The dynamics of species divergence in South Africa's megadiverse Cape region.

Seth D. Musker [1,2,✉], **Michael D. Pirie** [3], **and Nicolai M. Nürk** [1]

[1]Department of Plant Systematics, Bayreuth Centre of Ecology and Environmental Research (BayCEER), University of Bayreuth, Universitätsstraße 30, 95447, Bayreuth, Germany

[2]Department of Biological Sciences, University of Cape Town, Private Bag, Rondebosch 7700, South Africa

[3]University Museum, The University of Bergen, Postboks 7800, N-5020, Bergen, Norway

Summary

Deciphering the ecological and geographic factors that influence the dynamics of population divergence can aid in understanding why some groups of organisms diversify more prolifically than others. One such diverse group is the heathers (*Erica*, Ericaceae), whose exceptional species richness in the Cape Floristic Region is enigmatic. Here, we study *Erica abietina*, a small but highly variable species complex with four described subspecies differing in geographic range, habitat, and floral characters. To understand the factors and forces that shaped its evolution, we evaluate the status of the subspecies and test for hybridisation, introgression, pollinator-driven divergence, and geographic population structure using genotyping-by-sequencing on samples across the entire distribution. We find that the four subspecies form variably distinct genetic groups, however, the most widespread subspecies exhibits cryptic diversity, comprising two independent lineages that are geographically isolated and occur on different soil types. We find that populations with different pollination syndromes are genetically distinct. However, for one pair of genetically distinct populations ($F_{ST} \approx 0.06$) with different floral phenotypes we uncover several individuals of hybrid origin at a site where they occur sympatrically. Together, these results suggest that floral differentiation acts in concert with geographic isolation to maintain reproductive isolation. Finally, we show that a reticulate history involving "ghost" introgression best explains the group's evolution. Our results reveal a highly dynamic system whose diversity has been shaped by a variety of interacting forces, and we suggest that such systems are likely to have contributed substantially to the diversity of *Erica* and the Cape flora in general.

Biodiversity | Speciation | Population Genetics | Cape Floristic Region | Ericaceae

Correspondence: *Seth.Musker@uni-bayreuth.de*

Introduction

Patterns of biological diversity are shaped by a variety of geographic, ecological, and genetic factors. At the interface of micro- and macro-evolution, these factors interact in multifaceted ways to promote or inhibit population divergence and eventual speciation (Templeton, 1981). This means that for any group of organisms, teasing apart the relative importance of the factors driving diversification requires combining genetic data with information on the group's distribution, ecology, and environment (Via, 2009; Donoghue and Sanderson, 2015). This can be challenging in highly dynamic systems—those in which species limits are unclear and a variety of processes, such as hybridisation, introgression, genetic drift, and selection, may act together. Such dynamic systems should be prevalent in global biodiversity hotspots (Myers et al., 2000), which are often typified by high rates of speciation (e.g., Madriñán et al., 2013). The Cape Floristic Region (CFR) is one such hotspot, hosting a megadiverse flora of over 9000 vascular plant species of which a remarkable 70% are endemic (Manning and Goldblatt, 2012). Considerable research effort has focused on the macroevolutionary factors responsible for the CFR's exceptional diversity (Verboom et al., 2009). In contrast, relatively few studies have investigated the dynamics of species divergence in its early stages using genomic methods and representative population sampling (cf. Lexer et al., 2014; Prunier and Holsinger, 2010; Prunier et al., 2017). Here, we present the first genomic study of highly dynamic early-stage diversification in *Erica* L. (Ericaceae), which, with almost 700 species, is easily the largest genus in the CFR (Manning and Goldblatt, 2012).

Along with Restionaceae and Proteaceae, *Erica* is a defining feature of the fynbos vegetation that dominates much of the CFR (Bergh et al., 2014), and which is typified by its occurrence on extremely nutrient-poor soils (Bradshaw and Cowling, 2014) and exceptionally high species turnover at small spatial scales (Kruger and Taylor, 1980). *Erica* underwent rapid and prolific diversification immediately following its arrival in the CFR around 10-15 million years ago (Ma) (Pirie et al., 2016, 2019), though what drove this "Cape clade" radiation remains unclear. Frequent pollinator shifts appear to have contributed to plant diversification in the CFR in general (e.g., Johnson, 1996; van der Niet and Johnson, 2009), and this might be the case for *Erica*: throughout the phylogeny of the Cape clade, floral traits appear to to be exceptionally labile (Pirie et al., 2011). In particular, shifts between insect- and bird-pollination may have been especially influential (van Der Niet et al., 2014). At least sixty *Erica* species are thought to be pollinated by birds, the vast majority of which are confined to the CFR (Rebelo et al., 1985), and almost all are pollinated by the CFR-endemic Orange-breasted Sunbird (*Anthobaphes violacea*; Rebelo et al., 1984; Coetzee, 2016; Coetzee et al., 2020). Sunbird-pollinated *Erica* tend to have long (18-26 mm) corolla tubes that match the dimensions of the birds' bills (Rebelo et al., 1984, 1985; Barnes et al., 1995), whereas most insect-pollinated species have short corollas (Rebelo et al., 1984). Bird-pollinated *Erica* are also unscented and come in a variety of flower colours, whereas insect-pollinated species are often sweetly scented and typically have pink or white flowers (Rebelo et al., 1984, 1985; Oliver and Oliver, 2002, 2005). Despite the distinctness of these pollination syndromes (Rebelo et al., 1985) they seem not to be strongly conserved among closely related species (Pirie et al., 2011). What remains unclear, however, is how and to what extent such pollinator shifts have stimulated speciation and contributed to Cape *Erica* diversity, and whether additional factors, such as geographic isolation and other forms of niche divergence, have played supporting or superior roles.

Here, we focus on the *Erica abietina* species complex, which comprises several phenotypically distinct forms varying in corolla tube length and flower colour (Table 1). On the basis of this floral variation the species has been classified

Table 1. Characteristics and geographic ranges of the subspecies of *Erica abietina*.

Subspecies	Corolla length	Corolla colour	Floral scent	Sepals	Ovary	Range
E. a. abietina	18–26 mm	Crimson to dark red	None	Lanceolate-attenuate/ acuminate; pilose	Elongate obovoid; pubescent	Northern CP: TM high plateau, N and W slopes
E. a. atrorosea	18–22 mm	Rose to deep rose pink	None	Lanceolate-acute; glabrous to sparsely puberulous	Ellipsoid; glabrous	Widespread on CP: E slopes of TM S to Cape Point
E. a. constantiana	8–11 mm	Pale to deep rose pink	Sweet, lemony	Lanceolate-attenuate/ acuminate; glabrous to sparsely puberulous	Squat obovoid; puberulent	Central CP: S slopes of TM to Chapman's Peak
E. a. diabolis	11–14 mm	Rose pink	None	Lanceolate-attenuate/ acuminate; pilose	Obovoid; pubescent	Northern CP: Devil's Peak

CP: Cape Peninsula; TM: Table Mountain.

into four subspecies (Oliver and Oliver, 2002; Pirie et al., 2017), all of which are confined to the Cape Peninsula (CP; Fig. 1). The CP is a 50 x 15 km mountain range of rugged topography at the south-western tip of South Africa that extends from Table Mountain in the north to Cape Point in the south. Being largely surrounded by ocean, the CP is isolated from the rest of the Cape Fold Mountains by an extensive sandy plain ($>$ 40 km wide) whose low elevation placed it below sea level during Pleistocene interglacials (Adamson, 1959), and which now holds the city of Cape Town. The CP is a hotspot within a hotspot of floristic diversity (Cowling et al., 1996) with over 2200 plant species (Trinder-Smith et al., 1996) of which 158 (including $\geq$ 39 *Erica* species) are endemic (Helme and Trinder-Smith, 2006). All of this implies that *E. abietina* may be experiencing diversifying forces in a system akin to a "continental archipelago" (cf. Hughes and Eastwood, 2006), making it an ideal system for studying the dynamics of the early stages of species diversification.

To illuminate the evolutionary history of the *E. abietina* species complex, we use genotyping-by-sequencing (GBS; Elshire et al., 2011) coupled with population-genetic and phylogenetic analyses. Firstly, we examine whether genetic differentiation in the species complex is correlated with geography and to what extent it corroborates current taxon boundaries. Secondly, we infer phylogenetic relationships between individuals and populations to test whether shifts between bird and insect pollination syndromes have occurred multiple times, so contributing to lineage divergence. Finally, we test for evidence of ongoing and ancient gene flow in order to infer the influence of hybridisation and introgression on species divergence, both past and present.

Materials and Methods

Sample collection and sequencing

We collected fresh leaf material from at least six individuals of each formally recognised taxon in addition to several individuals of putative hybrid origin (Table 1) and attempted to sample from across the full geographic range of each taxon (Fig. 1). We followed the taxon concepts (Meier, 2017) outlined by Oliver and Oliver (2002) and refined by Pirie et al. (2017) when identifying specimens. In addition, we sampled from the two putatively closest outgroups of *E. abietina*, *E. nevillei* (one sample) and *E. quadrisulcata* (three samples), both of which are endemic to the Cape Peninsula (Pirie et al., 2017). DNA extraction in *Erica* is challenging (Bellstedt et al., 2010), and we adapted the protocol of Inglis et al. (2018) to retrieve sufficient DNA for sequencing (see Supplemetary Material).

GBS library preparation and sequencing was done by Novogene Genome Sequencing Company Ltd. (Beijing, China). Protocols followed the original Elshire et al. (2011) pipeline except that two restriction enzymes were used: MseI and

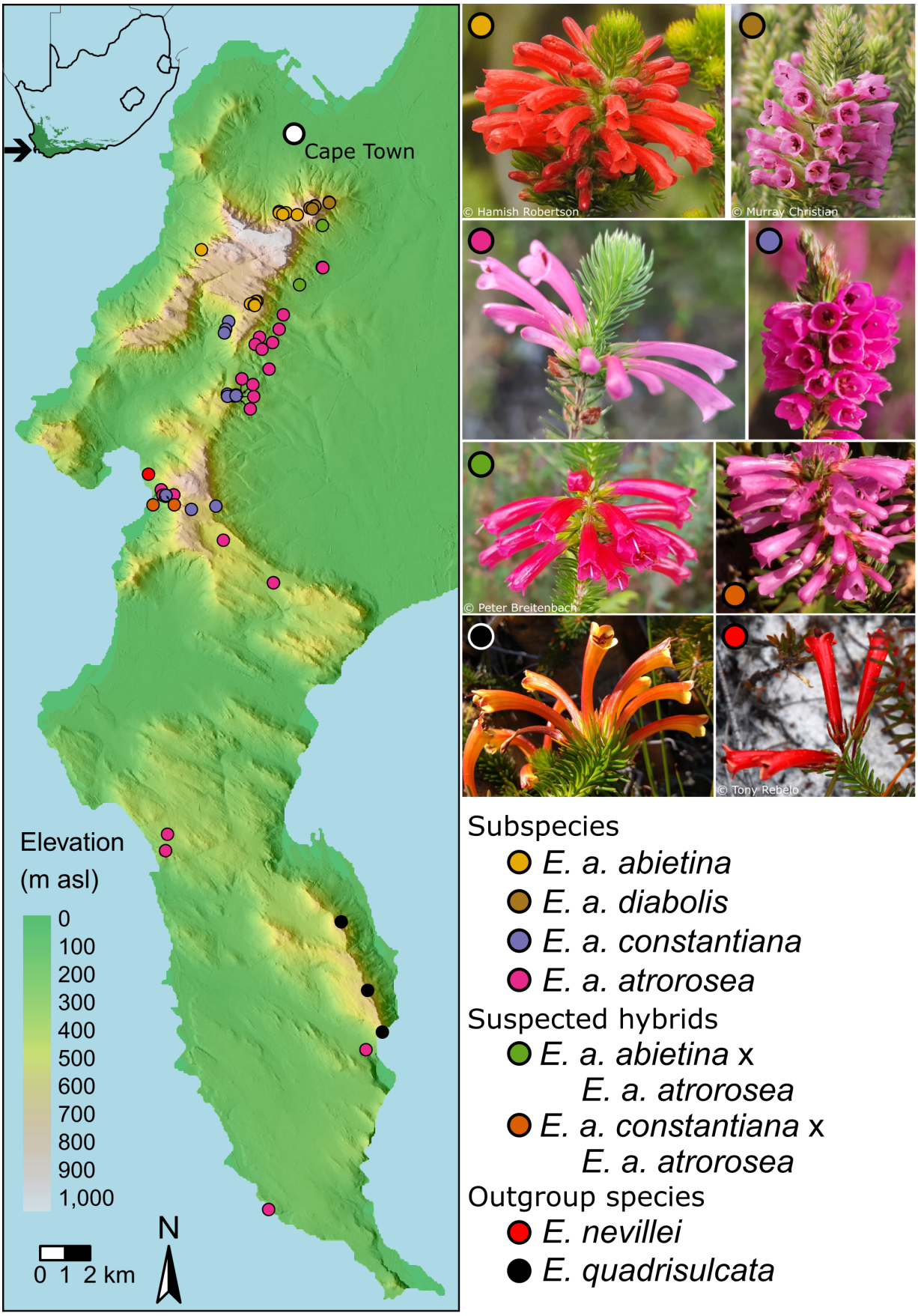

Figure 1. *Left*: Map of the Cape Peninsula showing elevation and hillshade with localities of collected samples. *Inset*: Map of South Africa showing the extent of the Fynbos biome and location of the Cape Peninsula. *Right*: Images of the focal taxa including putative hybrids.

HaeIII. Libraries were paired-end sequenced in two separate batches using an Illumina NovaSeq 6000 instrument to produce 144 bp-long reads after barcode removal. The two batches differed in the amount of sequencing effort per sample, though we found this to have little impact on the quality of the results. One sample was included in both batches to test the effect of sequencing effort on genotype calling accuracy, which we found to be negligible. Full details are available in the Supplemetary Material.

The quality of the raw reads was checked with FastQC (`http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc`) and MultiQC (Ewels et al., 2016). Adapter removal, quality trimming and filtering of reads was done with fastp (Chen et al., 2018, parameters: *–overrepresentation_analysis –trim_poly_g –qualified_quality_phred 20 –unqualified_percent_limit 30 –average_qual 20 –length_required 100*). Reads passing these filters were aligned to a draft reference genome of *Erica cerinthoides* (Musker et al., unpublished) using BWA-MEM (Li, 2013) with default parameters, followed by alignment sorting and indexing using SAMtools (Danecek et al., 2021). Alignments were merged using BamTools (Barnett et al., 2011).

Variants were called for all samples simultaneously using Freebayes v. 1.3.4 (Garrison and Marth, 2012, parameters: *–min-base-quality 3 –min-mapping-quality 10 –skip-coverage 10000 –use-best-n-alleles 4*). Table S1 outlines the variant calling results and the effects of quality filtering. Genotypes with read depth < 3 were recoded as missing using vcftools (Danecek et al., 2011). Using BCFtools (Danecek et al., 2021) we then kept variants with missingness (proportion of missing genotypes) < 0.5, minor allele frequency (MAF) > 0.01, allele balance < 0.01 or between 0.25 and 0.75, and MQM/MQMR (ratio of forward to reverse read mapping qualities) between 0.9 and 1.1. We then used *vcfallelicprimitives* from vcflib (Garrison, 2012) to decompose complex variants into SNPs where possible, and then removed non-SNP variants and non-biallelic SNPs. We used the BCFtools *fill-tags* plugin to test for excess heterozygosity, and removed SNPs with a p-value ≤ 0.2. We applied further filters as appropriate and used a naming scheme based on these filters (see Table S1). Filtering based on linkage disequilibrium (LD) was done using PLINK (Purcell et al., 2007, parameters: *–indep-pairwise 50 5 0.2*).

For calculating sequence diversity and absolute sequence divergence as well as for individual-level phylogeny inference, we generated an "all sites" VCF file by supplementing the SNP_m10 set (SNPs-only, missingness $< 10\%$; Table S1) with invariant sites. This was done by re-running FreeBayes as above but adding the "*–report-monomorphic*" switch, followed by removing sites with missingness > 0.1, decomposing complex variants as above, keeping only invariant sites, and finally combining the invariant sites with the SNP_m10 set using BCFtools. Several of the above processes were parallelised using GNU Parallel (Tange, 2020). The majority of the following analyses were conducted in the R statistical environment (R Core Team, 2021) using R v.4.1.2. VCF files were read using vcfR (Knaus and Grünwald, 2017).

Analysis of population structure

To assess population structure without any prior assumptions about group membership, we employed three complementary methods. Firstly, we ran a Principal Component Analysis on the SNP_m10_LD_maf04 set using the adegenet v.2.1.5 (Jombart and Ahmed, 2011) function *glPca*. Secondly, we employed the admixture model (Pritchard et al., 2000) for values of K (the number of ancestral populations) ranging from 2 to 8 using the sparse nonnegative matrix factorization (sNMF) algorithm (Frichot et al., 2014) implemented in LEA (Frichot and François, 2015), with default parameter values. For this we used the SNP_m10_LD_maf04 set

and excluded the outgroups *E. nevillei* and *E. quadrisulcata*. For each K we ran 100 independent repetitions of the algorithm and summarised the outputs using the CLUMPAK method (Kopelman et al., 2015) implemented in STARMIE (`https://github.com/sa-lee/starmie`). We generated bar plots of ancestry proportions using POPHELPER (Francis, 2017). Lastly, to better visualise the connections between individuals as well as explore the hierarchical structure of genetic variation, we employed network analysis using NETVIEW (Steinig et al., 2016, `https://github.com/esteinig/netview`). With the ALL_m10 set, we used PIXY (Korunes and Samuk, 2021) to calculate the harmonic mean of absolute sequence divergence (dxy) between individuals to generate the genetic distance matrix required by NETVIEW. We ran the network inference algorithm with values of k (the maximum number of mutual nearest neighbours) of 5, 10, 15 and 20, each time also estimating the minimum spanning tree to ensure a connected network was returned. We visualised each network with IGRAPH (Csardi and Nepusz, 2006) using the Kamada-Kawai spring-based layout algorithm (Kamada et al., 1989) to make the lengths of the connecting edges proportional to their associated genetic distance, and coloured the edges based on whether they were unique to the minimum spanning tree.

Based on these analyses we identified two genetically and geographically distinct northern and southern clusters within *E. a. atrorosea* (see Results), which we refer to as as *E. a. atrorosea* (North) and *E. a. atrorosea* (South). We distinguished these clusters for further analyses that required individuals to be grouped.

Detecting recent hybrids

To investigate the presence of hybrids between genetically distinct clusters, we sequenced four individuals of putative hybrid origin which we identified based on a combination of morphological features and geographic location (Fig. 1). Firstly, we identified two individuals from the mid-elevation eastern slopes of Table Mountain (TM) from populations showing a range of intermediate flower colours (magenta to cerise) between *E. a. abietina* (light red; TM plateau) and *E. a. atrorosea* (North) (pink; TM lower eastern slopes). Secondly, we identified two individuals from Blackburn Ravine with floral tube lengths intermediate between *E. a. atrorosea* (South) (18–22 mm) and *E. a. constantiana* (8–11 mm). To test the hybrid origin of these individuals we employed NEWHYBRIDS (Anderson, 2008) in combination with sNMF. For each set of putative parents and hybrids we subset the SNP_m10_LD_maf04 set to only the relevant individuals, removed resulting monomorphic sites, and ran sNMF with $K = 2$, again repeating the algorithm 100 times and summarising ancestry coefficients across runs as above. We then identified putatively non-admixed individuals as those whose maximum individual ancestry coefficient was ≥ 0.9, setting these as "P0" or "P1" in NEWHYBRIDS. To maximise the information content of the SNPs used, we calculated per-SNP F_{ST} (Weir and Cockerham, 1984) between P0 and P1 using the HIERFSTAT (Goudet, 2005) *basic.stats* function and chose the 500 SNPs with the highest F_{ST} while also only keeping one SNP per contig to avoid linkage effects. We used DARTR (Gruber et al., 2018) to convert the data into NEWHYBRIDS format. We then ran NEWHYBRIDS for 150,000 MCMC iterations, discarding the first 50,000 as burn-in. Based on these analyses we identified individuals showing evidence of recent hybrid origin (including backcrosses). To refer to data sets excluding these individuals, we append the suffix "_noHybrids".

Table 2. Assignment of taxa to newly defined "populations".

Taxon	Population	n
E. a. abietina	ABI	11
E. a. atrorosea (South)	$ATRO_S$	7
E. a. atrorosea (North)	$ATRO_N$	7
E. a. atrorosea (Blackburn Ravine)	$ATRO_{BLACKBURN}$	3
E. a. constantiana	$CONS$	8
E. a. diabolis	$DIAB$	5

Phylogenetic analysis

We inferred two individual-level phylogenies, one including all individuals (n = 65) and one excluding putative hybrid individuals (n = 45). We used IQ-TREE v.2.0.6 (Minh et al., 2020) using the full concatenated alignments. The ALL_m10 and ALL_m10_noHybrids VCF files were converted to fasta format using VCF2PHYLIP v.2.0 (Ortiz, 2019). For each analysis we chose the best-fitting substitution model using MODELFINDER (Kalyaanamoorthy et al., 2017) and the default Bayesian Information Criterion (BIC). We estimated node certainty with 1000 ultrafast bootstrap replicates (Hoang et al., 2018) and 1000 SH approximate likelihood ratio test (SH-alrt; Guindon et al., 2010) replicates, and ran four independent runs to improve the search of the likelihood space. We plotted the maximum-likelihood tree with GGTREE v.3.2.1 (Yu et al., 2017) after rooting the tree at *E. quadrisulcata* with APE v.5.0 (Paradis and Schliep, 2019).

Summary statistics

Individual-level summary statistics. To quantify genetic diversity we calculated per-sample multi-locus heterozygosity (MLH) using the R package INBREEDR (Stoffel et al., 2016) with the SNP_m10_LD_maf04 set. This revealed three anomalous samples within *E. abietina*, two with unusually low MLH which may be inbred (one from *E. a. abietina* and one from *E. a. atrorosea* [North]), and one with unusually high MLH (from *E. a. diabolis*) which may have been a chimeric sample stemming from a collection mishap in which two adjacent individuals were assumed to be one (Fig. S1).

Isolation-by-distance. To test for genetic isolation-by-distance (IBD) at the individual level, Mantel tests (Mantel, 1967) were performed separately for each taxon and for the two clusters within *E. a. atrorosea*. Geographic distances were determined after projecting collection localities onto the Hartebeesthoek94 coordinate reference system, and dxy was used as the genetic distance. Mantel tests were performed using the R package VEGAN v.2.6-4 (Oksanen et al., 2019), testing for statistical significance using 1000 permutations.

Grouping individuals into populations. Confidently estimating population-level summary statistics typically necessitates grouping individuals into putatively monophyletic, panmictic and outbred "populations". Based on the previous analyses, we removed 17 individuals with potentially mixed ancestry and three with anomalous MLH, and assigned the remainder to five distinct populations within *E. abietina*. Finally, we split *E. a. atrorosea* (South) into two groups, creating a separate group for three individuals collected from Blackburn Ravine whose inclusion would make the population paraphyletic according to the phylogenetic analyses (see Fig. 5). We distinguish these populations from the previously named entities as described in Table 2. All of the analyses described below used this assignment scheme.

Population-level summary statistics. To estimate the magnitude of pairwise genomic differentiation between the five populations plus the two outgroup taxa, we calculated pairwise Hudson's F_{ST} (Hudson et al., 1992; Bhatia et al., 2013) using ADMIXTOOLS2 (Maier et al., 2022).

Investigating population history

We next aimed to investigate the branching history of the group and test hypotheses of ancient introgression between populations. To estimate the population phylogeny while accounting for incomplete lineage sorting, we used a polymorphism-aware model (PoMo; Schrempf et al. 2019) implemented in IQ-TREE v.2.0.6 (Minh et al., 2020) with the SNP_m10_maf01 set. We set the substitution model to GTR+G4 (the GTR model [Tavaré 1986] with four discrete gamma rate categories), conducted 1000 ultrafast bootstrap replicates, and left all other parameters at their default values.

To search for evidence of ancient introgression, we used the *findGraphs* method implemented in ADMIXTOOLS2 (Maier et al., 2022) which conducts a heuristic search of the graph space by iteratively proposing modifications to the current graph – such as modifying the tree topology and adding admixed edges – each time evaluating the new graph's likelihood score, which is determined by comparing the f3-statistics (Peter, 2016) predicted by the graph to those estimated from the data. The algorithm attempts to find the best-fitting graph for the lowest number of allowed admixed edges (N_{ADMIX}) before adding more admixed edges one at a time.

We ran this analysis in two ways: first by specifying the population tree estimated by PoMo as the starting tree, and second by setting *E. quadrisulcata* as the outgroup without specifying any other restrictions to the randomly generated starting tree. In each case we ran the analysis five times. We started the graph searches at $N_{ADMIX} = 0$ and set the maximum value of N_{ADMIX} to 2. To search the likelihood space more exhaustively, we set the total number of generations after which to stop to 5000; the number of generations without improvement after which to stop to 100; the number of graphs evaluated in each generation to 30; and the *plusminus_generations* parameter (which helps to break out of local optima) to 20. All other parameters were kept at their default values. We manually inspected the ten resulting best graphs for each value of N_{ADMIX} to check for concordance in the graph topology, edge weights, and admixture proportions.

We compared alternative best-fitting graphs in a pairwise manner using a resampling procedure implemented in ADMIXTOOLS2 and described in Maier et al. (2022), which aims to test whether two graphs have similar predictive power. We used *qpgraph_resample_multi* to generate 100 replicate bootstrap resampled SNP block training and test sets and then evaluate each graph by estimating its weights using the training set and calculating its "out-of-sample" likelihood score using the (unseen) test set. This procedure allows to test the null hypothesis that the differences between two graphs' out-of-sample scores for each bootstrap replicate are equal to zero, i.e., that both graphs have equivalent predictive power. This was done using *compare_fits*, which conducts a two-sided z-test on the score differences, assuming a normal distribution of values and known standard deviation.

Results

Read mapping. After read mapping, a mean of 67.4% (±2.94% SD) of read pairs mapped properly to the *E. cerinthoides* draft genome. There was a small but statistically significant difference in mapping rate between the two batches (batch 1: 65.1%, batch 2: 67.9%; linear model, $F(1,64) = 9.84$, $p = 0.00258$) suggesting there were more

Table 3. Results of the Mantel tests of isolation-by-distance.

Population	Mantel's *r*	*p*-value	No. perm.
E. a. abietina	0.686	0.001	999
E. a. atrorosea (South)	0.449	0.005	999
E. a. atrorosea (North)	0.876	0.001	999
E. a. constantiana	0.718	0.001	999
E. a. diabolis	0.573	0.021	719

duplicated reads in batch 2. Mean mapping depth was 5.62 ($\pm$0.50 SD) and 9.10 ($\pm$1.63 SD) for batches 1 and 2, respectively.

Variant calling. Variant calling followed by stringent quality filtering resulted in a rich and highly complete SNP data set (Table S1). Individuals sequenced in batch 1 had more missing genotypes in the SNP_m10 set (mean = 8.81%, SD = 3.09%) than batch 2 individuals (mean = 2.50%, SD = 1.86 %), while the highest missingness in any individual was 13.4%. Genotype calls had lower read counts for batch 1 individuals (mean = 10.3, range = 8.1–12.7) than batch 2 individuals (mean = 22.6, range = 12.5–32.5). Missingness was generally higher in the outgroups (*E. nevillei*: 9.14%; *E. quadrisulcata*: 7.01–11.2%), though this was not appreciable, suggesting that allele dropout did not affect library preparation or variant detection.

Population structure

The Mantel tests for isolation-by-distance indicated that geographic distance was significantly correlated with genetic relatedness for all taxa, though the strength of IBD, measured as Mantel's *r*, was variable (Table 3). All analyses of population structure indicated strong genetic clustering corresponding broadly to the current taxonomic treatment of the species complex, with some noteworthy exceptions. The PCA eigenvalues (Fig. 2, *inset*) exhibited a steep decline in explained variance from axes 1 to 4 followed by a plateau from axes 5 to 7, after which they declined gradually. This pattern suggested the existence of either five or eight clusters in the data, as $n-1$ axes are required to distinguish n clusters. In contrast, and despite the exclusion of outgroups, the sNMF-based cross-entropy criterion suggested $K = 1$ or $K = 2$ to be optimal given a 5% genotype masking rate, although higher values of K generally recovered sensible groupings of individuals in concordance with the PCA and NetView analyses (Figs. 3,S2).

The first principal component axis (PC1; Fig. 2) primarily distinguished the outgroups from *E. abietina* and showed *E. nevillei* to be the closer of the two outgroups. Within *E. abietina*, PC2 and, to a lesser extent, PC1 distinguished *E. a. abietina* plus *E. a. diabolis* from the rest of the subspecies, and the sNMF results also recovered these two groups as the most important ancestral clusters at $K = 2$ (Fig. S2). Notably, there was consistent support for two distinct genetic clusters within *E. a. atrorosea*: One group (*E. a. atrorosea* [South]) consisted of individuals from the southern parts of the Cape Peninsula ranging from Cape Point to Silvermine, while the other (*E. a. atrorosea* [North]) comprised northern individuals collected along the lower eastern slopes of Table Mountain. In PCA space, *E. a. atrorosea* (North) fell between the two major clusters while still being closer to *E. a. atrorosea* (South) and *E. a. constantiana*. The NetView analysis (Fig. 3) revealed more fine-scale patterns of genetic structure and relatedness, especially at lower values of k (the maximum number of connections allowed between individuals). Within *E. a. abietina*, individuals sampled from different parts of Table Mountain were clearly recovered as belonging to distinct network clusters at $k = 5$. Within *E. a. atrorosea* (North), the four southernmost individuals were consistently

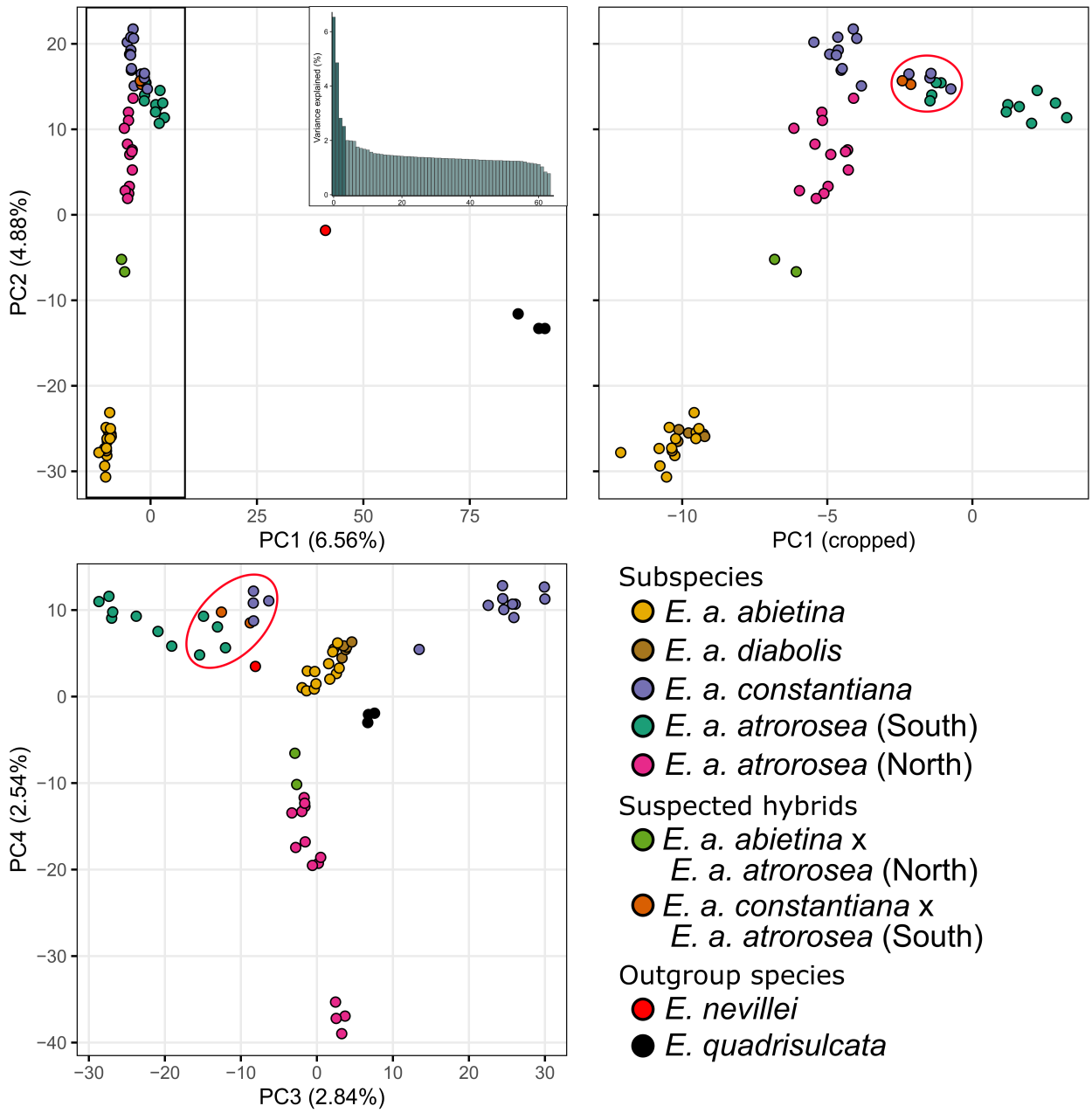

Figure 2. Ordination of the genotype matrix by PCA. The box in the first plot indicates the area covered by the second plot. Samples from Blackburn Ravine are indicated by red ellipses. The inset shows a plot of the variance explained by each PCA axis (also shown in brackets in axis titles). Note especially the close relationship between *E. a. abietina* and *E. a. diabolis* and their distinctness from the other subspecies; the positioning of the two samples identified as *E. a. abietina* x *E. a. atrorosea* hybrids based on morphology; and the positioning of samples from Blackburn Ravine.

recovered as distinct from and unconnected to any other populations, unlike the rest of *E. a. atrorosea* (North). NetView also more consistently recovered *E. a. abietina* and *E. a. diabolis* as distinct from each other than the PCA (Fig. 2) or sNMF (Fig. S2) analyses did.

Evidence for recent and ongoing hybridization

There was widespread evidence for recent hybridisation within *E. abietina*. The two individuals originally suspected to have hybrid ancestry between *E. a. atrorosea* and *E. a. abietina* were consistently recovered by sNMF as sharing ancestry predominantly from clusters corresponding to *E. a. abietina* and *E. a. atrorosea* (North) (Fig. S2), and

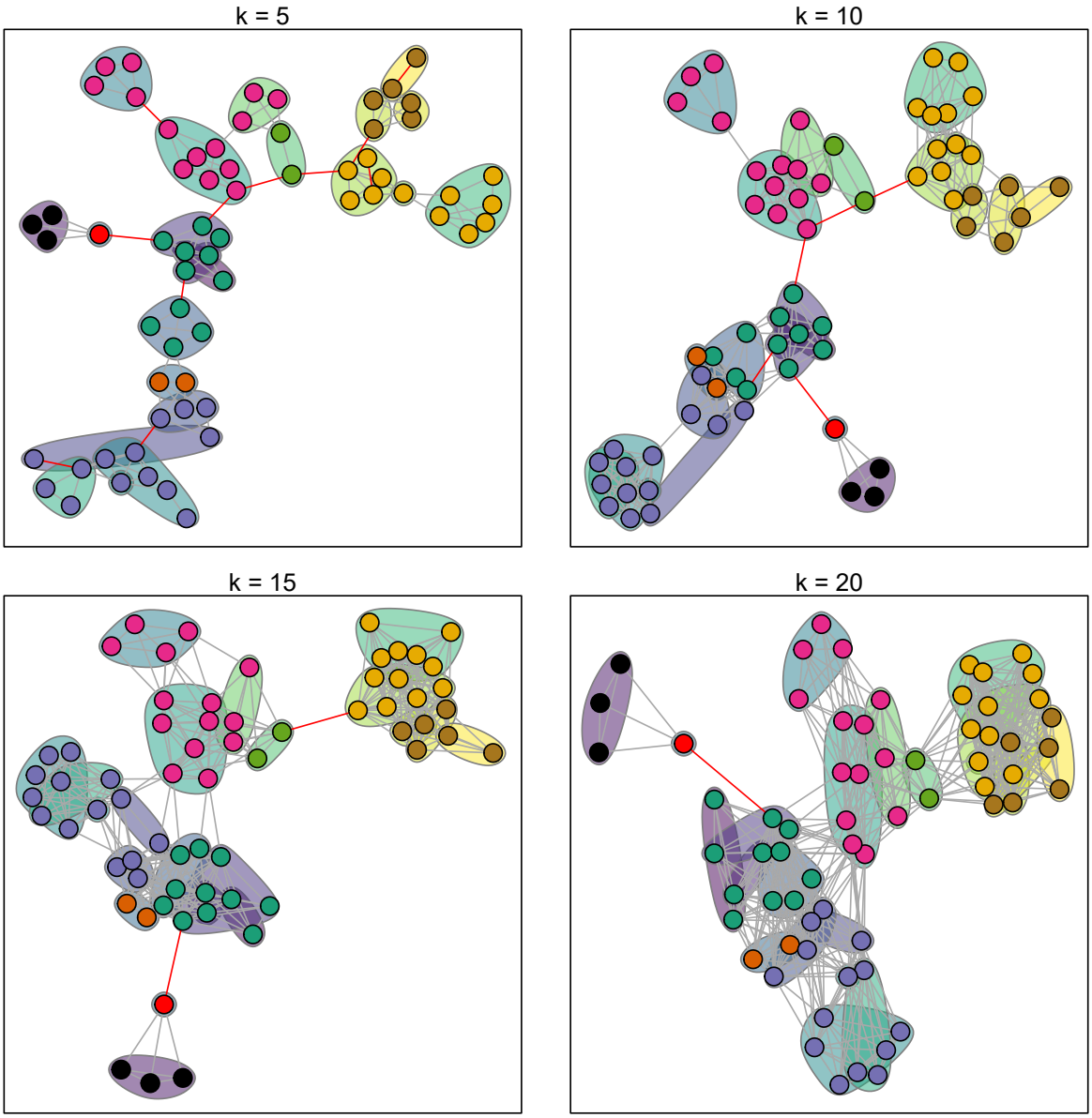

Figure 3. Genetic distance network visualisation. Panels show the NetView networks for different values of k (no. of allowed connections per node) based on dxy. Nodes represent individuals and colours reflect prior population assignments (Fig. 2). The shaded regions envelope individuals located geographically close to each other, with the shading representing mean latitude (Viridis colour scale; lighter colours are more northerly). Red edges are unique to the minimum spanning tree. Note how the finer scale differences are more noticeable for small values of k, becoming obscured by the broader patterns as k increases.

these were the only individuals recovered as F2 hybrids between these groups by NewHybrids (Fig. 4). These individuals also fell between *E. a. abietina* and *E. a. atrorosea* (North) in PCA space and in all NetView graphs. Additionally, while sNMF strongly supported the distinctness of *E. a. abietina* from *E. a. atrorosea* (North), the reverse did not apply. Only six *E. a. atrorosea* (North) individuals had <10% of their ancestry derived from the *E. a. abietina* parental population, and these consisted of the five southernmost *E. a. atrorosea* (North) samples plus one from Cecelia Forest in the central part of the *E. a. atrorosea* (North) range. The other eight *E. a. atrorosea* (North) individuals, which were not originally suspected of having mixed ancestry, had >10% of their ancestry derived from the *E. a. abietina* parental population. Of these, six individuals – including three from the northernmost sampling site at Newlands Forest and three from Cecelia Forest – were inferred by NewHybrids to be *E. a. atrorosea* (North)

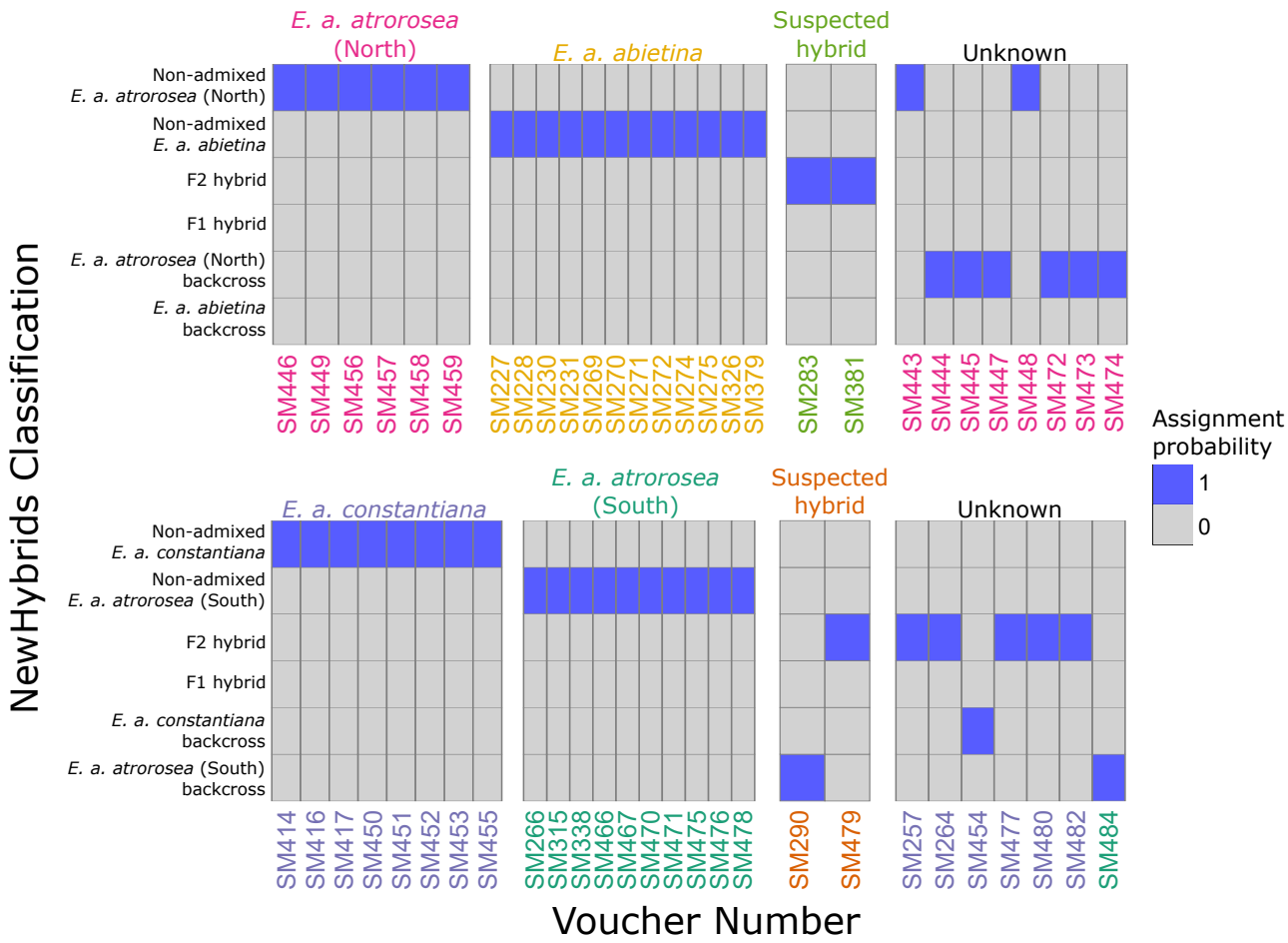

Figure 4. Hybrid assignment results: Posterior assignment probabilities of the various hybrid classes identifiable by NewHybrids. *Top*: *E. a. atrorosea* (North) x *E. a. abietina*; *Bottom*: *E. a. atrorosea* (South) x *E. a. constantiana*. Text colours follow Fig. 2.

backcrosses, while the remaining two individuals were inferred to be non-admixed *E. a. atrorosea* (North) (Fig. 4).

The results were also complex for the analysis of hybridisation between *E. a. atrorosea* (South) and *E. a. constantiana*. According to sNMF, only one out of 11 individuals originally identified as *E. a. atrorosea* (South) had mixed ancestry, which NewHybrids recovered as an *E. a. atrorosea* (South) backcross. In contrast, six of the 14 individuals identified as *E. a. constantiana* had mixed ancestry. Five of these were recovered as F2 hybrids, including all from Blackburn Ravine and two from nearby Silvermine area just to the east, while one individual collected from the more northerly Vlakkenberg area was recovered as a *E. a. constantiana* backcross. Of the two individuals originally suspected to be hybrids between the two subspecies, one was recovered as an F2 hybrid and the other as an *E. a. atrorosea* (South) backcross.

Phylogenetic relationships

All individuals included. Figure S3 depicts the results of the phylogenetic analysis with admixed individuals included. All four IQ-TREE runs returned virtually identical log-likelihood values. Overall, the tree showed successive branching forming a highly unbalanced ("ladder-like") tree, which appeared to be significantly influenced by the presence of individuals inferred to have mixed ancestry based on sNMF and NewHybrids results. At the same time, populations identified by the analyses of population structure were readily apparent and often formed monophyletic

clades. *E. a. abietina* and *E. a. diabolis* formed a well-supported clade and were confidently resolved as reciprocally monophyletic. Of the two individuals identified as *E. a. atrorosea* (North) *x E. a. abietina* F2 hybrids, one occupied a position clearly intermediate between the two populations, while the other fell within a clade containing most of the *E. a. atrorosea* (North) backcrosses and one non-admixed *E. a. atrorosea* (North) individual. The rest of the *E. a. atrorosea* (North) individuals (seven non-admixed and two backcrosses) formed a clade further towards more distant, older nodes. The individuals identified as non-admixed *E. a. constantiana* formed a well-supported clade with the inclusion of the single *E. a. constantiana* backcross individual, while all the non-admixed *E. a. atrorosea* (South) individuals not collected from Blackburn Ravine also formed a well-supported clade. A relatively poorly supported clade (bootstrap = 90%, SH-alrt = 75%) grouping between the non-admixed *E. a. atrorosea* (South) and *E. a. constantiana* clades consisted of all individuals from Blackburn Ravine regardless of prior identification. Within this clade, however, individuals identified as *E. a. atrorosea* (South) and *E. a. constantiana* each formed sub-clades that had good bootstrap support but poor SH-alrt support, and which each contained one of the individuals identified a priori as being of hybrid origin.

Admixed individuals excluded. Figure 5 depicts the results of the phylogenetic analysis with admixed individuals excluded. The two best IQ-TREE runs returned similar log-likelihood values (run 2: -8720997.954, run 3: -8721638.908, difference = 640.954). Most nodes of the maximum-likelihood tree received high bootstrap and SH-alrt support, particularly at deeper phylogenetic levels at which populations were distinguished. The most notable exception was the node subtending (*E. a. atrorosea* [North],(*E. a. abietina*,*E. a. diabolis*)), which had very low support values, meaning that the placement of *E. a. atrorosea* (North) could not be resolved. The three *E. a. atrorosea* (South) individuals from Blackburn Ravine that were not identified as being of recent hybrid origin were nevertheless recovered in an intermediate position between non-admixed *E. a. atrorosea* (South) and *E. a. constantiana*. Given that all other individuals collected from this locality were marked as putative hybrid-origin, this may indicate that these individuals have mixed ancestry of too ancient origin to have been detected by the sNMF and NewHybrids analyses.

Population differentiation and phylogeny

The PoMo tree (Fig. 6) had the same topology as the individual-level tree with respect to population-level groupings, except that $ATRO_N$ was recovered as sister to $CONS$, but with low support (bootstrap support = 72). F_{ST} was on average much higher at the species level than it was between populations within *E. abietina* (Fig. 6). Within *E. abietina*, differentiation was lowest between ABI and $DIAB$ and highest between $CONS$ and $DIAB$.

Evidence for ancient introgression

The ADMIXTOOLS2 graph search analysis results were effectively identical regardless of whether the starting tree was specified or not. When the number of allowed admixed edges (N_{ADMIX}) was zero the best graph topology was identical to that of the PoMo tree in all ten runs (Fig. 7, A). With $N_{ADMIX} = 1$, all runs converged on the same optimal topology and essentially identical edge weights. This graph showed the ancestor of ABI and $DIAB$ as an admixed edge with most (76%) of its ancestry derived from the ancestor of (i.e., the edge subtending) $ATRO_N$ and the rest derived from the ancestor of *E. abietina* as a whole (Fig. 7, B).

With $N_{ADMIX} = 2$, the independent runs found two distinct optimal graphs. The best-scoring graph

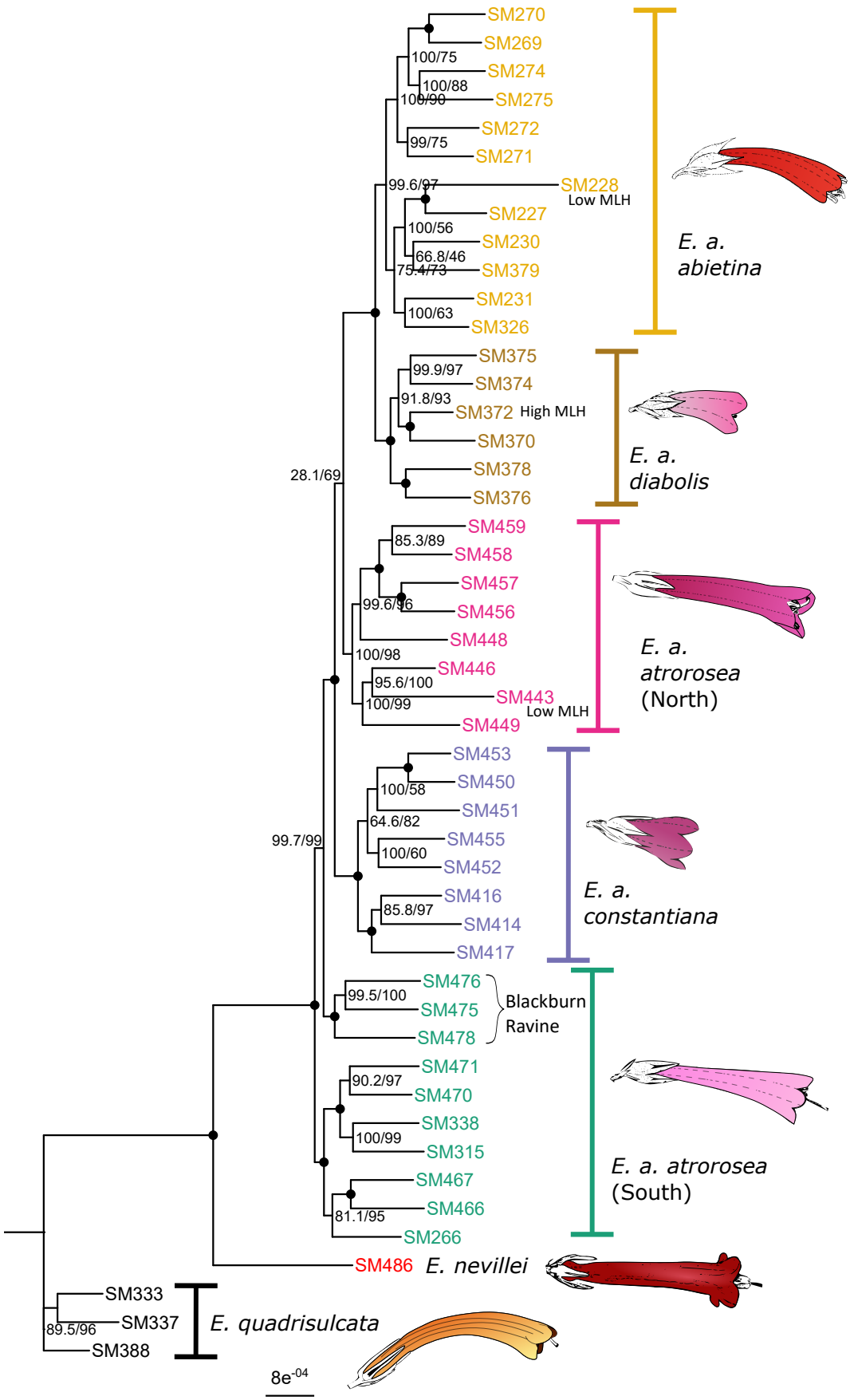

Figure 5. The maximum-likelihood phylogeny inferred by IQ-TREE with admixed individuals excluded. Node values indicate bootstrap/SH-alrt support, and black circles indicate full support from both measures. Text colours follow Fig. 2. Note that *E. a. atrorosea* (North) individuals form a well-supported clade but that the clade's position in the phylogeny is unresolved. Individuals with anomalous heterozygosity and those from Blackburn Ravine are annotated.

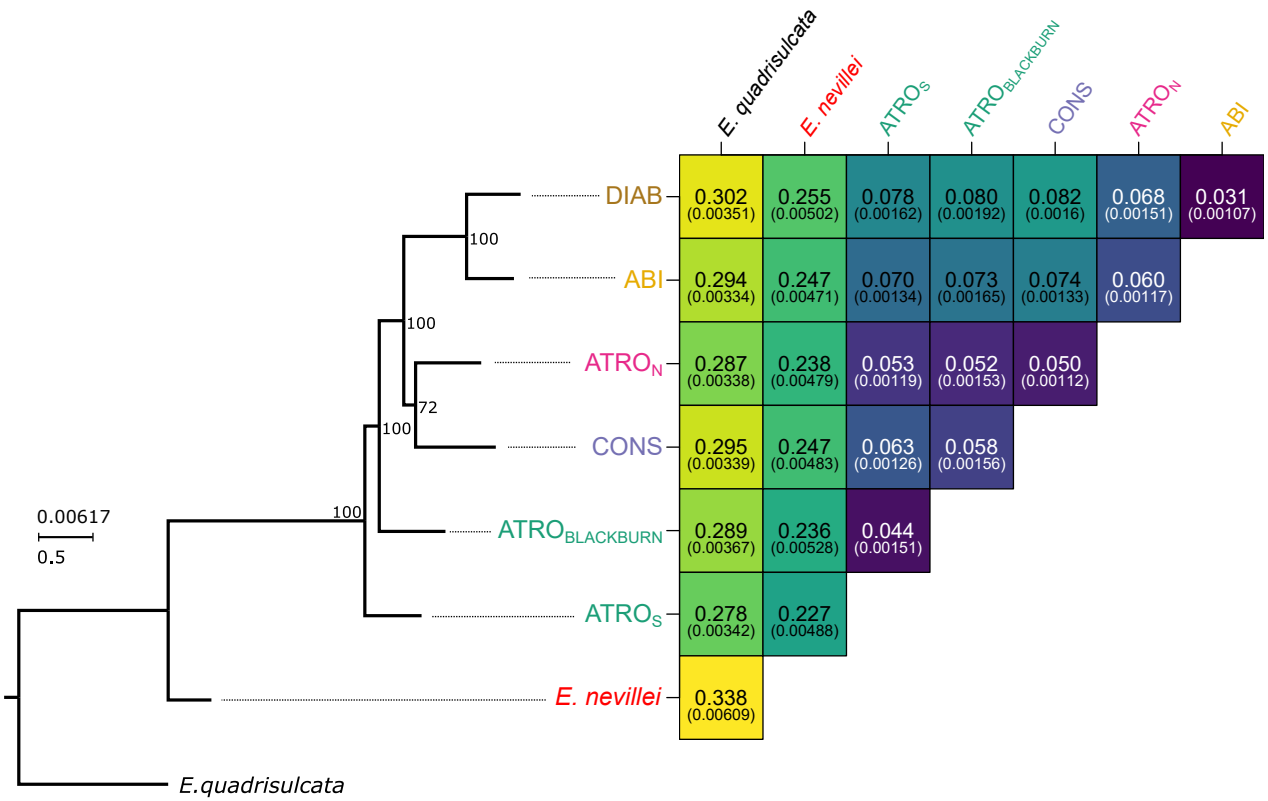

Figure 6. *Left*: population-level phylogeny estimated by PoMo, an incomplete lineage sorting-aware tree inference method. The tree scale shows expected substitutions per site (top) and coalescent units (bottom). Branch supports are ultrafast bootstrap. Note that the tree topology differs from that in Fig. 5 in the position of *E. a. atrorosea* (North), likely due to the exclusion from this analysis of putatively admixed individuals. *Right*: Pairwise F_{ST} between populations, with standard deviation in brackets. Darker colours indicate lower values.

(Fig. 7, C), found in two runs, retained the admixed edge found with $N_{ADMIX} = 1$ and additionally recovered $ATRO_{BLACKBURN}$ as descending from an admixed edge with equal ancestry derived from the ancestors of $ATRO_S$ and $CONS$. The next-best graph (Fig. 7, D) depicted a more complex history of admixture, with one admixed edge subtending $CONS$ and $ATRO_N$ deriving 80% of its ancestry from the ancestor of $ATRO_{BLACKBURN}$ and 20% from a "ghost" lineage that was sister to the rest of *E. abietina*. This same "ghost" lineage contributed 32% of its ancestry to an admixed edge subtending ABI and $DIAB$, which derived the rest of its ancestry from the ancestor of $ATRO_N$.

The best $N_{ADMIX} = 1$ graph had greater predictive power than the best $N_{ADMIX} = 0$ graph (mean score difference = -150.8 ± 39.5 SD, z = -3.82, p < 0.001). The two best $N_{ADMIX} = 2$ graphs had similar predictive power (mean score difference = -11.7 ±11.3 SD, z = -1.04, p = 0.30), however, neither had better predictive power than the best $N_{ADMIX} = 1$ graph (best graph: mean score difference = -31.0 ±17.7 SD, z = -1.75, p = 0.08; second-best graph: mean score difference = -19.3 ±15.2 SD, z = -1.27, p = 0.20). These results point to a single admixed edge as the most parsimonious depiction of the reticulate history of the group.

Discussion

Taxonomy and cryptic diversity

Given that genetic divergence between established species (e.g., between *E. abietina* and *E. nevillei*) was found to be much higher than any pairwise divergence between the various populations of *E. abietina* (Fig. 6), it is reasonable to conclude that *E. abietina* constitutes a single – if highly variable – species. On the other hand, there is mixed support for the current subspecific treatment of the complex. In particular, what is currently recognised as *E. a. atrorosea* appears to consist of two non-sister lineages that are geographically isolated but phenotypically very similar. Such "cryptic" diversity has come to be detected frequently in a range of organisms (Bickford et al., 2007), particularly since the advent of Next Generation Sequencing-based genotyping in non-model organisms (e.g., Hinojosa et al., 2019; Blair et al., 2019; Lutsak, 2020; Boucher et al., 2021; Daïnou et al., 2016). The *E. a. atrorosea* (South) lineage appears to be widespread on sandstone-derived soils south of Table Mountain, whereas *E. a. atrorosea* (North) is seemingly restricted to the granite- and shale-derived soils of the eastern slopes of Table Mountain and Constantiaberg (see Fig. 1). None of our results suggested current or ancient hybridization between these two lineages, lending support to their independence.

Floral trait evolution

Despite the close relationships and weak reproductive barriers between the populations of *E. abietina*, shifts between pollination syndromes (as approximated by flower tube length) occurred at least twice in the complex. Given the phylogenetic results, the most parsimonious reconstruction of floral trait evolution is that long tubes are plesiomorphic and short tubes were derived in *E. a. constantiana* and *E. a. diabolis* independently. Such a pattern, combined with incomplete reproductive isolation between sympatric short- and long-tubed populations (in the case of *E. a. constantiana* and *E. a. atrorosea* [South]), implies that strong selective forces drive and maintain shifts to insect pollination.

What might drive shifts to different modes of pollination? At least for *E. a. constantiana*, the Cape Honeybee (*Apis mellifera* subsp. *capensis*) seems likely to be the primary pollinator based on numerous personal observations (e.g., `https://www.inaturalist.org/observations/26399879`, `https://www.inaturalist.org/observations/131065480`). Shifting from bird to bee pollination could be interpreted as a shift to a less specialised pollination system (Cronk and Ojeda, 2008). Johnson (1996), for example, suggested that sunbirds are more reliable pollinators than insects and that this, in addition to the relatively low abundance of insects in fynbos, especially at higher elevations, drove frequent shifts to sunbird pollination in many CFR lineages. However, van der Niet et al. (2020) showed that individual honeybees visiting two co-occurring Cape *Erica* species exhibited remarkable floral constancy, in that during foraging bouts they tended to consistently prefer one or the other species rather than visiting both. Such constancy presumably explained the authors' finding of extremely low rates of interspecific pollen transfer. Based on a detailed analysis of a community of co-occurring insect-pollinated *Erica*, Bouman et al. (2017) suggested that, rather than incurring a fitness cost, bee pollination may instead benefit co-occurring species by enabling them to collectively attract pollinators while simultaneously avoiding cross-species pollen flow. Such benefits may explain the independent evolution of short-tubed flowers in *E. a. constantiana* and *E. a. diabolis*. Overall, our results highlight the variability of selection on floral traits and point to *Erica* as a whole being highly sensitive to such selection.

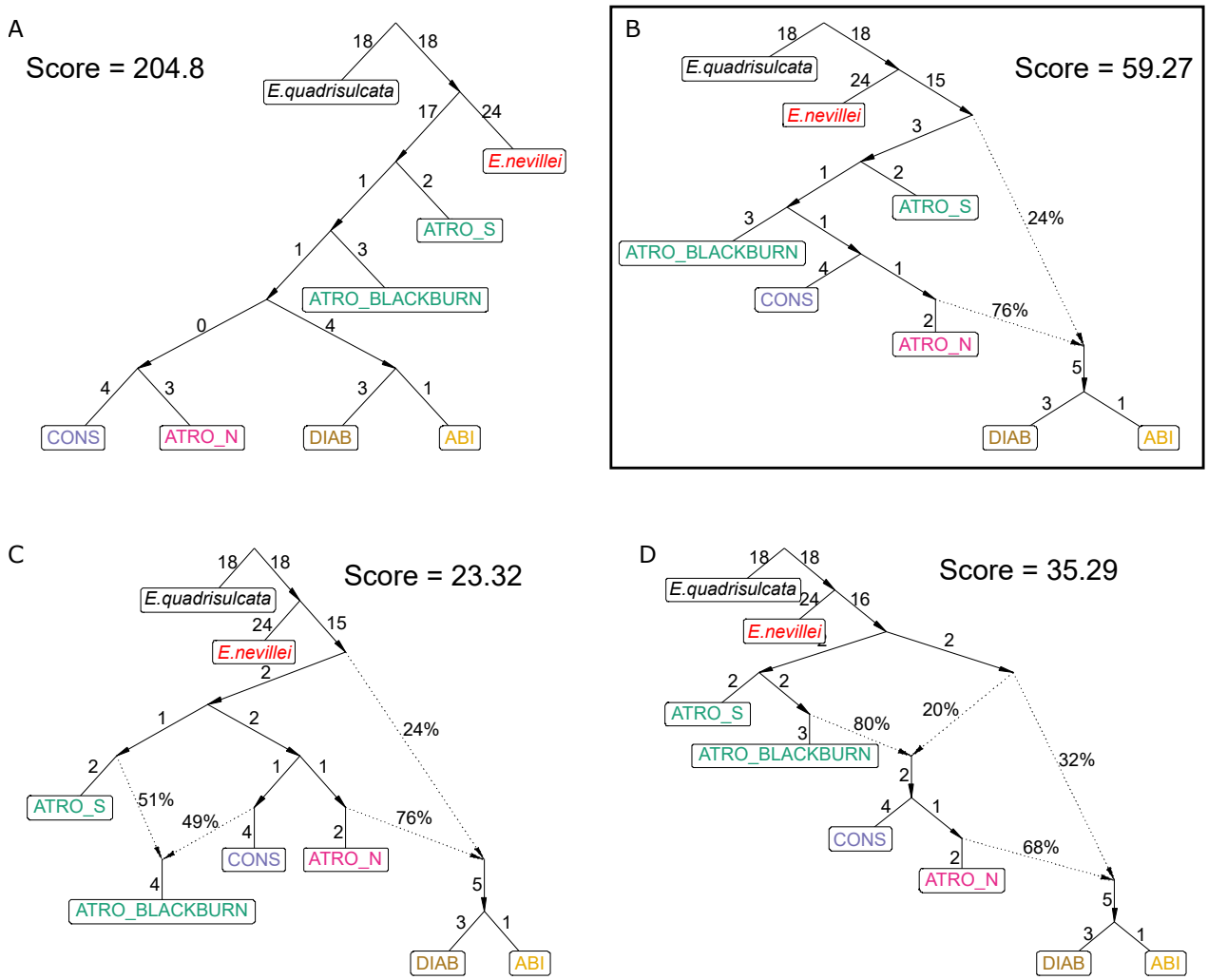

Figure 7. The four top-scoring admixture graphs according to ADMIXTOOLS2 for $N_{ADMIX} = 0$ (A), 1 (B), and 2(C, D). Numbers above solid non-admixed edges are proportional to their weight, while percentages above dashed admixture edges indicate their contribution to the admixed edge. The graph in panel B had the best performance based on its predictive power being similar to those in panels C and D while only having one admixture edge.

Despite the evident lability of floral traits in *Erica*, their link to the genus's diversification is less certain. Our results showed that pollination syndrome differentiation cannot be assumed to be a reliable indicator of reproductive isolation. In the CFR, floral tube length variation has been shown to cause intraspecific reproductive isolation in *Erica* (Newman and Johnson, 2021) and other angiosperm groups (e.g., Minnaar et al., 2019). However, a global-scale meta-analysis showed that floral trait divergence on its own is a poor predictor of speciation, and that it instead almost always acts in concert with other factors – such as geographic and habitat isolation – to effect speciation (Kay and Sargent, 2009). Our results support this assertion in that all but one individual showing mixed ancestry between *E. a. constantiana* and *E. a. atrorosea* (South) occurred at the same locality (Blackburn Ravine), whereas putatively non-admixed individuals occurred in areas where only short- or long-tubed individuals, but not both, are present. We therefore suggest that floral trait divergence between *E. a. constantiana* and *E. a. atrorosea* (South) and between *E. a. diabolis* and *E. a. abietina* resulted from a combination of pollination-driven selection and geographic isolation resulting in genetic drift. The cryptic diversity within *E. a. atrorosea* – which we suggest was driven by geographic isolation coupled with adaptation of the northern lineage to relatively nutrient-rich soils – further illustrates that divergence can occur without a pollinator shift.

Hybridization and introgression

Although *E. abietina* can be grouped into genetically distinct, geographically separated and phenotypically recognisable units, these cannot be said to represent completely isolated lineages. The discovery of several putative hybrids of recent origin – which occurred at localities where more than one subspecies was present or where their ranges met (Fig. 1, 4) – shows that gene flow between certain populations is ongoing and probably frequent. Interspecific hybridisation is well-known in Cape *Erica* (Oliver and Oliver, 2002, 2005), and does not always involve closely related species (e.g., Oliver, 1986). Interestingly, in this study none of the putative hybrid individuals was inferred to be first-generation and most were backcrosses. This implies that hybrids are fertile and liable to generate "hybrid swarms". Such a hybrid swarm may exist at Blackburn Ravine where *E. a. atrorosea* (South) and *E. a. constantiana* appear to have formed a population with a large proportion of admixed individuals capable of backcrossing with either parental lineage.

In contrast, recent introgression between *E. a. abietina* and *E. a. atrorosea* (North) seems to have been unidirectional. This was particularly evident from the sNMF analyses, in which all *E. a. abietina* individuals were inferred to be genetically non-admixed, whereas *E. a. atrorosea* (North) individuals showed a pattern of increasingly mixed ancestry from south to north. This could reflect a case of recent secondary contact between these lineages; however, this would imply recent range expansion in one or both of these populations. This seems unlikely given that genetic diversity (measured by heterozygosity) did not show a clear relationship with geography in either of them (Fig. S1). Nevertheless, denser geographic sampling of both populations enabling formal analyses aimed at detecting range expansions (e.g., He et al., 2017; Peter and Slatkin, 2013) would be a valuable endeavour. Alternatively, the pattern may instead reflect long-standing gene flow from *E. a. abietina* into *E. a. atrorosea* (North) that has not been sufficient to cause gene swamping (Lenormand, 2002; Bridle and Vines, 2007). *Erica a. abietina* is likely to be adapted to extremely nutrient-poor sandstone-derived soils such as exist on Table Mountain's upper plateau (Compton, 2004), and such specialisation has been shown to limit the ability of fynbos plants to adapt to more nutrient-rich soils (Verboom et al., 2017). This may mean that *E. a. abietina* alleles are maladaptive for *E. a. abietina* x *E. a. atrorosea* (North) hybrids occupying the lower eastern slopes – whose soils are relatively rich in nutrients (Compton, 2004; Cramer et al., 2018, Fig. 1) – resulting in what has been termed "migration-selection equilibrium" (Lenormand, 2002). Studies investigating the factors that determine the geographic range limits of these populations (e.g., local adaptation, competition, dispersal limitation; Gaston, 2009) would help to differentiate these possibilities.

Evidence of present-day hybridisation does not necessarily mean that introgression has influenced diversification (e.g., Kessler et al., 2022; Jordan et al., 2017; Westbury et al., 2019), however, our analysis based on *f3*-statistics provides good evidence that *E. abietina* does have a reticulate evolutionary history. Perhaps the most biologically sensible interpretation of the best-fitting admixture graph (Fig. 7, B) is that *E. a. abietina*/*E. a. diabolis* split from the rest of the complex while simultaneously deriving some (roughly one quarter) of their genetic makeup from a more distant relative ancestral to or close to *E. nevillei*. This interpretation implies the influence of a "ghost" lineage: one that is unknown, unsampled or extinct and which interbred historically with an extant lineage (Ottenburghs, 2020). Ghost introgression appears to be common across the tree of life (e.g., Meyer et al., 2012; Green et al., 2010; Barlow et al., 2018; Maier et al., 2022) and may be an important driver of diversification in rapidly evolving lineages (Ottenburghs, 2020). *Erica nevillei* and *E. quadrisulcata* are both narrow-range endemics confined to rocky outcrops

at relatively high elevations (Oliver and Oliver, 2002). Given this, it is surprising that the seemingly suitable upper plateau of Table Mountain does not host these or a related species with a similar niche. However, assuming that this absence is the result of extinction could explain why ghost introgression is inferred. Such speculation could be tested with improved sampling of the relevant populations and a more thorough genotyping method (such as whole-genome sequencing) which would allow for more detailed and powerful analyses (e.g., Mondal et al., 2019), including those that can detect adaptive introgression of genomic regions (e.g., Racimo et al., 2015).

Conclusions

This study adds to the small but growing and much-needed body of literature focused on understanding speciation in action in the Fynbos flora (Ellis et al., 2014; Barraclough, 2006; Prunier and Holsinger, 2010; Prunier et al., 2017; Lexer et al., 2014). We find that the subspecific classification of *E. abietina* is only a partial reflection of its evolutionary history. Notably, there is evidence of cryptic diversity in the complex, which may have gone unrecognised due to a focus on floral traits in the taxonomic literature (Oliver and Oliver, 2002). We also find that floral trait divergence is likely to be driven by strong selective forces but that it is unlikely to drive lineage divergence without the action of additional factors, particularly geographic isolation. Finally, apart from ongoing gene flow in the complex revealed by the presence of subspecific hybrids, we find evidence that ancient gene flow may have also influenced the complex's evolution in the form of "ghost" introgression from a now-extinct lineage. Exploring the possibility of links between introgression and floral trait evolution in *Erica* may provide important insights into the spectacular diversification of the genus (see e.g., Nelson et al., 2021). Overall our results paint a picture of a highly dynamic system whose evolutionary history has been shaped by diverse, interacting forces (Donoghue and Sanderson, 2015). Refining this picture will undoubtedly further our understanding of diversification in *Erica* and beyond.

Acknowledgements

Primary funding for the project was provided by the Deutsche Forschungsgemeinschaft (PI 1169/1-2). The Genomics Core Facility (GCF) at the University of Bergen, which is a part of the NorSeq consortium, provided services on whole-genome sequencing of *Erica cerinthoides*; GCF is supported in part by major grants from the Research Council of Norway (grant no. 245979/F50), Bergen Research Foundation (BFS) (grant no. BFS2017TMT04 and BFS2017TMT08), and Trond Mohn Foundation (TMS). Computing facilities were provided by the University of Cape Town's ICTS High Performance Computing team (`hpc.uct.ac.za`) and the Centre for High Performance Computing at the University of Bayreuth (`bzhpc.uni-bayreuth.de`). Collections were made under permits from CapeNature (CN35-31-8281) and SANParks (CRC/2019-2020/004–2019/V1). Voucher specimens were deposited at the Compton Herbarium (NBG).

Competing interests

The authors declare no competing interests.

Author contributions

NN and SM conceived the study. SM collected samples, did the lab work, and analysed the data. SM wrote the initial draft with input by MP and NN. All authors reviewed the final manuscript.

Data availability

Raw reads will be deposited in GenBank and other associated data including genotypic data, metadata and code will be made available upon publication.

References

Adamson, R. (1959). Notes on the Phytogeography of the Flora of the Cape Peninsula. *Transactions of the Royal Society of South Africa*, 35(5):443–462.

Anderson, E. C. (2008). Bayesian inference of species hybrids using multilocus dominant genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1505):2841–2850.

Barlow, A., Cahill, J. A., Hartmann, S., Theunert, C., Xenikoudakis, G., Fortes, G. G., Paijmans, J. L., Rabeder, G., Frischauf, C., Grandal-d'Anglade, A., et al. (2018). Partial genomic survival of cave bears in living brown bears. *Nature Ecology & Evolution*, 2(10):1563–1570.

Barnes, K., Nicolson, S. W., and Van Wyk, B. E. (1995). Nectar sugar composition in *Erica*. *Biochemical Systematics and Ecology*, 23(4):419–423.

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., and Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, 27(12):1691–1692.

Barraclough, T. G. (2006). What can phylogenetics tell us about speciation in the Cape flora? *Diversity & Distributions*, 12(1):21–26.

Bellstedt, D. U., Pirie, M. D., Visser, J. C., de Villiers, M. J., and Gehrke, B. (2010). A rapid and inexpensive method for the direct PCR amplification of DNA from plants. *American Journal of Botany*, 97(7).

Bergh, N. G., Verboom, G. A., Rouget, M., and Cowling, R. M. (2014). Vegetation types of the Greater Cape Floristic Region. In *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region*. Oxford University Press.

Bhatia, G., Patterson, N., Sankararaman, S., and Price, A. L. (2013). Estimating and interpreting FST: The impact of rare variants. *Genome Research*, 23(9):1514–1521.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., and Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, 22(3):148–155.

Blair, C., Bryson, R. W., Linkem, C. W., Lazcano, D., Klicka, J., Mccormack, J. E., Klicka, J., Lazcano, D., Linkem, C. W., Bryson, R. W., and Blair, C. (2019). Cryptic diversity in the Mexican highlands: Thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: Crotalus). *Molecular Ecology Resources*, 19(2):349–365.

Boucher, F. C., Dentant, C., Ibanez, S., Capblancq, T., Boleda, M., Boulangeat, L., Smyčka, J., Roquet, C., and Lavergne, S. (2021). Discovery of cryptic plant diversity on the rooftops of the Alps. *Scientific Reports*, 11(1).

Bouman, R. W., Steenhuisen, S. L., and Van Der Niet, T. (2017). The role of the pollination niche in community assembly of *Erica* species in a biodiversity hotspot. *Journal of Plant Ecology*, 10(4):634–648.

Bradshaw, P. L. and Cowling, R. M. (2014). Landscapes, rock types, and climate of the Greater Cape Floristic Region. In *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region*. Oxford University Press.

Bridle, J. R. and Vines, T. H. (2007). Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution*, 22(3):140–147.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17):i884–i890.

Coetzee, A. (2016). *Nectar distribution and nectarivorous bird foraging behaviour at different spatial scales*. PhD thesis, Stellenbosch: Stellenbosch University.

Coetzee, A., Spottiswoode, C. N., and Seymour, C. L. (2020). Post-pollination barriers enable coexistence of pollinator-sharing ornithophilous *Erica* species. *Journal of Plant Research*, 133(6):873–881.

Compton, J. S. (2004). *The rocks and mountains of Cape Town*. Juta and Company Ltd.

Cowling, R. M., MacDonald, I., and Simmons, M. (1996). The Cape Peninsula, South Africa: physiographical, biological and historical background to an extraordinary hot-spot of biodiversity. *Biodiversity & Conservation*, 5(5):527–550.

Cramer, M. D., Power, S. C., Belev, A., Gillson, L., Bond, W. J., Hoffman, M. T., and Hedin, L. O. (2018). Are forest-shrubland mosaics of the Cape Floristic Region an example of alternate stable states? *Ecography*, 42(4):717–729.

Cronk, Q. and Ojeda, I. (2008). Bird-pollinated flowers in an evolutionary and molecular context. *Journal of Experimental Botany*, 59(4):715–727.

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *Interjournal*, Complex Systems:1695.

Daïnou, K., Blanc-Jolivet, C., Degen, B., Kimani, P., Ndiade-Bourobou, D., Donkpegan, A. S. L., Tosso, F., Kaymak, E., Bourland, N., Doucet, J.-L., and Hardy, O. J. (2016). Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences – a case study in the tree genus *Milicia*. *BMC Evolutionary Biology*, 16(1):259.

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., and Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, 10(2):giab008.

Donoghue, M. J. and Sanderson, M. J. (2015). Confluence, synnovation, and depauperons in plant diversification. *New Phytologist*, 207(2):260–274.

Ellis, A. G., Verboom, G. A., van der Niet, T., Johnson, S. D., Linder, H. P., Allsopp, N., Colville, J. F., and Marrone, P. (2014). Speciation and extinction in the greater Cape Floristic Region. In Allsopp, N., Colville, J. F., and Verboom, G. A., editors, *Fynbos: Ecology, evolution, and conservation of a Megadiverse Region*, chapter 6, pages 119–141. Oxford University Press, Oxford.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One*, 6(5):1–10.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048.

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, 17(1):27–32.

Frichot, E. and François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8):925–929.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, 196(4):973–983.

Garrison, E. (2012). Vcflib: A C++ library for parsing and manipulating VCF files. *Github Https://github. Com/ekg/vcflib*.

Garrison, E. and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv:1207.3907*.

Gaston, K. J. (2009). Geographic range limits: achieving synthesis. *Proceedings of the Royal Society B: Biological Sciences*, 276(1661):1395–1406.

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F -statistics. *Molecular Ecology Notes*, 5:184–186.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., et al. (2010). A draft sequence of the Neandertal genome. *Science*, 328(5979):710–722.

Gruber, B., Unmack, P. J., Berry, O. F., and Georges, A. (2018). DartR: An R package to facilitate analysis of SNP data

generated from reduced representation genome sequencing. *Molecular Ecology Resources*, 18(3):691–699.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3):307–321.

He, Q., Prado, J. R., and Knowles, L. L. (2017). Inferring the geographic origin of a range expansion: Latitudinal and longitudinal coordinates inferred from genomic data in an ABC framework with the program X-ORIGIN. *Molecular Ecology*, 26(24):6908–6920.

Helme, N. A. and Trinder-Smith, T. H. (2006). The endemic flora of the Cape Peninsula, South Africa. *South African Journal of Botany*, 72(2):205–210.

Hinojosa, J. C., Koubínová, D., Szenteczki, M., Pitteloud, C., Dincă, V., Alvarez, N., and Vila, R. (2019). A mirage of cryptic species: genomics uncover striking mito-nuclear discordance in the butterfly *Thymelicus sylvestris*. *Molecular Ecology*, 28(17):3857–3868.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2):518–522.

Hudson, R. R., Slatkin, M., and Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, 132(2):583–589.

Hughes, C. and Eastwood, R. (2006). Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences*, 103(27):10334–10339.

Inglis, P. W., Pappas, M. d. C. R., Resende, L. V., and Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *Plos One*, 13(10):e0206085.

Johnson, S. D. (1996). Pollination, adaptation and speciation models in the Cape flora of South Africa. *TAXON*, 45(1):59–66.

Jombart, T. and Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21):3070–3071.

Jordan, C. Y., Lohse, K., Turner, F., Thomson, M., Gharbi, K., and Ennos, R. A. (2017). Maintaining their genetic distance: little evidence for introgression between widely hybridising species of *Geum* with contrasting mating systems. *Molecular Ecology*, 27:1214–1228.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6):587–589.

Kamada, T., Kawai, S., et al. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15.

Kay, K. M. and Sargent, R. D. (2009). The Role of Animal Pollination in Plant Speciation: Integrating Ecology, Geography, and Genetics. *The Annual Review of Ecology Evolution and Systematics*, 40(September):637–656.

Kessler, C., Wootton, E., and Shafer, A. B. (2022). Speciation without gene-flow in hybridizing deer. *Biorxiv Evolutionary Biology*.

Knaus, B. J. and Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1):44–53.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, pages 1179–1191.

Korunes, K. L. and Samuk, K. (2021). PIXY: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources*, 21(4):1359–1368.

Kruger, F. and Taylor, H. (1980). Plant species diversity in cape fynbos: gamma and delta diversity. *Vegetatio*, 41:85–93.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4):183–189.

Lexer, C., Wüest, R. O., Mangili, S., Heuertz, M., Stölting, K. N., Pearman, P. B., Forest, F., Salamin, N., Zimmermann, N. E., and Bossolini, E. (2014). Genomics of the divergence continuum in an African plant biodiversity hotspot, I: Drivers of population divergence in *Restio capensis* (Restionaceae). *Molecular Ecology*, 23(17):4373–4386.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Arxiv:1303.3997*.

Lutsak, T. (2020). Coalescence-based species delimitation using genome-wide data reveals hidden diversity in a cosmopolitan

group of lichens. *Organisms Diversity and Evolution*, 20(2):189–218.

Madriñán, S., Cortés, A. J., and Richardson, J. E. (2013). Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Frontiers in Genetics*, 4.

Maier, R., Flegontov, P., Flegontova, O., Changmai, P., and Reich, D. (2022). On the limits of fitting complex models of population history to genetic data. *bioRxiv*.

Manning, J. and Goldblatt, P. (2012). *Plants of The Greater Cape Floristic Region 1: The Core Cape Flora*, volume 29. South African National Biodiversity Institute, Pretoria.

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, 27(2):209–220.

Meier, R. (2017). Citation of taxonomic publications: the why, when, what and what not. *Systematic Entomology*, 42(2):301–304.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., De Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 338(6104):222–226.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., Lanfear, R., and Teeling, E. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5):1530–1534.

Minnaar, C., de Jager, M. L., and Anderson, B. (2019). Intraspecific divergence in floral-tube length promotes asymmetric pollen movement and reproductive isolation. *New Phytologist*, 224(3):1160–1170.

Mondal, M., Bertranpetit, J., and Lao, O. (2019). Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nature Communications*, 10(1):1–9.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772):853–858.

Nelson, T. C., Stathos, A. M., Vanderpool, D. D., Finseth, F. R., wu Yuan, Y., and Fishman, L. (2021). Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation *Mimulus* section *Erythranthe*. *PLoS Genetics*, 17(2):e1009095.

Newman, E. and Johnson, S. D. (2021). A shift in long-proboscid fly pollinators and floral tube length among populations of *Erica junonia* (Ericaceae). *South African Journal of Botany*, 142:451–458.

Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O'Hara, R., Simpson, G., and Solymos, P. (2019). vegan: Community Ecology Package. R package version 2.5-4.

Oliver, E. G. H. (1986). The identity of *Erica vinacea* and notes on hybridization in *Erica*. *Bothalia*, 16(1):35–38.

Oliver, E. G. H. and Oliver, I. M. (2002). The genus *Erica* (Ericaceae) in southern Africa: Taxonomic notes 1. *Bothalia*, 32(1):37–61.

Oliver, E. G. H. and Oliver, I. M. (2005). The genus *Erica* (Ericaceae) in southern Africa: taxonomic notes 2. *Bothalia*, 35(2):121–148.

Ortiz, E. M. (2019). vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. https://doi.org/10.5281/zenodo.2540861.

Ottenburghs, J. (2020). Ghost Introgression: Spooky Gene Flow in the Distant Past. *BioEssays*, 42(6):2000012.

Paradis, E. and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35:526–528.

Peter, B. M. (2016). Admixture, Population Structure, and F-Statistics. *Genetics*, 202(4):1485–1501.

Peter, B. M. and Slatkin, M. (2013). Detecting range expansions from genetic data. *Evolution*, 67(11):3274–3289.

Pirie, M. D., Kandziora, M., Nürk, N. M., Le Maitre, N. C., Mugrabi De Kuppler, A., Gehrke, B., Oliver, E. G., and Bellstedt, D. U. (2019). Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotemperate by *Erica*. *BMC Evolutionary Biology*, 19(1):0–26.

Pirie, M. D., Oliver, E. G., Gehrke, B., Heringer, L., De Kuppler, A. M., Le Maitre, N. C., and 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.

Pirie, M. D., Oliver, E. G. H., and Bellstedt, D. U. (2011). A densely sampled ITS phylogeny of the Cape flagship genus *Erica* L. suggests numerous shifts in floral macro-morphology. *Molecular Phylogenetics and Evolution*, 61(2):593–601.

Pirie, M. D., Oliver, E. G. H., Kuppler, A. M. D., Gehrke, B., Maitre, N. C. L., Kandziora, M., and Bellstedt, D. U. (2016). The biodiversity hotspot as evolutionary hot-bed: spectacular radiation of *Erica* in the Cape Floristic Region. *Bmc Evolutionary Biology*.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155:945–959.

Prunier, R., Akman, M., Kremer, C. T., Aitken, N., Chuah, A., Borevitz, J., and 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.

Prunier, R. and 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.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., and Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3):559–575.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Racimo, F., Sankararaman, S., Nielsen, R., and Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, 16(6):359–371.

Rebelo, A., Siegfried, W., and Crowe, A. (1984). Avian pollinators and the pollination syndromes of selected mountain fynbos plants. *South African Journal of Botany*, 3(5):285–296.

Rebelo, A., Siegfried, W., and Oliver, E. (1985). Pollination syndromes of *Erica* species in the south-western Cape. *South African Journal of Botany*, 51(4):270–280.

Schrempf, D., Minh, B. Q., von Haeseler, A., and Kosiol, C. (2019). Polymorphism-Aware Species Trees with Advanced Mutation Models, Bootstrap, and Rate Heterogeneity. *Molecular Biology and Evolution*, 36(6):1294–1301.

Steinig, E. J., Neuditschko, M., Khatkar, M. S., Raadsma, H. W., and Zenger, K. R. (2016). NETVIEW P: a network visualization tool to unravel complex population structure using genome-wide SNPs. *Molecular Ecology Resources*, 16(1):216–227.

Stoffel, M. A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., and Hoffman, J. I. (2016). inbreedR: an R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution*, 7(11):1331–1339.

Tange, O. (2020). GNU Parallel 20200522 ('Kraftwerk'). Zenodo: https://doi.org/10.5281/zenodo.3841377.

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, 17(2):57–86.

Templeton, A. R. (1981). Mechanisms of Speciation – A Population Genetic Approach. *Annual Review of Ecology and Systematics*, 12:23–48.

Trinder-Smith, H., Cowling, R., and Linder, H. (1996). Profiling a besieged flora: endemic and threatened plants of the Cape Peninsula, South Africa. *Biodiversity & Conservation*, 5(5):575–589.

van der Niet, T. and Johnson, S. D. (2009). Patterns of plant speciation in the Cape floristic region. *Molecular Phylogenetics and Evolution*, 51(1):85–93.

van der Niet, T., Pires, K., and Steenhuisen, S. L. (2020). Flower constancy of the Cape honey bee pollinator of two co-flowering *Erica* species from the Cape Floristic Region (South Africa). *South African Journal of Botany*, 132:371–377.

van Der Niet, T., Pirie, M. D., Shuttleworth, A., Johnson, S. D., and Midgley, J. J. (2014). Do pollinator distributions underlie the evolution of pollination ecotypes in the Cape shrub *Erica plukenetii*? *Annals of Botany*, 113(2):301–315.

Verboom, G. A., Archibald, J. K., Bakker, F. T., Bellstedt, D. U., Conrad, F., Dreyer, L. L., Forest, F., Galley, C., Goldblatt, P., Henning, J. F., Mummenhoff, K., Linder, H. P., Muasya, A. M., Oberlander, K. C., Savolainen, V., Snijman, D. A., van der Niet, T., and Nowell, T. L. (2009). Origin and diversification of the Greater Cape flora: Ancient species repository, hot-bed of recent radiation, or both? *Molecular Phylogenetics and Evolution*, 51(1):44–53.

Verboom, G. A., Stock, W. D., and Cramer, M. D. (2017). Specialization to Extremely Low-Nutrient Soils Limits the Nutritional Adaptability of Plant Lineages. *The American Naturalist*, 189(6):684–699.

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, 106(supplement_1):9939–9946.

Weir, B. S. and Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6):1358–1370.

Westbury, M. V., Petersen, B., and Lorenzen, E. D. (2019). Genomic analyses reveal an absence of contemporary introgressive admixture between fin whales and blue whales, despite known hybrids. *Plos One*, 14(9):e0222004.

Yu, G., Smith, D., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8:28–36.

Supporting Information

Figure S1. Multilocus heterozygosity (MLH) of all individuals in the study.

Figure S2. Mean individual ancestry proportions estimated by sNMF.

Figure S3. Maximum-likelihood phylogeny inferred by IQ-TREE with all individuals included.

Figure S4. Read redundancy comparison between sequencing batches.

Table S1. Table summarising variant filtering procedures and features of the genotype matrices.

Table S2. Voucher information.