Botanical Journal of the Linnean Society, 2025, **XX**, 1–13
https://doi.org/10.1093/botlinnean/boaf015
Advance access publication 6 March 2025
Original Article

Original Article

Interploidy gene flow does not prevent adaptive genetic differentiation in sympatric populations of *Epidendrum fulgens* and *E. puniceoluteum* (Orchidaceae)

Bárbara Simões Santos Leal[1,2,*], Guilherme Bovi Ambrosano[3], Gabriel Rodrigues Alves Margarido[3], Clarisse Palma-Silva[1], Fabio Pinheiro[1]

[1]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP 13083-862, Brazil
[2]Instituto Tecnológico Vale, Belém, PA 66055-090, Brazil
[3]Departamento de Genética, Escola Superior de Agricultura 'Luiz de Queiroz' - Universidade de São Paulo, Piracicaba, SP 13418-900, Brazil

[*] Corresponding author. Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP 13083-862, Brazil.
E-mail: bssleal@gmail.com

ABSTRACT

Polyploids often exhibit ecological divergence from diploid parents, but the relative importance of selection in speciation by polyploidy remains to be tested in most systems. Here we use transcriptome-derived single nucleotide polymorphisms (SNPs) to test whether increased gene flow between the diploid *Epidendrum fulgens* and tetraploid *E. puniceoluteum* (Orchidaceae) in sympatry could prevent adaptation to contrasting habitats (sand dunes and swamps, respectively), and to infer genes probably under differential selection. Additionally, we used species distribution data to test for climatic niche divergence between species and a subset of synonymous SNPs to test for past demographic signatures. We found no evidence of introgression in the transcribed portion of their genomes. For the most differentiated loci between species, we annotated biological processes related to replication machineries and also to differential responses to habitat features. We also found that climatic niches slightly diverge due to increased tolerance to lower temperatures and wider amplitude of precipitation in *E. fulgens*, which probably explains the species' distinct signatures of past demographic changes. By combining ecological transcriptomics with climatic niche comparisons, we shed light on the potential role of adaptive processes in originating and maintaining plant biodiversity in Neotropical coastal environments.

Keywords: interploidy gene flow; hybridization; Orchidaceae; transcriptome-derived SNPs; speciation by polyploidy

INTRODUCTION

Hybridization is a pervasive evolutionary force influencing diversification of plants (Soltis and Soltis 2009, Mallet *et al.* 2016, Payseur and Rieseberg 2016). This process may lead to speciation by either introducing adaptive variation or by forming hybrid taxa through homoploid hybrid speciation or allopolyploidization (Abott *et al.* 2013, Seehausen *et al.* 2014). Although allopolyploidization is thought to confer strong and instantaneous reproductive isolation from diploid progenitors (postzygotic barriers; Levin 2002, Ramsey and Schemske 2002), studies have shown that gene flow across distinct ploidy levels sometimes persist (e.g. Pinheiro *et al.* 2010, Jørgensen *et al.* 2011, Zohren *et al.* 2016). Increasing evidence also supports that polyploids have relatively higher genetic diversity due to the prevalence of gene flow from conspecific diploids (Zohren *et al.* 2016) and also from other closely related polyploids (Masburger *et al.* 2019). Therefore, despite the existence of genomic instabilities, polyploidy may increase the chance of adaptive introgression, thereby representing an important source of diversity in nature (Marburger *et al.* 2019, Novikova *et al.* 2020).

Ecological niche differentiation is increasingly well supported as being broadly important for establishment and long-time persistence of polyploids (Duchoslav *et al.* 2017, Schmickl and Yant 2021). Indeed, polyploids often exhibit ecological divergence from diploid parents (Soltis and Soltis 2000, Ramsey and Ramsey 2014), suggesting an additional role of prezygotic barriers for reproductive isolation or at least for the initial establishment of the polyploid lineage and further coexistence with diploid parents and their derived polyploids (Levin 2002, Pegoraro *et al.* 2016). Furthermore, polyploids are particularly

Received 23 August 2024; revised 8 October 2024; accepted 24 January 2025

prone to occupy heterogeneous, stressful environmental conditions (Wei *et al.* 2019, Schmickl and Yant 2021). The timing and magnitude of interspecific gene flow are therefore important in the context of polyploid speciation, because it determines if gene flow can introduce adaptive mutations and therefore promote diversification (see Schmickl and Yant 2021) or, rather, be restricted to prevent adaptive differentiation across ploidy levels.

Ecologically divergent lineages forming hybrid zones have been examined to determine general mechanisms that reduce gene flow and promote diversification (Abbott *et al.* 2013, Gompert *et al.* 2017). For plants, sharp environmental changes may occur over short distances between hybridizing species due to differences in soil composition, temperature, moisture, and light regime. Although much can be learned from the available literature, the limited number of multiploidy hybridizing systems examined so far has prevented us from generalizing the process of adaptation and evolution of the relatively frequent plant polyploids. Indeed, most studies on patterns of interploidy gene flow have so far been concentrated on plants from temperate regions (e.g. Clark *et al.* 2015, Han *et al.* 2015, Zohren *et al.* 2016, Farhat *et al.* 2020), particularly from the model genus *Arabidopsis* (e.g. Arnold *et al.*, 2015; Monnahan *et al.* 2019). Polyploid model systems from tropical regions are scarce and studies rarely consider ploidy levels (but see Mergeay *et al.* 2008, Lopes *et al.* 2020, Diallo *et al.* 2022), despite the putative importance on driving speciation in these regions.

Genomic analyses provide excellent opportunities to test the effects of selection on distinct habitats in multiploidy hybridizing systems. Reduced representation approaches, for instance, have been widely used to infer the hybridization dynamics and identify genes evolved under divergent selection within hybrid zones (e.g. Chapman *et al.* 2013, Rheindt *et al.* 2014, Tavares *et al.* 2022). Single nucleotide polymorphism (SNP) genotyping from transcriptome sequencing, in particular, provides a cost-effective way to reduce the complexity of the genome while still retaining functionally relevant information, as the method directly targets genic regions (Wolf 2013, Alvarez *et al.* 2015, De Wit *et al.* 2015). It has also been claimed as a useful approach to study adaptation in species with no available reference genome (Andrews *et al.* 2016, Hoban *et al.* 2016). Indeed, when combined with gene annotation, such an approach can reveal loci with high levels of differentiation between species, and suggest candidate genes driving adaptation and reproductive isolation (Faria *et al.* 2014, Rajakaruna 2017). Furthermore, it allows for powerful testing of specific evolutionary histories (McCoy *et al.* 2014, Roux and Pannell 2015), given the high number of markers generally obtained. Nevertheless, investigating patterns of gene flow, drift, and selection across ploidy levels remains challenging for nonmodel species (Dufresne *et al.* 2014, Meirmans *et al.* 2018). Despite the many difficulties in modelling evolutionary forces in polyploids (such as determining allelic dosage or distinguishing paralogous sequences) and the lack of tools that allow for genetic analyses across different ploidy levels, recent effort has reshaped the field of polyploid population genetics and allowed new insights into their evolution (Meirmans *et al.* 2018, Schmickl and Yant 2021).

Here we use thousands of transcriptome-derived SNPs to assess the importance of adaptive processes in the origin and maintenance of multiploidy hybridizing systems, using two closely related neotropical orchids, the diploid *Epidendrum fulgens* Brongn and tetraploid *E. puniceoluteum* F.Pinheiro & F.Barros, as a model. Both species belong to *Epidendrum* subg. *Amphyglottium* and diverged ~1.9 Mya (Cardoso-Gustavson *et al.* 2018). A hybrid origin for *E. puniceoluteum*, with *E. fulgens* as one of its parents, has been previously hypothesized due to shared chromosome sequences (Pinheiro *et al.* 2010, Moraes *et al.* 2013). Studies have confirmed that these species hybridize in the Atlantic coastal plains in southeast Brazil, a narrow vegetation physiognomy known as *restingas*, and form triploid hybrids able to backcross with *E. puniceoluteum* (Pinheiro *et al.*, 2010; Moraes *et al.*, 2013). In sympatric zones, *E. fulgens* is predominantly found in sandy dunes, while *E. puniceoulutem* is found in swampy areas further inland (Sujii *et al.* 2019). A previous study showed that edaphic conditions differ between species within sympatric areas, and that candidate genes underlying salt and waterlogging adaptation are differently expressed (Leal *et al.* 2020). Although not explicitly tested, previous studies have also suggested that reinforcement is unlikely to affect divergence in these species, once hybrids share both habitats with the parental species (Sujii *et al.* 2019) and many genes of ecological importance (Leal *et al.* 2020).

By using RNA sequencing data collected from sympatric and allopatric populations of the diploid *E. fulgens* and tetraploid *E. puniceoluteum* (Leal *et al.*, 2020) and data on climatic niche divergence, we aim to (i) evaluate if previously detected introgression is less extensive in genic regions which are more subject to divergent selection, (ii) test whether increased gene flow in sympatry could prevent adaptation to contrasting habitats due to homogenization of part of the genes subject to selection, (iii) identify outlier variants potentially associated with selection to divergent habitats between species, and (iv) infer whether species are adapted to distinct climatic conditions and therefore show distinct signatures of past demographic changes using a subset of synonymous SNPs. We hypothesize that adaptive processes may have played an additional role in the origin and maintenance of the tetraploid *E. puniceluteum*, and that between-species gene flow does not prevent selection for distinct habitats. Our results shed light on the potential role of selection during speciation by polyploidy and help to disentangle the drivers of plant speciation in the Neotropics.

MATERIAL AND METHODS

Plant system

The diploid *E. fulgens* and tetraploid *E. puniceoluteum* are distributed in the *restingas* vegetation along the southeastern and southern Brazilian Atlantic coast, with a small area of geographical overlap located from southern São Paulo to Paraná state (Fig. 1). This ecosystem covers sandy soils of marine and fluvial–marine origin dated to the Quaternary (Scarano 2002). Although these species frequently hybridize forming triploid hybrids (Moraes *et al.* 2013), studies have documented ecological divergence between these species that probably results from the adaptation to different soil conditions and helps to maintain the reproductive isolation between them in the zone of overlap (Pinheiro *et al.* 2010, Sujii *et al.* 2019, Leal *et al.* 2020). Indeed, *E. fulgens* inhabits sand dunes not subject to flooding, while *E. puniceoluteum* is generally found in sedge swamps, i.e. depressions between successive beach ridges that are seasonally

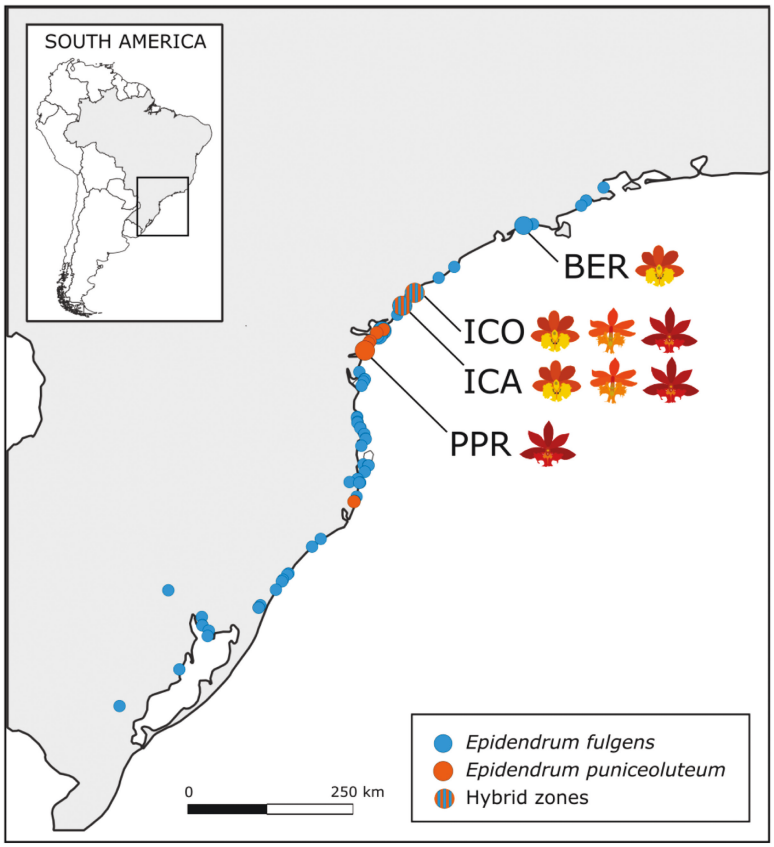

Figure 1. Distribution of occurrence records of the diploid *Epidendrum fulgens* and tetraploid *E. puniceoluteum* (Orchidaceae). Sampled sites are shown in the map as BER (Bertioga—SP) = *Epidendrum fulgens* in allopatry; ICO (Ilha Comprida—SP) and ICA (Ilha do Cardoso—SP) = *E. fulgens* and *E. puniceoluteum* in sympatry; and PPR (Pontal do Paraná—PR) = *E. puniceoluteum* in allopatry.

flooded (Pinheiro *et al.* 2010). Other abiotic and biotic characteristics of their habitats, such as microclimate and symbiotic interactions (i.e. mycorrhiza), although not characterized, may also show contrasting patterns in dunes and swamps.

De novo assembly and functional annotation

For this study, we used previously generated RNA sequencing data (available at the NCBI SRA database, BioProject ID: PRJNA678229; Leal *et al.* 2020) obtained from root tissues of 31 individuals of *E. fulgens*, *E. puniceoluteum*, and their hybrids (three to five individuals per species per location) distributed in two hybrid zones (sympatric areas), named ICO and ICA, and one allopatric locality of each species, named BER and PPR (Fig. 1; Supporting Information, Table S1). Geographical distance between localities varied from 41.39 km (ICA and ICO, Fig. 1) to 327.53 km (PPR and BER, Fig. 1). As *E. fulgens* and *E. puniceoluteum* occur in distinct microhabitats mostly delimited by soil variation (dunes versus swamps, respectively) but are similarly exposed to high solar radiation, we expected to capture more variants under divergent selection in root tissue than in leaves.

While a previous study focused on differences at gene expression level, in this study we looked at genetic variation between these species. To do so, we *de novo* assembled a 'mixed species transcriptome' to be used as a new reference to map the raw RNA sequencing reads. We first examined quality metrics from the downloaded Illumina paired-end raw reads using FastQC v.0.11.9 (Andrews 2019). After removing erroneous k-mers from reads using Rcorrector v.1.0.4 (Song and Florea 2015) with default settings, we trimmed adaptors and low-quality bases using TrimGalore v.0.6.0 (http://www.bioinformatics.babraham.ac.uk/projects). We removed sequences from mycorrhiza contaminants by aligning reads against the genome of *Serendipita indica*, *Serendipita vermifera*, and *Tulasnella calospora* (NCBI BioProjects: PRJEA76339, PRJNA207844, and PRJNA207843) using Bowtie2 v.2.3.5.1 (Langmead and Salzberg 2012), followed by SAMtools v.1.10 (Danecek *et al.* 2021) and BEDTools (Quinlan and Hall 2010). We also removed rRNAs listed in the SILVA database (Quast *et al.* 2012) using the bbduk.sh script from the BBMap package v.38.90 (https://jgi.doe.gov/data-and-tools/bbtools/).

Filtered reads from a subset of four samples of each species (from BER and PPR allopatric populations, see Fig. 1) were further used for *de novo* assembly of the reference transcriptome using Trinity v.2.8.5 (Grabherr *et al.* 2011) following the protocol of Haas *et al.* (2013) with the following parameters: '–normalize_reads –CuffFly –min_glue 4'. General assembly statistics were obtained with the TrinityStats.pl script made available in the Trinity package. The quality of the resulting assembly was assessed by aligning reads from each library back to the assembled transcriptome using Bowtie2. We also accessed the completeness of the transcriptome by aligning assemblies against the Embryophyta database and tested for benchmarking universal single-copy orthologue content using BUSCO v.5.0.0 (Simão *et al.* 2015).

To prevent redundancy in the *de novo* assembled transcriptome, we first clustered transcripts using CD-HIT-EST v.4.8.1

(Li and Godzik 2006) with a sequence similarity cut-off of 0.95. We then clustered isoforms assembled for each gene into SuperTranscripts (Davidson *et al.* 2017) using the Trinity toolkit script Trinity_gene_splice_modeler.py, which implements Lace software v.1.0. In our final transcriptome, each gene is therefore represented by a single sequence. Although these SuperTranscripts do not represent a true biological molecule, they provide a good substitute for a reference genome and allow for the use of splice-aware tools to perform the alignment step for SNP calling (Davidson *et al.* 2017).

We annotated the *de novo* assembled transcriptome using the Trinotate v.3.2.1 pipeline (Bryant *et al.* 2017). The first step consisted of using TransDecoder v.5.5.0 (https://github.com/TransDecoder) to predict protein-coding regions within transcripts. We then conducted a BLASTP (Camacho *et al.* 2009) search (with e ≤ 1e-3) using the predicted protein sequences against Swiss-Prot (Boeckmann *et al.* 2005) and UniRef90 (The UniProt Consortium 2014) protein databases. The Pfam databases (Finn *et al.* 2014) were also used to predict conserved protein domains using HMMER v.3.3.1 (Potter *et al.* 2018). We further used Signalp (Nielsen 2017) and TMHMM (Sonnhammer *et al.* 1998) tools to predict signal peptides (secretion signals) and transmembrane domains within these protein sequences. Similarities to known proteins were also detected by conducting BLASTX and BLASTP searches (Camacho *et al.* 2009) with the entire transcripts as queries against the same protein databases and using the same cutoff (i.e. e ≤ 1e-3).

Variant calling

The filtered reads from 31 individuals of *E. fulgens*, *E. puniceoluteum*, and their hybrids were mapped to the reference transcriptome using STAR v.2.7.9 (Dobin and Gingeras 2015), which has been shown to have superior SNP sensitivity in a comparative study evaluating mapping tools (Engström *et al.* 2013). Duplicated reads were further marked and read groups were added using 'Picard' (http://broadinstitute.github.io/picard). We used the SplitNCigarReads tool of the Genome Analysis Toolkit v.4.2.0.0 (GATK; DePristo *et al.* 2011, Van der Auwera *et al.* 2013) to split reads into exon segments (removing Ns but maintaining grouping information) and hard-clip any sequences overhanging into the intronic regions. The GATK module 'HaplotypeCaller' was used to call genotypes by specifying the ploidy of each sample with the -ploidy argument, i.e. '-ploidy 2' for *E. fulgens*, '-ploidy 4' for *E. puniceoluteum*, and '-ploidy 3' for the hybrids. We further used the 'GenomicsDBImport' and 'GenotypeGVCFs' GATK modules to call variants among samples using two distinct approaches: first, we called variants common to all sampled individuals (full dataset), then to the subset of 24 samples belonging to the parental species, i.e. *E. fulgens* and *E. puniceoluteum* (parentals dataset). To run the 'GenomicsDBImport' and 'GenotypeGVCFs' modules within a feasible time span, we concatenated the contigs into 50 artificial blocks/chromosomes with each contig apart from each other by 100 Ns and translated the joint variant locations back to the original using a custom python script. Following recommendations of the GATK workflow, we filtered out clusters of three or more SNPs within a window of 35 bp, and any variant with Fisher Strand values (FS) <30.0 and quality by depth values (QD) <2.0 using the 'VariantFiltration' module. We further used the R-script filter-vcf.R of Blischak *et al.* (2018a) to retain a single biallelic SNP per contig with minimum read depth of five and maximum of 10% missing individuals. The pipeline used for transcriptome assembly and SNPs calling is summarized in Supporting Information, Fig. S1.

Genetic diversity and structure

To infer the genetic structure from the full and parentals datasets, we used a Bayesian approach implemented in the software Structure v.2.3.4 (Pritchard *et al.* 2021, Falush *et al.* 2021) which is suitable for mixed ploidy datasets. The software Structure was run under the admixture model allowing for correlated allele frequencies, using a burn-in of 20 000 followed by 100 000 Markov chain Monte Carlo (MCMC) iterations. As Structure accepts only uniform ploidy as input, we added one or two rows of missing data for our triploid and diploid samples, respectively, making them pseudotetraploids. Ten replicate runs were performed with values of K ranging from 1 to 8. We inferred the most likely K as described by Evanno *et al.* (2005) using Structure Harvester v.0.6.94 (Earl and vonHoldt 2011) and extracted individual admixture proportions for the best K averaged across runs using the main pipeline of the CLUMPAK server (Kopelman *et al.* 2015). We also ran the admixture model from Entropy v.1.2 (Shastry *et al.* 2021) to obtain Bayesian estimates of genotypes for mixed ploidy data using the full dataset at $K = 2$. To to so, we used three replicated runs of 30 000 steps following 10 000 steps that were discarded as burn-in, sampled every 10th step. We then conducted a principal component analysis (PCA) using Singular Value Decomposition (SVD) on the genotype estimates matrix obtained from Entropy scaled by the ploidy (i.e. divided by 2, 3, or 4 in diploids, triploids, and tetraploids, respectively), as indicated in the software manual.

We calculated F_{ST} (Weir and Cockerham 2006) between *E. fulgens* and *E. puniceoluteum* within each sympatric zone and between species from allopatric populations for each of the 18 772 filtered contigs of the parentals dataset (excluding hybrids) using the R package 'StAMPP' v.1.6.3 (Pebleton *et al.* 2013). This package allows for differentiation estimates based on SNPs datasets from mixed ploidy levels. We further compared patterns of distribution of F_{ST} values across loci between sympatric and allopatric conditions to test whether differences between species would be accentuated in a few regions under strong selection in sympatric zones, while being more homogeneously distributed in allopatry.

Detection of outlier SNPs

As traditional methods for detection of outlier SNPs do not allow the entry of polyploid data (Meirmans *et al.* 2018), we identified highly differentiated outlier SNPs that may result from divergent selection by contrasting the F_{ST} of individual SNPs to the empirical genome-wide distribution of F_{ST}, as calculated by 'StAMPP' (Pembleton *et al.* 2013). Loci that showed F_{ST} values greater than the upper 98% quantile calculated upon the entire empirical distribution in all *E. fulgens* versus *E. puniceoluteum* comparisons were retained as outliers. We extracted Gene Ontology (GO) terms from the Trinotate output for the set of loci identified as outliers in all comparisons between species and

summarized them with REVIGO (Supek *et al.* 2011), using the simRel score and the Uniprot database.

Demographic patterns

To prevent bias in demographic analysis, we filtered out nonsynonymous SNPs (i.e. any variant that resulted in a different amino acid being incorporated into the protein) to generate a subset of synonymous SNPs from the parental species dataset. For this, we used the GFF3 file generated by TransDecoder during the annotation procedure, converted the GFF3 file to a GTF file using the R package 'rtracklayer' (Lawrence *et al.* 2009) and then used SNPdat 1.0.5 (Doran and Creevey 2013) to annotate variants from the parentals dataset as synonymous or nonsynonymous.

The folded site frequency spectrum (SFS) was built using the SNPs that SNPdat unambiguously called as synonymous for each population from *E. fulgens* and *E. puniceoluteum* using the down-projection method implemented in easySFS (https://github.com/isaacovercast/easySFS). The software Stairway Plot v.2 (Liu and Fu 2020) was then used to infer effective population size (N_E) changes through time based on the inferred SFS. For this, we applied the approximate mutation rate of 10^{-9} for plants (see Lynch 2010) and considered the generation time of both species as 5 years, in accordance with previous studies of *Epidendrum* subg. *Amphyglottium* (Devadas *et al.* 2010; Pessoa *et al.* 2022).

Climatic niche differentiation

We used a spatially explicit analysis to estimate the climatic niche overlap between hybridizing *E. fulgens* and *E. puniceoluteum*. For this, we obtained occurrence records of *E. fulgens* and *E. puniceoluteum* from previous field expeditions and from the Global Biodiversity Information Facility (GBIF 2019a, 2019b) and speciesLink platform (CRIA—Centro de Referência e Informação Ambiental). We further filtered 65 and nine spatially unique records with precise coordinates and reliable identification for each species, respectively (Fig. 1).

We calculated uncorrected and corrected environmental niche overlap based on the climatic data from the filtered occurrence records of each species using a kernel smoother method implemented in the R package 'ecospat' v.3.5.1 (Di Cola *et al.* 2017). Climatic variables were downloaded from the Paleoclim database (Brown *et al.* 2018) for the current period with 30″ resolution (~1 km^2) and then cut for the area spanning from −15 to −36° of latitude and −38 to −58° of longitude. To avoid multicollinearity, we retained seven low-correlation (<.7) variables as detected by the vifcor function using the 'usdm' R package (Naimi 2023): maximum temperature of warmest month (Bio 5), minimum temperature of coldest month (Bio 6), mean temperature of warmest quarter (Bio 10), mean temperature of coldest quarter (Bio 11), annual precipitation (Bio 12), precipitation of wettest month (Bio 13), and precipitation of warmest quarter (Bio 18).

Using the 'ecospat' package, we applied a background based on randomly sampled points within a buffer radius of 90 km using the ENMTools R package (Warren *et al.* 2021). A PCA was then used to reduce the set of seven bioclimatic variables to two principal component axes defining the gridded environmental space for each species. We further tested if niches of *E. fulgens* and *E. puniceoluteum* are more divergent ($P < .05$) than expected by chance using the niche equivalency and the niche similarity functions based on these same occurrence grids and backgrounds. The niche equivalency test is considered more conservative as it tests whether the species' niches diverge by only considering occurrences (Warren *et al.* 2008). The niche similarity test, in turn, also accounts for the differences in the surrounding environmental conditions (background) to test for niche divergence (Broennimann *et al.* 2012). Simulations for both niche similarity and equivalency tests were performed in 10 000 replicates each.

RESULTS

Transcriptome assembly and SNP datasets

We assembled 445 707 332 trimmed reads into a combined *E. fulgens* and *E. puniceolutem* transcriptome assembly containing a total of 500 116 contigs. This new assembled reference transcriptome had a contig N50 of 899 bp and an average contig length of 617 bp. BUSCO indicates a well-assembled transcriptome with 94.9% complete genes (92.3% single-copy plus 2.6% duplicated), 2.5% fragmented genes, and 2.6% missing genes. Such a high percentage of complete BUSCOs indicates that our reference transcriptome contains most of the highly conserved, single-copy orthologous genes expected to be present in all Embryophytes.

The average read mapping across libraries was 97.35%, ranging from 96.85 to 98% for *E. fulgens* individuals and from 96.2 to 97.76% for *E. puniceoluteum* individuals. After calling and filtering SNPs among samples, we retained 686 836 SNPs distributed in 21 031 contigs (full dataset) and 244 512 distributed in 18 772 contigs (parentals dataset). The additional filtering step used to retain only variants unambiguously called as synonymous for demographic analyses yielded a total of 115 221 synonymous variants from 14 514 contigs (parentals dataset subset synonymous).

Genetic structure patterns

We retrieved $K = 2$ as the best K for the Structure analysis using the parentals dataset. Structure showed no pattern of admixture in the codifying part of the genome of the parental species *E. fulgens* and *E. puniceoluteum* (Fig. 2A). When adding the hybrids to such analysis (full dataset), we also retrieved $K = 2$ as the best K, with a clear separation between species and admixed hybrids, except for a single individual from ICA (Fig. 2B). Results from the PCA (first two components) executed from genotype estimates of parental species and hybrids mirrored those from Structure, showing all but one putative hybrid as intermediate individuals (Fig. 2C).

Transcriptome-wide differentiation between the two species was high in both allopatric and sympatric conditions (mean $F_{ST} = 0.382 \pm 0.154$ and 0.378 ± 0.153, respectively; Fig. 3A). We also found similar patterns of distribution of F_{ST} values between species in allopatry and sympatry, as both F_{ST} density curves have similar means and variances (Fig. 3A).

Detection of outlier SNPs

We recovered a total of 428 outlier SNPs from the parentals dataset using the 98% F_{ST} cut-off approach for each *E. fulgens*

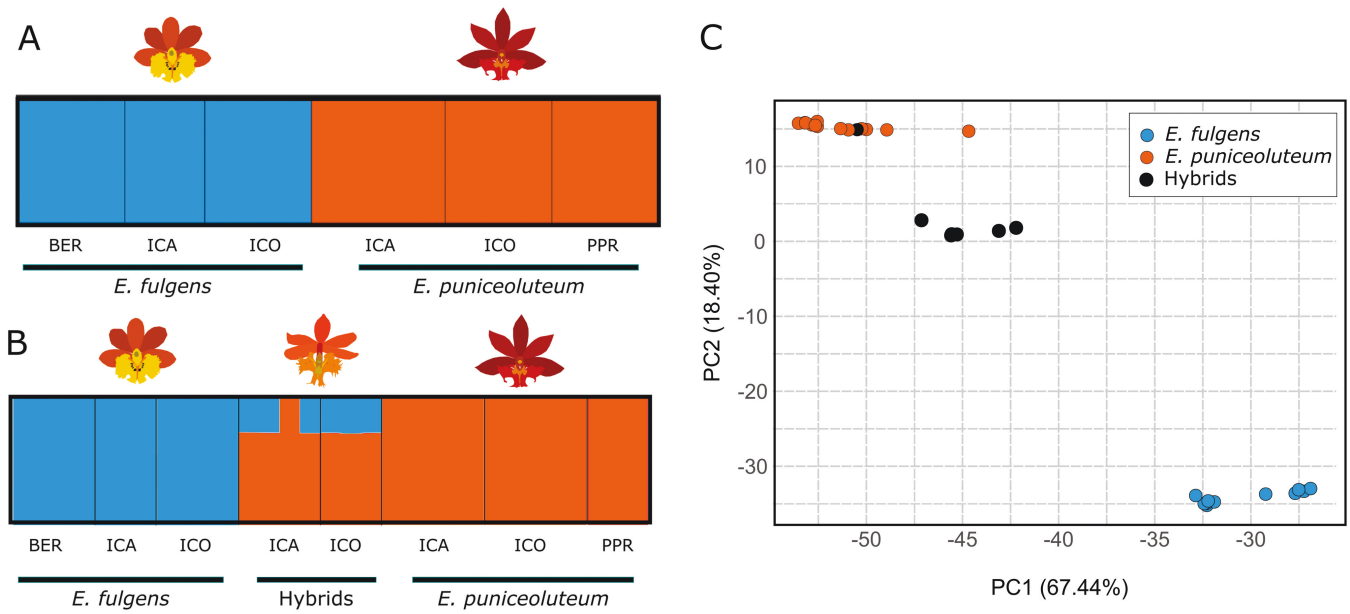

Figure 2. Genetic structure between diploid *Epidendrum fulgens* and tetraploid *E. puniceoluteum* (A) and among these parental species and their hybrids (B and C) using 18 772 (parentals dataset) or 21 031 (full dataset) transcriptome-derived SNPs, respectively, as indicated by the Bayesian analysis executed in Structure (A and B) and the first two axes of a principal component analysis (PCA, C). *Epidendrum fulgens* is shown in blue, while *E. puniceoluteum* is shown in orange. Hybrids are shown as the admixed individuals in Structure plots (B) and black points in the PCA plot (C).

versus *E. puniceoluteum* comparison. We focused our interpretation on a total 183 loci that showed consistent signatures of high genetic divergence ($F_{ST} \geq 0.72$) between parental species across the three comparisons (see Veen diagram; Fig. 3B). From these, we assigned GO terms for 84 genes and obtained a final list of 63 nonredundant terms (dispensability < 0.5). Some outlier SNPs are associated with nucleus activities such as chromatin organization (GO:0006325; $F_{ST} = 0.960$) and karyogamy (GO:0000741; $F_{ST} = 0.899$) (Supporting Information, Table S2), but also responses to environmental conditions, such as responses to abscisic acid (GO:0009737, $F_{ST} = 0.969$), wounding (GO:0009611; $F_{ST} = 0.969$), fungus (GO:0050832, $F_{ST} = 0.898$), and heat (GO:0009408, $F_{ST} = 0.859$) (Fig. 3C; Table S2). Notably, the two fixed loci (i.e. $F_{ST} = 1$) between species are responsible for vesicle-mediated transport and RNA modification (GO:0016192 and GO:0009451) (Fig. 3C; Table S2).

Demographic patterns

Stairwayplot 2 analysis evidenced contrasting signatures of changes in N_E though time in *E. fulgens* and *E. puniceoluteum* (Fig. 4). Such analysis showed that populations of *E. fulgens* had a slight increase (two-fold) of N_E from ~70 000 years ago (Fig. 4A-B). However, populations of *E. puniceoluteum* have probably passed by drastic reductions in N_E over time and today correspond to ~1/1000th of the ancestral population size (Fig. 4A, B).

Climatic niche divergence

The divergence tests using climatic data derived from current records of species revealed that the climatic niches of *E. fulgens* and *E. puniceoluteum* are not equivalent (Schoener's $D = 0.171$, $P = .0089$; Warren's $I = 0.408$, $P = .012$), despite being more similar than expected by chance ($P > .05$, Fig. 5B). Within the climatic space, differences between species' niche are more closely correlated with the minimum temperature of coldest month (Bio 6) and precipitation of wettest month (Bio 13, Fig. 5A). The realized climatic niche of *E. fulgens* is ~12 times larger than that of *E. puniceoluteum* (Fig. 5B).

DISCUSSION

Inferring the causes of speciation and the consequences of hybridization is one of the central goals of evolutionary biology (Seehausen *et al.* 2014), though it presents particular challenges when examining polyploid systems (Meirmans *et al.* 2018). In our study, we refined previous knowledge on the speciation and hybridization between *E. fulgens* and *E. puniceoluteum*, two neotropical orchids exhibiting varying ploidy (diploid and tetraploid, respectively) and occurring in contrasting microhabitats in *restingas* (sand dunes and sedge swamps, respectively). By combining a population genetics approach using thousands of transcriptome-derived SNPs and climatic niche analysis, we show that gene flow might not be sufficient to prevent adaptive genetic differentiation between sympatric populations of *E. fulgens* and *E. puniceoluteum*, and that these species differ in several genes potentially linked to habitat divergence and maintain distinct climatic niches despite their large geographical distribution overlap. Below, we discuss the significance of our findings in improving our understanding of the general implications of interploidy gene flow in plants and providing insights into the role of adaptive genetic variation during speciation by polyploidization.

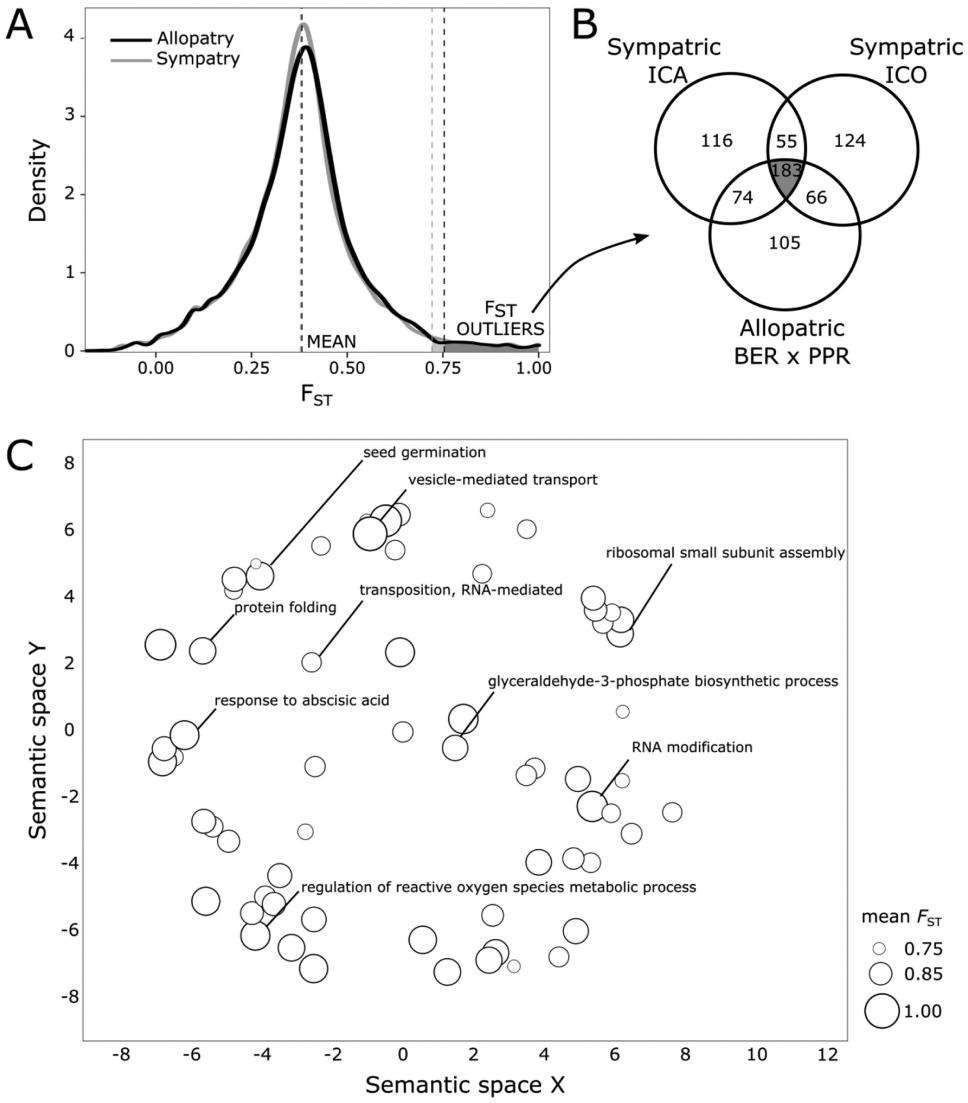

Figure 3. Patterns of genomic differentiation between diploid *Epidendrum fulgens* and tetraploid *E. puniceoluteum* across 18 772 transcriptome-derived SNPs in allopatric versus sympatric conditions. A, loci with F_{ST} values greater than the upper 98% quantile were retained as outliers. B, Venn diagram highlighting the subset of outliers commonly retrieved in comparisons between species in sympatry (ICA and ICO) and allopatry (ALO). C, semantic plot obtained in REVIGO summarizing the Gene Ontology (GO) terms for the subset of 84 out of 183 annotated loci. The more general GO terms associated with higher mean F_{ST} values (circle size) are shown in the plot.

Our results show a strong genetic divergence between hybridizing orchids *E. fulgens* and *E. puniceolutem*, with no evidence of introgression in the transcribed portion of their genomes. It contrasts with previous evidence from neutral microsatellite markers and experimental crosses showing that the tetraploid *E. puniceoluteum* experienced gene flow from diploid *E. fulgens*, but not the reverse (Pinheiro *et al.* 2010, Moraes *et al.* 2013), which resulted in clear signatures of introgression (Pinheiro *et al.* 2010). This discrepancy could be attributed, in part, to the notable variance in mutation rates between SNPs and microsatellites. However, we propose that the varying scenarios of reproductive isolation observed in these hybridizing species, contingent upon the molecular markers utilized (i.e. microsatellites often found in nongenic regions versus SNPs derived from genic regions), can be explained primarily by the disproportionate effects of divergent selection in genic versus nongenic regions. So far, the accumulated evidence suggests that reproductive isolation between *E. fulgens* and *E. puniceoluteum* is probably of adaptive nature since introgression is more extensive at nongenic regions (Pinheiro *et al.* 2010). Despite growing evidence that introgression represents a source of novel variation that transfers traits of adaptive value in multiploidy lineages (e.g. Monnahan *et al.* 2019, Novikova *et al.* 2020), the implications of hybridization probably depend on the involved species (Sutherland and Galloway 2017). Introgression between *E. fulgens* and *E. puniceoluteum* might be limited by selection for distinct habitats, particularly if F2 hybrids and introgressed individuals experience a fitness decline compared to parental species, which demands further investigation.

Our data indicate that interspecific gene flow might not prevent adaptation to contrasting habitats in sympatric locations. We found similar patterns of distribution of F_{ST} values between

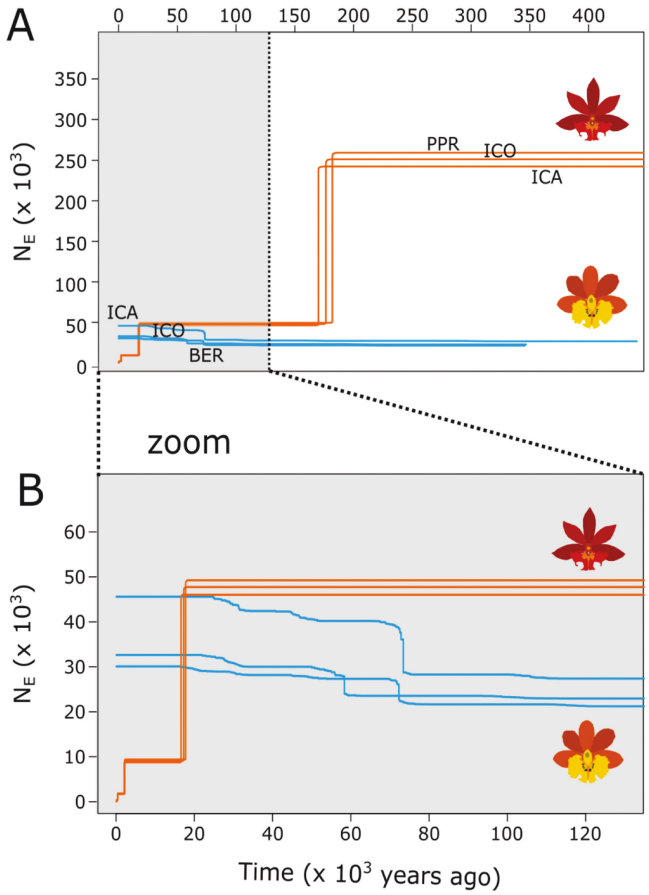

Figure 4. Changes in effective population sizes (N_E) through time for diploid *Epidendrum fulgens* (in blue) and tetraploid *E. puniceoluteum* (in orange) during the last 400 000 years (A) and 120 000 years (B) as estimated from the site frequency spectrum (SFS) based on the subset of 14 514 synonymous SNPs using StairwayPlot 2.

species across the genome in sympatry and allopatry. We also found that a significant proportion of the more highly differentiated loci (F_{ST} > 0.72) are commonly identified as outliers in comparisons between these species in allopatric or sympatric conditions. Some of these loci are responsible, for instance, for the transcriptional and replication machineries (i.e. GO:0009451: RNA modification, GO:0006325: chromatin organization, GO:0006357: regulation of transcription by RNA polymerase II, GO:0000741: karyogamy, GO:0031120: snRNA pseudouridine synthesis) that might be important for the physical chromosome reorganization following polyploidization (see Zhang *et al.* 2019, Concia *et al.* 2020, Garcia-Lozano *et al.* 2021). Recent evidence has indeed indicated that multiple genes are involved in the stabilization of replication following whole genome duplication in *Arabidopsis* (Yant *et al.*, 2013; Bohutínská *et al.*, 2021), but the identification of targets of divergent selection that result in reproductive isolation across ploidy levels remains sparse to characterize the generality of the process in plants. Furthermore, we cannot disregard alternative explanations for the presence of F_{ST} outliers, such as the association of these loci with islands of reduced linkage and low diversity within species (Cruickshank and Hahn 2014). A more comprehensive understanding of the genetic basis of speciation by polyploidy in this system includes contrasting neutral versus selection scenarios (see Clo *et al.* 2022) using data-informed simulations. Testing whether divergent selection has acted on the same suite of genes in other pairs of species in *Epidendrum* or, more probably, in distinct genes affecting the same core pathways would also bring robust evidence in favour of selection (see Bohutínská *et al.* 2021).

Our analysis also revealed some enriched processes that might be linked to responses to several habitat features (e.g. GO:0009737, GO:0009611 and GO:0009408: responses to abscisic acid, wounding, and heat, respectively, and GO:0050832: defence response to fungus) among the most differentiated loci between *E. fulgens* and *E. puniceoluteum*. Niche differentiation plays an important role in species' divergence due to its impact on both pre- and postmating isolation (Sobel *et al.* 2019) and is increasingly supported as important for polyploid establishment (Ramsey 2011, Marchant *et al.* 2016) mitigating the so-called minority cytotype exclusion (Levin 1975). A recent study has indeed shown that habitat preferences is an important component for reproductive isolation between two closely related *Epidendrum* species with distinct ploidy (Arida *et al.* 2021). Nevertheless, it is unclear whether habitat differences between the studied species have been achieved immediately following the raising of the tetraploid *E. puniceoluteum* or have been shaped for long periods of time after polyploidization. By phasing the subgenomes of the putative allopolyploid *E. puniceoluteum*, it would be possible to estimate divergence time and infer timing of gene flow between such species (see Kryvokhyzha *et al.* 2019), allowing further insights into the dynamics of speciation between these hybridizing taxa.

In agreement with divergent selection potentially affecting genes linked to the habitat features we have observed, we demonstrated in this study that the realized climatic niches of *E. fulgens* and *E. puniceoluteum* diverge slightly in environmental space due to increased tolerance to lower temperature and wider amplitude of precipitation in *E. fulgens*. In addition to previous evidence linking divergence to soil features (see Leal *et al.* 2020), our study has provided useful insights into ecological divergence driven by climate conditions between these closely related diploid and allotetraploid species, which is consistent with evidence from other plant lineages (e.g. Marchant *et al.* 2016, Baniaga *et al.* 2020). Nevertheless, our data do not support previous evidence showing expanded niches in tetraploids compared to their diploid progenitors (e.g. López-Jurado *et al.* 2019, Molina-Henao and Hopkins 2019, Kiedrzyński *et al.* 2021), since *E. puniceoluteum* is geographically restricted and its niche is climatically limited when compared to *E. fulgens*. When accounting for microhabitat features, however, *E. puniceoluteum* shows wider tolerance to flooding, which is responsible for determining some ecological differentiation between species (Pinheiro *et al.* 2010, Leal *et al.* 2020). The literature indeed shows that polyploidization usually increases stress tolerance in both plants and animals (e.g.e.g. Chao *et al.* 2013, Arnold *et al.* 2016, Novikova *et al.* 2020), but the extent to which genome duplication itself and the combination of two divergent genomes via hybridization facilitates the evolution of ecological novelty in allopolyploids needs further investigation (see Fort *et al.* 2016).

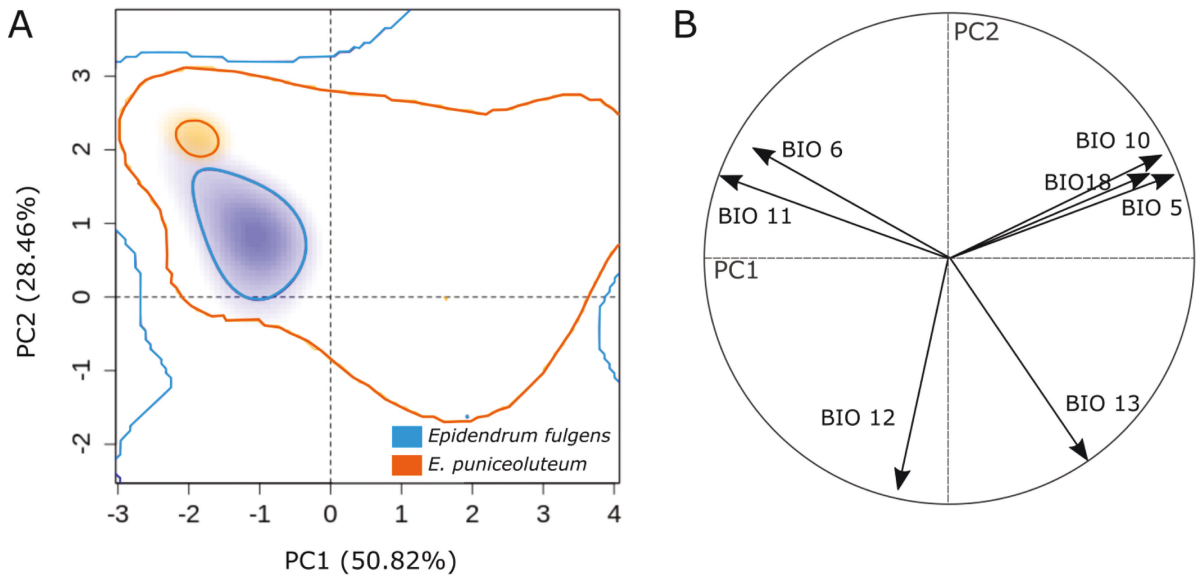

Figure 5. Current climatic niches of diploid *E. fulgens* (in blue) and tetraploid *E. puniceoluteum* (in orange) in the environmental space spanned by two PCA axes. A, overlap between background environments and realized climatic niches of *E. fulgens* (blue) and *E. puniceoluteum* (orange). Shading represents the density of the occurrences by cell, with darker shading indicating higher density of occurrences. Solid contour lines visualize 100% of the available environment. B, correlation circle showing the contribution of the seven bioclimatic variables—maximum temperature of warmest month (Bio 5), minimum temperature of coldest month (Bio 6), mean temperature of warmest quarter (Bio 10), mean temperature of coldest quarter (Bio 11), annual precipitation (Bio 12), precipitation of wettest month (Bio 13), and precipitation of warmest quarter (Bio 18), on the two first PCA axes (PC1 and PC2).

Future work that builds on the spatial and temporal distribution of diploids and allopolyploids in the subgenus to which *E. fulgens* and *E. puniceoluteum* belong (i.e. subg. *Amphyglottium*, Pinheiro *et al.*, 2009) will probably help to uncover the importance of polyploidization for climate adaptation (see Blischak *et al.* 2018b) and occupancy of harsh and disturbed habitats by these species (Pinheiro and Cozzolino 2013).

Finally, our data indicate distinct signatures of demographic changes across time between species, with *E. puniceoluteum* showing population declines following the last glacial maximum (LGM, 21 kya) and *E. fulgens* exhibiting a slight increase in population sizes from ~70 kya followed by stability since the LGM. This analysis also recovered an older event (170–180 kya) of population size decline in *E. puniceoluteum*, but the interpretation requires caution since SFS-based analysis, such as the one used in our study, performs more poorly when estimating ancient than recent population size history (Liu and Fu 2015, Baharian and Gravel 2018, Patton *et al.* 2019). In addition, demographic inference from genic regions even using a subset of synonymous variants can be skewed by the effects of selection at linked loci (Ewing and Jensen 2016, Schrider *et al.* 2016). Taking into account these putative caveats, we assume that the distinct recent demographic history of *E. fulgens* and *E. puniceoluteum* might result from adaptation to distinct flooding regimes, which has been shown to be the leading factor in how plant species are distributed over the southern Brazilian Atlantic coast (Spier *et al.* 2016). According to our results, *E. puniceoluteum* might have been more affected by sea level increase since the LGM than *E. fulgens*, probably due to the loss of waterlogged sandy areas despite the wetter climatic conditions during the Holocene (last 6 kya) compared to the LGM (Behling and Negrelle 2001, Garcia *et al.* 2004, Pessenda *et al.* 2004). Besides being more widely distributed along the Brazilian Atlantic coast, *E. fulgens* also occurs inland on granitic and arenitic rocky outcrops at its extreme southern distribution. The variability of habitat occupied by this species can indeed underlie the higher resilience to past climate and sea level changes. Although *E. fulgens* and *E. puniceoluteum* become established in sites that are ecologically dissimilar (nonflooded and periodically flooded, respectively), their distinct demographic histories and current coexistence in parapatry/sympatry may also have resulted from displacement of the tetraploid due to expansion of the diploid species (see Buggs and Pannel 2007).

CONCLUSION

Taken together, our results suggest that hybridization may have a limited impact on the genetic integrity of parental species when restricted to nongenic regions. As one of the first studies to examine the genetic background of hybridizing species with different ploidy levels in the Neotropics, our results contribute to further elucidating the role of adaptation to different environments as a driver of variation in heterogeneous habitats, even in the presence of interspecific gene exchange. Here, we identify candidate loci possibly associated with distinct ecological preferences of parental species, but controlled experiments are still needed to confirm their adaptive role due to the possibility of alternative explanations for the observed high divergence. Our study also shows that the parental species has different demographic trends, which are in agreement with the levels of abundance and niche amplitude occupied by each taxon. For example, the population increase of *E. fulgens* may be coupled to its ability to thrive in different extreme environments, such as on rock outcrops and sand dunes. In contrast, *E. puniceoluteum*

is found only in patchy populations that are seasonally flooded, environments that have probably become rarer since the end of the LGM. To date, considering the scenario of demographic decline, habitat adaptation may have ensured the genomic integrity of *E. puniceoluteum*, which has a patchy distribution and populations with fewer individuals. Future studies should take advantage of the genomic data already available for these nonmodel species to understand which functional traits are associated with divergent selection in each specific habitat. The use of reciprocal transplant experiments coupled with common garden studies may also reveal the intensity of local adaptation and the levels of plasticity contributing to species coexistence. Furthermore, biotic interactions, such as mycorrhizal associations and pollinator-mediated selection, provide essential data in studies seeking to understand patterns of adaptation at small geographical scales. Considering all of these contributing factors jointly will give us a more comprehensive understanding of the genetic architecture of local adaptation in hyperdiverse regions, such as the Neotropics.

SUPPLEMENTARY DATA

Supplementary data are available at *Botanical Journal of the Linnean Society* online.

ACKNOWLEDGEMENTS

We are grateful to the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP) for providing computational and storage resources.

CONFLICT OF INTEREST

We have no conflicts of interest to disclose.

FUNDING

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; 2016/22785-8 and CBioClima 2021/10639-5). We also thank FAPESP for the postdoc fellowship (18/18967-9) granted to B.S.S.L. and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the fellowship productivity in research granted to F.P. (302849/2021-1) and C.P.S. (305398/2019-9).

DATA AVAILABILITY

The data underlying this article (i.e. assembled reference transcriptome and annotation file, as well as the final SNP datasets) are available in the FigShare repository (doi: 10.6084/m9.figshare.27183336).

REFERENCES

Abbott R, Albach D, Ansell S *et al.* Hybridization and speciation. *Journal of Evolutionary Biology* 2013;**26**:229–46. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Alvarez M, Schrey AW, Richards CL. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Molecular Ecology* 2015;**24**:710–25. https://doi.org/10.1111/mec.13055

Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (1 March 2019, date last accessed).

Andrews KR, Good JM, Miller MR *et al.* Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 2016;**17**:81–92. https://doi.org/10.1038/nrg.2015.28

Arida BL, Scopece G, Machado RM *et al.* Reproductive barriers and fertility of two neotropical orchid species and their natural hybrid. *Evolutionary Ecology* 2021;**35**:41–64. https://doi.org/10.1007/s10682-020-10095-5

Arnold B, Kim ST, Bomblies K. Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. *Molecular Biology and Evolution* 2015;**32**:1382–95. https://doi.org/10.1093/molbev/msv089

Arnold BJ, Lahner B, DaCosta JM *et al.* Borrowed alleles and convergence in serpentine adaptation. *Proceedings of the National Academy of Sciences of the United States of America* 2016;**113**:8320–5. https://doi.org/10.1073/pnas.1600405113

Baharian S, Gravel S. On the decidability of population size histories from finite allele frequency spectra. *Theoretical Population Biology* 2018;**120**:42–51. https://doi.org/10.1016/j.tpb.2017.12.008

Baniaga AE, Marx HE, Arrigo N *et al.* Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecology Letters* 2020;**23**:68–78. https://doi.org/10.1111/ele.13402

Behling H, Negrelle RR. Tropical rain forest and climate dynamics of the Atlantic lowland, Southern Brazil, during the Late Quaternary. *Quaternary Research* 2001;**56**:383–9.

Blischak PD, Kubatko LS, Wolfe AD. SNP genotyping and parameter estimation in polyploids using low-coverage sequencing data. *Bioinformatics* 2018a;**34**:407–15. https://doi.org/10.1093/bioinformatics/btx587

Blischak PD, Mabry ME, Conant GC *et al.* Integrating networks, phylogenomics, and population genomics for the study of polyploidy. *Annual Review of Ecology, Evolution, and Systematics* 2018b;**49**:253–78.

Boeckmann B, Blatter MC, Famiglietti L *et al.* Protein variety and functional diversity: Swiss-Prot annotation in its biological context. *Comptes Rendus Biologies* 2005;**328**:882–99. https://doi.org/10.1016/j.crvi.2005.06.001

Bohutínská M, Vlček J, Yair S *et al.* Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proceedings of the National Academy of Sciences of the United States of America* 2021;**118**:e2022713118. https://doi.org/10.1073/pnas.2022713118

Broennimann O, Fitzpatrick MC, Pearman PB *et al.* Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* 2012;**21**:481–97. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Brown JL, Hill DJ, Dolan AM *et al.* PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* 2018;**5**:1–9.

Bryant DM, Johnson K, DiTommaso T *et al.* A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Reports* 2017;**18**:762–76. https://doi.org/10.1016/j.celrep.2016.12.063

Buggs RJ, Pannell JR. Ecological differentiation and diploid superiority across a moving ploidy contact zone. *Evolution* 2007;**61**:125–40. https://doi.org/10.1111/j.1558-5646.2007.00010.x

Camacho C, Coulouris G, Avagyan V *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* 2009;**10**:1–9.

Cardoso-Gustavson P, Saka MN, Pessoa EM *et al.* Unidirectional transitions in nectar gain and loss suggest food deception is a stable evolutionary strategy in *Epidendrum* (Orchidaceae): insights from anatomical and molecular evidence. *BMC Plant Biology* 2018;**18**:1–13.

Chao DY, Dilkes B, Luo H *et al.* Polyploids exhibit higher potassium uptake and salinity tolerance in Arabidopsis. *Science* 2013;**341**:658–9. https://doi.org/10.1126/science.1240561

Chapman MA, Hiscock SJ, Filatov DA. Genomic divergence during speciation driven by adaptation to altitude. *Molecular Biology and Evolution* 2013;**30**:2553–67. https://doi.org/10.1093/molbev/mst168

Clark LV, Stewart JR, Nishiwaki A *et al.* Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression. *Journal of Experimental Botany* 2015;**66**:4213–25. https://doi.org/10.1093/jxb/eru511

Clo J, Padilla-García N, Kolář F. Polyploidization as an opportunistic mutation: The role of unreduced gametes formation and genetic drift in polyploid establishment. *Journal of Evolutionary Biology* 2022;**35**:1099–109. https://doi.org/10.1111/jeb.14055

Concia L, Veluchamy A, Ramirez-Prado JS *et al*. Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biology* 2020;**21**:1–20.

Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 2014;**23**:3133–57. https://doi.org/10.1111/mec.12796

Danecek P, Bonfield JK, Liddle J *et al*. Twelve years of SAMtools and BCFtools. *GigaScience* 2021;**10**:giab008. https://doi.org/10.1093/gigascience/giab008

Davidson NM, Hawkins AD, Oshlack A. SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. *Genome Biology* 2017;**18**:1–10.

DePristo MA, Banks E, Poplin R *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 2011;**43**:491–8. https://doi.org/10.1038/ng.806

Devadas R, Medhi RP, Das SP. Interspecific hybrid developed in *Epidendrum* orchid from the cross *E. radicans* Pav. ex. Lindl. x *E. xanthinum* Lindl. *Journal of Horticultural Sciences* 2010;**5**:144–7. https://doi.org/10.24154/jhs.v5i2.465

De Wit P, Pespeni MH, Palumbi SR. SNP genotyping and population genomics from expressed sequences–current advances and future possibilities. *Molecular Ecology* 2015;**24**:2310–23. https://doi.org/10.1111/mec.13165

Diallo AM, Kjær ED, Ræbild A *et al*. Coexistence of diploid and polyploid *Acacia senegal* (L. Willd.) and its implications for interploidy pollination. *New Forests* 2022;**54**:67–82. https://doi.org/10.1007/s11056-021-09901-x

Di Cola V, Broennimann O, Petitpierre B *et al*. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* 2017;**40**:774–87. https://doi.org/10.1111/ecog.02671

Dobin A, Gingeras TR. Mapping RNA-seq reads with STAR. *Current Protocols in Bioinformatics* 2015;**51**:11–4.

Doran AG, Creevey CJ. Snpdat: easy and rapid annotation of results from de novo snp discovery projects for model and non-model organisms. *BMC Bioinformatics* 2013;**14**:1–6.

Duchoslav M, Fialová M, Jandová M. The ecological performance of tetra-, penta-and hexaploid geophyte *Allium oleraceum* in reciprocal transplant experiment may explain the occurrence of multiple-cytotype populations. *Journal of Plant Ecology* 2017;**10**:569–80.

Dufresne F, Stift M, Vergilino R *et al*. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology* 2014;**23**:40–69. https://doi.org/10.1111/mec.12581

Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 2011;**4**:359–61. https://doi.org/10.1007/s12686-011-9548-7

Engström PG, Steijger T, Sipos B *et al*.; RGASP Consortium. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods* 2013;**10**:1185–91. https://doi.org/10.1038/nmeth.2722

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 2005;**14**:2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewing GB, Jensen JD. The consequences of not accounting for background selection in demographic inference. *Molecular Ecology* 2016;**25**:135–41. https://doi.org/10.1111/mec.13390

Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2021;**164**:1567–87. https://doi.org/10.1093/genetics/164.4.1567

Farhat P, Siljak-Yakovlev S, Valentin N *et al*. Gene flow between diploid and tetraploid junipers-two contrasting evolutionary pathways in two *Juniperus* populations. *BMC Evolutionary Biology* 2020;**20**:1–16.

Faria R, Renaut S, Galindo J *et al*. Advances in ecological speciation: an integrative approach. *Molecular Ecology* 2014;**23**:513–21. https://doi.org/10.1111/mec.12616

Finn RD, Bateman A, Clements J *et al*. Pfam: the protein families database. *Nucleic Acids Research* 2014;**42**:222–30.

Fort A, Ryder P, McKeown PC *et al*. Disaggregating polyploidy, parental genome dosage and hybridity contributions to heterosis in *Arabidopsis thaliana*. *The New Phytologist* 2016;**209**:590–9. https://doi.org/10.1111/nph.13650

Garcia MJ, De Oliveira PE, Siqueira E *et al*. A Holocene vegetational and climatic record from the Atlantic rainforest belt of coastal State of São Paulo, SE Brazil. *Review of Palaeobotany and Palynology* 2004;**131**:181–99.

Garcia-Lozano M, Natarajan P, Levi A *et al*. Altered chromatin conformation and transcriptional regulation in watermelon following genome doubling. *The Plant Journal* 2021;**106**:588–600. https://doi.org/10.1111/tpj.15256

GBIF. Occurrence download https://doi.org/10.15468/dl.dwlijc (25 March 2019a, date last accessed).

GBIF. Occurrence download https://doi.org/10.15468/dl.gjrl5u (25 March 2019b, date last accessed).

Gompert Z, Mandeville EG, Buerkle CA. Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics* 2017;**48**:207–29. https://doi.org/10.1146/annurev-ecolsys-110316-022652

Grabherr MG, Haas BJ, Yassour M *et al*. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* 2011;**29**:644–52. https://doi.org/10.1038/nbt.1883

Haas BJ, Papanicolaou A, Yassour M *et al*. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 2013;**8**:1494–512. https://doi.org/10.1038/nprot.2013.084

Han TS, Wu Q, Hou XH *et al*. Frequent introgressions from diploid species contribute to the adaptation of the tetraploid Shepherd's purse (*Capsella bursa-pastoris*). *Molecular Plant* 2015;**8**:427–38. https://doi.org/10.1016/j.molp.2014.11.016

Hoban S, Kelley JL, Lotterhos KE *et al*. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist* 2016;**188**:379–97. https://doi.org/10.1086/688018

Jørgensen MH, Ehrich D, Schmickl R *et al*. Interspecific and interploidal gene flow in Central European Arabidopsis (Brassicaceae). *BMC Evolutionary Biology* 2011;**11**:346–13. https://doi.org/10.1186/1471-2148-11-346

Kiedrzyński M, Zielińska KM, Jedrzejczyk I *et al*. Tetraploids expanded beyond the mountain niche of their diploid ancestors in the mixed-ploidy grass *Festuca amethystina* L. *Scientific Reports* 2021;**11**:18735. https://doi.org/10.1038/s41598-021-97767-6

Kopelman NM, Mayzel J, Jakobsson M *et al*. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 2015;**15**:1179–91. https://doi.org/10.1111/1755-0998.12387

Kryvokhyzha D, Ross-Ibarra J, Milesi P *et al*. Towards the new normal: transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (Capsella bursa-pastoris). *PLOS Genetics* 2019;**15**:e1008131. https://doi.org/10.1371/journal.pgen.1008131

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012;**9**:357–9. https://doi.org/10.1038/nmeth.1923

Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 2009;**25**:1841–2. https://doi.org/10.1093/bioinformatics/btp328

Leal BSS, Brandão MM, Palma-Silva C *et al*. Differential gene expression reveals mechanisms related to habitat divergence between hybridizing orchids from the Neotropical coastal plains. *BMC Plant Biology* 2020;**20**:1–14.

Levin DA. Minority cytotype exclusion in local plant populations. *Taxon* 1975;**24**:35–43. https://doi.org/10.2307/1218997

Levin DA. *The Role of Chromosomal Change in Plant Evolution*. New York: Oxford University Press, 2002.
Li W, Godzik A. CD-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;**22**:1658–9. https://doi.org/10.1093/bioinformatics/btl158
Liu X, Fu YX. Exploring population size changes using SNP frequency spectra. *Nature Genetics* 2015;**47**:555–9. https://doi.org/10.1038/ng.3254
Liu X, Fu YX. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology* 2020;**21**:1–9.
Lopes JM, de Carvalho HH, Zorzatto C *et al*. Genetic relationships and polyploid origins in the Lippia alba complex. *American Journal of Botany* 2020;**107**:466–76.
López-Jurado J, Mateos-Naranjo E, Balao F. Niche divergence and limits to expansion in the high polyploid *Dianthus broteri* complex. *New Phytologist* 2019;**222**:1076–87. https://doi.org/10.1111/nph.15663
Lynch M. Evolution of the mutation rate. *Trends in Genetics* 2010;**26**:345–52. https://doi.org/10.1016/j.tig.2010.05.003
Mallet J, Besansky N, Hahn MW. How reticulated are species? *Bioessays* 2016;**38**:140–9. https://doi.org/10.1002/bies.201500149
Marburger S, Monnahan P, Seear PJ *et al*. Interspecific introgression mediates adaptation to whole genome duplication. *Nature Communications* 2019;**10**:1–11.
Marchant DB, Soltis DE, Soltis PS. Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *New Phytologist* 2016;**212**:708–18.
McCoy RC, Garud NR, Kelley JL *et al*. Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Molecular Ecology* 2014;**23**:136–50. https://doi.org/10.1111/mec.12591
Meirmans PG, Liu S, van Tienderen PH. The analysis of polyploid genetic data. *The Journal of Heredity* 2018;**109**:283–96. https://doi.org/10.1093/jhered/esy006
Mergeay J, Aguilera X, Declerck S *et al*. The genetic legacy of polyploid Bolivian *Daphnia*: the tropical Andes as a source for the North and South American *D. pulicaria* complex. *Molecular Ecology* 2008;**17**:1789–800. https://doi.org/10.1111/j.1365-294X.2007.03679.x
Molina-Henao YF, Hopkins R. Autopolyploid lineage shows climatic niche expansion but not divergence in *Arabidopsis arenosa*. *American Journal of Botany* 2019;**106**:61–70. https://doi.org/10.1002/ajb2.1212
Monnahan P, Kolář F, Baduel P *et al*. Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nature Ecology & Evolution* 2019;**3**:457–68. https://doi.org/10.1038/s41559-019-0807-4
Moraes AP, Chinaglia M, Palma-Silva C *et al*. Interploidy hybridization in sympatric zones: the formation of *Epidendrum fulgens*× *E. puniceoluteum* hybrids (Epidendroideae, Orchidaceae). *Ecology and Evolution* 2013;**3**:3824–37. https://doi.org/10.1002/ece3.752
Naimi B. usdm: Uncertainty analysis for species distribution models, R package version 2023, **1**:1-19. https://cran.r-project.org/web/packages/usdm/usdm.pdf
Nielsen H. Predicting secretory proteins with SignalP. In: Kihara D (ed.), *Protein Function Prediction*. New York, NY: Humana Press, 2017, 59–73
Novikova PY, Brennan IG, Booker W *et al*. Polyploidy breaks speciation barriers in Australian burrowing frogs *Neobatrachus*. *PLoS Genetics* 2020;**16**:e1008769. https://doi.org/10.1371/journal.pgen.1008769
Patton AH, Margres MJ, Stahlke AR *et al*. Contemporary demographic reconstruction methods are robust to genome assembly quality: a case study in Tasmanian devils. *Molecular Biology and Evolution* 2019;**36**:2906–21. https://doi.org/10.1093/molbev/msz191
Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. *Molecular Ecology* 2016;**25**:2337–60. https://doi.org/10.1111/mec.13557
Pegoraro L, Cafasso D, Rinaldi R *et al*. Habitat preference and flowering-time variation contribute to reproductive isolation between diploid and autotetraploid *Anacamptis pyramidalis*. *Journal of Evolutionary Biology* 2016;**29**:2070–82. https://doi.org/10.1111/jeb.12930
Pembleton LW, Cogan NO, Forster JW. StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources* 2013;**13**:946–52. https://doi.org/10.1111/1755-0998.12129
Pessenda LCR, Gouveia SEM, Aravena R *et al*. Holocene fire and vegetation changes in southeastern Brazil as deduced from fossil charcoal and soil carbon isotopes. *Quaternary International* 2004;**114**:35–43.
Pessoa EM, Nollet F, Magalhães RF *et al*. Nuclear–plastid discordance indicates past introgression in *Epidendrum* species (Laeliinae: Orchidaceae) with highly variable chromosome numbers. *Botanical Journal of the Linnean Society* 2022;**199**:357–71.
Pinheiro F, Cozzolino S. *Epidendrum* (Orchidaceae) as a model system for ecological and evolutionary studies in the Neotropics. *Taxon* 2013;**62**:77–88. https://doi.org/10.1002/tax.621007.
Pinheiro F, Koehler S, Corrêa AM *et al*. Phylogenetic relationships and infrageneric classification of *Epidendrum* subgenus *Amphiglottium* (Laeliinae, Orchidaceae). *Plant Systematics and Evolution* 2009;**283**:165–77. https://doi.org/10.1007/s00606-009-0224-2
Pinheiro F, De Barros F, Palma-Silva C *et al*. Hybridization and introgression across different ploidy levels in the Neotropical orchids *Epidendrum fulgens* and *E. puniceoluteum* (Orchidaceae). *Molecular Ecology* 2010;**19**:3981–94. https://doi.org/10.1111/j.1365-294X.2010.04780.x
Potter SC, Luciani A, Eddy SR *et al*. HMMER web server: 2018 update. *Nucleic Acids Research* 2018;**46**:W200–4. https://doi.org/10.1093/nar/gky448
Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2021;**155**:945–59. https://doi.org/10.1093/genetics/155.2.945
Quast C, Pruesse E, Yilmaz P *et al*. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 2012;**41**:D590–6. https://doi.org/10.1093/nar/gks1219
Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2. https://doi.org/10.1093/bioinformatics/btq033
Rajakaruna N. Lessons on evolution from the study of edaphic specialization. *The Botanical Review* 2017;**84**:39–78. https://doi.org/10.1007/s12229-017-9193-2
Ramsey J. Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences of the United States of America* 2011;**108**:7096–101. https://doi.org/10.1073/pnas.1016631108
Ramsey J, Schemske DW. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* 2002;**33**:589–639. https://doi.org/10.1146/annurev.ecolsys.33.010802.150437
Ramsey J, Ramsey TS. Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 2014;**369**:20130352. https://doi.org/10.1098/rstb.2013.0352
Rheindt FE, Fujita MK, Wilton PR *et al*. Introgression and phenotypic assimilation in *Zimmerius* flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. *Systematic Biology* 2014;**63**:134–52. https://doi.org/10.1093/sysbio/syt070
Roux C, Pannell JR. Inferring the mode of origin of polyploid species from next-generation sequence data. *Molecular Ecology* 2015;**24**:1047–59. https://doi.org/10.1111/mec.13078
Scarano FR. Structure, function and floristic relationships of plant communities in stressful habitats marginal to the Brazilian Atlantic rainforest. *Annals of Botany* 2002;**90**:517–24. https://doi.org/10.1093/aob/mcf189
Schmickl R, Yant L. Adaptive introgression: how polyploidy reshapes gene flow landscapes. *New Phytologist* 2021;**230**:457–61. https://doi.org/10.1111/nph.17204
Schrider DR, Shanku AG, Kern AD. Effects of linked selective sweeps on demographic inference and model selection. *Genetics* 2016;**204**:1207–23. https://doi.org/10.1534/genetics.116.190223

Seehausen O, Butlin RK, Keller I *et al.* Genomics and the origin of species. *Nature Reviews Genetics* 2014;**15**:176–92. https://doi.org/10.1038/nrg3644

Shastry V, Adams PE, Lindtke D *et al.* Model-based genotype and ancestry estimation for potential hybrids with mixed-ploidy. *Molecular Ecology Resources* 2021;**21**:1434–51. https://doi.org/10.1111/1755-0998.13330

Simão FA, Waterhouse RM, Ioannidis P *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31**:3210–2. https://doi.org/10.1093/bioinformatics/btv351

Sobel JM, Stankowski S, Streisfeld MA. Variation in ecophysiological traits might contribute to ecogeographic isolation and divergence between parapatric ecotypes of *Mimulus aurantiacus*. *Journal of Evolutionary Biology* 2019;**32**:604–18. https://doi.org/10.1111/jeb.13442

Soltis PS, Soltis DE. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences of the United States of America* 2000;**97**:7051–7. https://doi.org/10.1073/pnas.97.13.7051

Soltis PS, Soltis DE. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 2009;**60**:561–88. https://doi.org/10.1146/annurev.arplant.043008.092039

Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* 2015;**4**:s13742–015.

Sonnhammer EL, Von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. In: Glasgow J *et al.* (eds.), *ISMB '98: Proceedings of the 6th International Conference on Intelligent Systems for Molecular Biology*. Vol. **6**. Palo Alto, California: AAAI Press, 1998, 175–82.

Spier D, Gerum HL, Noernberg MA *et al.* Flood regime as a driver of the distribution of mangrove and salt marsh species in a subtropical estuary. *Journal of Marine Systems* 2016;**161**:11–25. https://doi.org/10.1016/j.jmarsys.2016.05.004

Sujii PS, Cozzolino S, Pinheiro F. Hybridization and geographic distribution shapes the spatial genetic structure of two co-occurring orchid species. *Heredity* 2019;**123**:458–69. https://doi.org/10.1038/s41437-019-0254-7

Supek F, Gibas C, Bošnjak M *et al.* REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 2011;**6**:e21800. https://doi.org/10.1371/journal.pone.0021800

Sutherland BL, Galloway LF. Postzygotic isolation varies by ploidy level within a polyploid complex. *The New Phytologist* 2017;**213**:404–12. https://doi.org/10.1111/nph.14116

Tavares MM, Ferro M, Leal BSS *et al.* Speciation with gene flow between two Neotropical sympatric species (*Pitcairnia* spp.: Bromeliaceae). *Ecology and Evolution* 2022;**12**:e8834.

The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Research* 2014;**43**:D204–DD212. https://doi.org/10.1093/nar/gku989

Van der Auwera GA, Carneiro MO, Hartl C *et al.* From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics* 2013;**43**:11–0.

Warren DL, Glor RE, Turelli M. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution; International Journal of Organic Evolution* 2008;**62**:2868–83. https://doi.org/10.1111/j.1558-5646.2008.00482.x

Warren DL, Matzke NJ, Cardillo M *et al.* ENMTools 1.0: an R package for comparative ecological biogeography. *Ecography* 2021;**44**:504–11. https://doi.org/10.1111/ecog.05485

Wei N, Cronn R, Liston A *et al.* Functional trait divergence and trait plasticity confer polyploid advantage in heterogeneous environments. *The New Phytologist* 2019;**221**:2286–97. https://doi.org/10.1111/nph.15508

Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* 2006;**38**:1358. https://doi.org/10.2307/2408641

Wolf JB. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources* 2013;**13**:559–72. https://doi.org/10.1111/1755-0998.12109

Yant L, Hollister JD, Wright KM *et al.* Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Current Biology: CB* 2013;**23**:2151–6. https://doi.org/10.1016/j.cub.2013.08.059

Zhang H, Zheng R, Wang Y *et al.* The effects of Arabidopsis genome duplication on the chromatin organization and transcriptional regulation. *Nucleic Acids Research* 2019;**47**:7857–69. https://doi.org/10.1093/nar/gkz511

Zohren J, Wang N, Kardailsky I *et al.* Unidirectional diploid–tetraploid introgression among British birch trees with shifting ranges shown by restriction site-associated markers. *Molecular Ecology* 2016;**25**:2413–26. https://doi.org/10.1111/mec.13644