evolutionary radiation of grasshoppers. *Molecular Ecology*, *27*(5), 1229–1244. https://doi.org/10.1111/mec.14504

O'Meara, B. C., Ané, C., Sanderson, M. J., & Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*, *60*(5), 922–933. https://doi.org/10.1111/j.0014-3820.2006.tb01171.x

Padial, J. M., Castroviejo-Fisher, S., Köhler, J., Vilà, C., Chaparro, J. C., & De la Riva, I. (2009). Deciphering the products of evolution at the species level: the need for an integrative taxonomy. *Zoologica Scripta*, *38*(4), 431–447. https://doi.org/10.1111/j.1463-6409.2008.00381.x

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2019). *Biostrings: efficient manipulation of biological strings* [R package version 2.50.2].

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Pirie, M. D., Oliver, E. G. H., Gehrke, B., Heringer, L., Mugrabi de Kuppler, A., Le Maitre, N. C., & Bellstedt, D. U. (2017). Underestimated regional species diversity in the Cape Floristic Region revealed by phylogenetic analysis of the Erica abietina/E. viscaria clade (Ericaceae). *Botanical Journal of the Linnean Society*, *184*(2), 185–203. https://doi.org/10.1093/botlinnean/box021

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Prunier, R., Akman, M., Kremer, C. T., Aitken, N., Chuah, A., Borevitz, J., & Holsinger, K. E. (2017). Isolation by distance and isolation by environment contribute to population differentiation in Protea repens (Proteaceae L.) a widespread South African species. *American Journal of Botany*, *104*(5), 674–684. https://doi.org/10.3732/ajb.1600232

Prunier, R., & Holsinger, K. E. (2010). Was it an explosion? Using population genetics to explore the dynamics of a recent radiation within Protea (Proteaceae L.). *Molecular Ecology*, *19*(18), 3968–3980. https://doi.org/10.1111/j.1365-294x.2010.04779.x

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). *dDocent*: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, *2*, e431. https://doi.org/10.7717/peerj.431

R Core Development Team. (2018). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Rannala, B., & Yang, Z. (2017). Efficient Bayesian species tree inference under the multispecies coalescent. *Systematic Biology*, *66*(5), 823–842. https://doi.org/10.1093/sysbio/syw119

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Rubin, B. E. R., Ree, R. H., & Moreau, C. S. (2012). Inferring phylogenies from RAD sequence data. *PLoS One*, *7*(4), 1–12. https://doi.org/10.1371/journal.pone.0033394

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592–593. https://doi.org/10.1093/bioinformatics/btq706

Schnitzler, J., Barraclough, T. G., Boatwright, J. S., Goldblatt, P., Manning, J. C., Powell, M. P., Rebelo, T., & Savolainen, V. (2011). Causes of plant diversification in the Cape biodiversity hotspot of South Africa. *Systematic Biology*, *60*(3), 343–357. https://doi.org/10.1093/sysbio/syr006

Semple, J. C., & Watanabe, K. (2009). A review of chromosome numbers in Asteraceae with hypotheses on chromosomal base number evolution. In V. A. Funk, A. Susanna, T. F. Stuessy, & R. J. Bayer (Eds.), *Systematics, evolution and biogeography of Compositae* (pp. 61–72). International Association for Plant Taxonomy.

Slingsby, P., & Bond, W. J. (1985). The influence of ants on the dispersal distance and seedling recruitment of Leucospermum conocarpodendron (L.) Buek (Proteaceae). *South African Journal of Botany*, *51*(1), 30–34. https://doi.org/10.1016/s0254-6299(16)31698-2

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *Sci Pap Univ Kansas Nat Hist Mus*, *38*(22), 1409–1438.

Solís-Lemus, C., Knowles, L. L., & Ané, C. (2014). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, *69*(2), 492–507. https://doi.org/10.1111/evo.12582

Spriggs, E. L., Eaton, D. A. R., Sweeney, P. W., Schlutius, C., Edwards, E. J., & Donoghue, M. J. (2019). Restriction-site-associated DNA sequencing reveals a cryptic Viburnum species on the North American coastal plain. *Systematic Biology*, *68*(2), 187–203. https://doi.org/10.1093/sysbio/syy084

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., Kistenich, S., Larsson, K.-H., Liow, L. H., Nowak, M. D., Stedje, B., Bachmann, L., & Dimitrov, D. (2018). Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, *33*(3), 153–163. https://doi.org/10.1016/j.tree.2017.11.007

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, *114*(7), 1607–1612. https://doi.org/10.1073/pnas.1607921114

Tang, H., Peng, J., Wang, P., & Risch, N. J. (2005). Estimation of individual admixture: analytical and study design considerations. *Genetic Epidemiology*, *28*(4), 289–301. https://doi.org/10.1002/gepi.20064

Tassone, E. E., Miles, L. S., Dyer, R. J., Rosenberg, M. S., Cowling, R. M., & Verrelli, B. C. (2020). Evolutionary stability, landscape heterogeneity, and human land-usage shape population genetic connectivity in the Cape Floristic Region biodiversity hotspot. *Evolutionary Applications*, *14*(4), 1109–1123. https://doi.org/10.1111/eva.13185

Treurnicht, M., Colville, J. F., Joppa, L. N., Huyser, O., & Manning, J. (2017). Counting complete? Finalising the plant inventory of a global biodiversity hotspot. *PeerJ*, *5*, e2984. https://doi.org/10.7717/peerj.2984

Twyford, A. D., Wong, E. L. Y., & Friedman, J. (2020). Multi-level patterns of genetic structure and isolation by distance in the widespread plant Mimulus guttatus. *Heredity*, *125*(4), 227–239. https://doi.org/10.1038/s41437-020-0335-7

Uninett Sigma2. (2020, October 12). *About Sigma2*.

Unmack, P. J., Adams, M., Hammer, M. P., Johnson, J. B., Gruber, B., Gilles, A., Young, M., & Georges, A. (2021). Plotting for change: an analytical framework to aid decisions on which lineages are candidate species in phylogenomic species discovery. *Biological Journal of the Linnean Society*, *135*(1), 117–137. https://doi.org/10.1093/biolinnean/blab095

Vallès, J., Canela, M. Á., Garcia, S., Hidalgo, O., Pellicer, J., Sánchez-Jiménez, I., Siljak-Yakovlev, S., Vitales, D., & Garnatje, T. (2013). Genome size variation and evolution in the family Asteraceae. *Caryologia*, *66*(3), 221–235. https://doi.org/10.1080/00087114.2013.829690

van Mazijk, R., Cramer, M. D., & Verboom, G. A. (2021). Environmental heterogeneity explains contrasting plant species richness between the South African Cape and southwestern Australia. *Journal of Biogeography*, *48*(8), 1875–1888. https://doi.org/10.1111/jbi.14118

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). Springer New York. https://doi.org/10.1007/978-0-387-21706-2

Verboom, G. A., Bergh, N. G., Haiden, S. A., Hoffmann, V., & Britton, M. N. (2015). Topography as a driver of diversification in the Cape Floristic Region of South Africa. *New Phytologist*, *207*(2), 368–376. https://doi.org/10.1111/nph.13342

Wang, I. (2019). *topoDistance: calculating topographic paths and distances* [R package version 1.0.1]. https://CRAN.R-project.org/package=topoDistance

Watts, M., Winkler, I. S., Daugeron, C., de Carvalho, C. J. B., Turner, S. P., & Wiegmann, B. M. (2016). Where do the neotropical Empidini lineages (Diptera: Empididae: Empidinae) fit in a worldwide context? *Molecular Phylogenetics and Evolution*, *95*, 67–78. https://doi.org/10.1016/j.ympev.2015.10.019

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wiley, E. O. (1978). The evolutionary species concept reconsidered. *Syst Biol*, *27*(1), 17. https://doi.org/10.2307/2412809

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159. https://doi.org/10.1093/genetics/16.2.97

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138. https://doi.org/10.1093/genetics/28.2.114

Wright, S. (1946). Isolation by distance under diverse systems of mating. *Genetics*, *31*(1), 39–59. https://doi.org/10.1093/genetics/31.1.39

Yang, Z., & Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, *31*(12), 3125–3135. https://doi.org/10.1093/molbev/msu279

Yang, Ziheng. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854–865. https://doi.org/10.1093/czoolo/61.5.854

Yang, Ziheng, & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, *107*(20), 9264–9269. https://doi.org/10.1073/pnas.0913022107