GBE

Unraveling the Genome of *Epidendrum Fulgens*: Demographic History and Gene Family Dynamics in a Resilient Neotropical Orchid

Jacqueline S. de Mattos [1,2,3], Paulo Aecyo [1], Kyle Keepers [2], Marília M. Tavares [1], Fernando Della-Rosa [4], Clarisse Palma-Silva [1], Nolan C. Kane [2], Fábio Pinheiro [1,*], Diego M. Riaño-Pachon [4]

[1]Laboratory of Evolutionary Ecology and Plant Genomics, Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Brazil

[2]Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, CO, USA

[3]Instituto Tecnológico Vale Desenvolvimento Sustentável, Departamento de Genômica Ambiental, Belém, PA, Brazil 66055-090

[4]Laboratory of Computational, Evolutionary and Systems Biology, Center for Nuclear Energy in Agriculture/CENA, University of São Paulo, Piracicaba, Brazil

*Corresponding author: E-mail: biopin@unicamp.br

Accepted: February 23, 2026

ABSTRACT

Epidendrum L. (Orchidaceae) is one of the largest plant genera in the neotropics, accounting for more than 1,500 species and a high variety of morphological, physiological, and ecological adaptations. However, addressing evolutionary and ecological questions in the genus is a challenge due to the lack of publicly available genomic resources. Here, we present the first chromosome-scale genome assembly from the orchid genus *Epidendrum*, *E. fulgens* ($2n = 24$, $2C = 2.4$ pg), a perennial-terrestrial that occurs in coastal restingas and inland granitic rock outcrops along the Brazilian Atlantic Forest. The genome sequencing was performed using PacBio HiFi and Omni-C technologies, and the final assembly comprised 12 chromosome-scale scaffolds, representing the 12 pseudochromosomes expected for *E. fulgens*. The primary assembly pseudochromosomes captured 97.3% of BUSCO genes, with an N50 of 88.6 Mbp and L50 of 5; similar metrics were observed for both locally phased haplotypes. Approximately 77% of the genome consisted of transposable elements, and the annotation identified 30,830 protein-coding genes and 6,141 associated Gene Ontology (GO) terms. We also found two events of population expansion and retraction in *E. fulgens* demographic history, and many expanded and contracted gene families. GO enrichment analysis showed functional categorization for the expanded and contracted gene sets (cell cycle regulation, cell shape and morphology, osmoregulation, environmental plasticity, and stress response), evidencing the traits that are functionally important for this species' adaptation and persistence in nutrient-poor soils and extreme environments. With this study, we highlight the need to further explore the genomic space in the Orchidaceae family and in the Neotropical region, providing new resources for future studies aiming to investigate the origin and resilience of Neotropical flora.

Key words: genome assembly, genome evolution, orchid biology, bioinformatics, Orchidaceae, neotropical biodiversity.

Significance

Epidendrum stands as one of the largest genera both within the orchid family and within angiosperms as a whole; however, genomic resources for this group remain scarce, limiting advances in plant breeding, evolutionary biology, and conservation. Here, we present a high-quality, chromosome-level genome assembly for *Epidendrum fulgens*, which uncovers detailed and dated events of population contraction and expansion, as well as specific gene families functionally important for the persistence of plant species in nutrient-poor soils and extreme environments. The genomic data presented here provide critical insights into orchid evolution in harsh environments such as coastal sand dune vegetation and establish a valuable foundation for future research, particularly in the Neotropics.

Introduction

Many advances in genome sequencing and assembly technologies have been recently witnessed by the scientific community (e.g. Hi-C and Omni-C; Porubsky et al. 2020; Sun et al. 2022), enabling a huge increase in our knowledge about genome structure and function. Despite such advances, plant genomics research has been mainly focused on model (Bevan and Walsh 2005) and crop species (Morrell et al. 2012), followed by some major forest trees (Myburg et al. 2011; Nystedt et al. 2013) and medicinal plants (Zhang et al. 2017; Xu et al. 2022) with several taxonomic gaps throughout plant lineages and geographical distribution (Marks et al. 2021). Indeed, large Angiosperms groups such as Orchidaceae, Begoniaceae, and Asteraceae are sparse in genetic resources compared to the number of species in each family, precluding deep molecular and evolutionary studies (Cornwell et al. 2018). The assembly of plant genomes is a key component in the generation of genomic resources that are able to aid in conservation practices, improving breeding systems (Goodstein et al. 2012; Morrell et al. 2012) and understanding evolutionary processes in less-studied regions (Pont et al. 2019; Loiseau et al. 2021), and therefore, it is essential to apply it to less studied systems within high conservation priority areas, such as those in the neotropical region.

Viridiplantae (green plants) contains around 500,000 species (Corlett 2016), but approximately 1,800 plant species have their genome sequenced as of this study (Kress et al. 2022; Sun et al. 2022; Schwacke et al.2025), representing an outstandingly low percentage of the diversity of green plants. Additionally, plant genomes might be tricky to assemble, considering the scale to which their genomes may vary—the largest plant genome (*Tmesipteris oblanceolata*) is 160 Gbp (Fernández et al. 2024), whereas one of the smallest (*Genlisea margaretae*) is only 0.065 pg (Tsai et al. 2017). Additional assembly challenges include complex genomic features like chromosomal rearrangements (Sun et al. 2022) and large amounts of repetitive sequences, heterozygosity, and frequent polyploidy (Kyriakidou et al. 2018; Sun et al. 2022). Among Angiosperms, or flowering plants, Orchidaceae is one of the largest families, accounting for more than 880 genera and 25,000 species, which are able to grow in a wide variety of habitats and occupy many different niches (Givnish et al. 2015). Orchids present extremely diverse floral morphology and adaptations to different kinds of pollinators, even among close relatives, and most of them inhabit tropical forests as epiphytes (Cozzolino and Widmer 2005; Givinish et al. 2015). However, much of the Orchidaceae diversity is under threat, with more than half of all orchid species estimated to be at risk of extinction (Antonelli et al. 2023; Bachman et al 2023). Currently, only a few chromosome-scale genome assemblies are publicly available for the family, with only one other orchid genome (*Vanilla planifolia*) from the Neotropics, the region that harbors most of orchid diversity (Pérez-Escobar et al. 2024a, 2024b) in the world, which leads to an urgent need for more high quality assembled genomes of this ecologically important plant group to support basic science, conservation and evolutionary biology studies.

Among orchids, *Epidendrum* L. is one of the largest genera of neotropical plants, with approximately 2,000 species described (Karremans 2021). This megadiverse genus also exhibits remarkable diversity in morphological features and ecological adaptations (Pinheiro and Cozzolino 2013). Because of their great diversity in ecological interactions, *Epidendrum* is considered a good model for ecological and evolutionary studies, mainly on speciation (Pessoa et al. 2012; Cardoso-Gustavson et al. 2018), reproductive isolation (Pinheiro et al. 2016; Arida et al. 2021), and habitat selection (Leal et al. 2020, 2025). *Epidendrum fulgens* is a perennial-terrestrial, which is pollinated by butterflies (Arida et al. 2025) and is found on sand dunes along the coast and in rocky outcrops in inland regions of south and southeastern Brazil, in the Brazilian Atlantic Forest (BAF) (de Mattos et al. 2023), considered one of the world's hotspots for biodiversity (Kareiva and Kareiva 2017). The number and lengths of chromosomes for many species

Table 1 Assembly statistics comparing the primary (collapsed) genome with both haplophases (Hap1, Hap2), including the quality assessment from the Embryophyta BUSCO gene set. Curation methods described in Supplementary Methods

Post-curation assembly	Primary	Phased haplotype 1	Phased haplotype 2
Number of scaffolds	12	12	12
Total length (bp)	1,166,745,331	1,158,411,174	1,150,982,409
Largest contig (bp)	182,107,267	180,353,035	179,214,105
N50	96,305,604	95,686,049	95,684,482
L50	5	5	5
Complete BUSCOs	1,571 (97.3%)	1,571 (97.3%)	1,563 (96.8%)
Single-copy BUSCOs	1,547 (95.8%)	1,545 (95.7%)	1,538 (95.3%)
Duplicated BUSCOs	24 (1.5%)	26 (1.6%)	25 (1.5%)
Fragmented BUSCOs	11 (0.7%)	11 (0.7%)	14 (0.9%)
Missing BUSCOs	32 (2.0%)	32 (2.0%)	37 (2.3%)
Total BUSCOs searched	1614	1614	1614

of *Epidendrum* have been previously estimated and are extremely variable (2n = 24 in *E. fulgens* and 2n = 240 in *E. cinnabarinum*) (de Assis et al. 2013; Moraes et al. 2013), whereas the *E. fulgens* diploid genome size was estimated at 2.4 pg by flow cytometry (unpublished results). In this work, we use a combination of sequencing strategies, involving PacBio Single-Molecule Real-Time (SMRT) HiFi sequencing and chromatin conformation capture (Dovetail OmniC) technology to provide a chromosome-level genome assembly of the diploid neotropical orchid *E. fulgens* (2C = 2.4pg, 2*n* = 24; Moraes et al. 2013). RNA-seq from different plant tissues was also used to aid in gene prediction and functional annotation. Despite being a megadiverse genus and one of the most speciose genera of orchids, *Epidendrum* genomic resources are still scarce. Here, we reduce this genomic resource gap in orchid research by generating the first chromosome-scale genome assembly for the megadiverse *Epidendrum* genus and for any orchid species in South America. This high-quality assembled genome will be an important resource, providing a foundation for future studies involving conservation genetics, comparative genomics, and other applied and basic research in this genus and related genera.

Results

Chromosome-scale Genome Assembly

We generated 28× sequence coverage for PacBio Hifi sequencing, with a total of 38.7 gigabase-pairs reads and 2,741,988 total reads. For Omni-C sequencing, data presented approximately 30× sequence coverage and 82 gigabase-pairs of reads, with reads of approximately 150 bp in length. Any contamination was detected in the raw sequencing reads (Fig. S1). We assembled two phased haplotypes (consisting of both haploid genomes) and the primary assembly (a collapsed haploid genome used for downstream analyses). Our primary assembly (prior to curation) consisted of 712 scaffolds and a genome size of 1.2 Gbp, which agrees with the expected haploid genome size estimated by flow cytometry (Moraes et al. 2013). We recovered a total of 12 chromosome-scale scaffolds, which ranged from 182.11 Mb to 75.70 Mb in length (Fig. S2) and contained around 97.5% of the genome, presenting a scaffold N50 value of 88.65 Mb. After manual curation (PreTextView and Jupiterplot, Fig. S3), the primary assembly measured 1.17 Gb, with the largest contig of 182 Mbp. Manual curation proved challenging due to numerous small contigs that lacked Omni-C signal, preventing their chromosomal placement. These remained as unplaced contigs in the primary assembly.

Benchmarking Universal Single-Copy Ortholog (BUSCO) analyses indicated that the completeness of the gene set based on the embryophyta_odb10 database was 97.3%, with 95.7% being single-copy genes and 1.6% duplicated genes (Table 1). For the liliopsida_odb10 database, BUSCO scores were 93.3%, with 91.5% being single-copy genes. Most chromosomes in our assembly presented telomeres on both ends, except for chromosome 1 and chromosome 6, which presented a telomeric repeat in the middle of the sequence, even after the manual curation procedures (Fig. S6). The fully phased haplotypes spanned 1.16 Gb (haplotype 1: 1,158,411,174 bp) and 1.15 Gb (haplotype 2: 1,150,982,409 bp) (Table 1). The assemblies showed high completeness scores according to BUSCO: 97.3% (haplotype1) and 96.8% (haplotype2) for embryophyta_odb10 (Table 1), and 93.0% (haplotype 1) and 92.8% (haplotype 2) for liliopsida_odb10. Our assessment of contamination with BlobTools indicated that 49% of our raw reads belonged to Orchidaceae, whereas 44.9% were classified as unmapped. The remaining families, where the analyses indicated BLAST similarities with, were Cucurbitaceae (3.75%), Vitaceae (0.92%), Malvaceae (0.32%), and Arecaceae (0.19%), whereas only 0.10% of the reads had absolutely no BLAST hits.

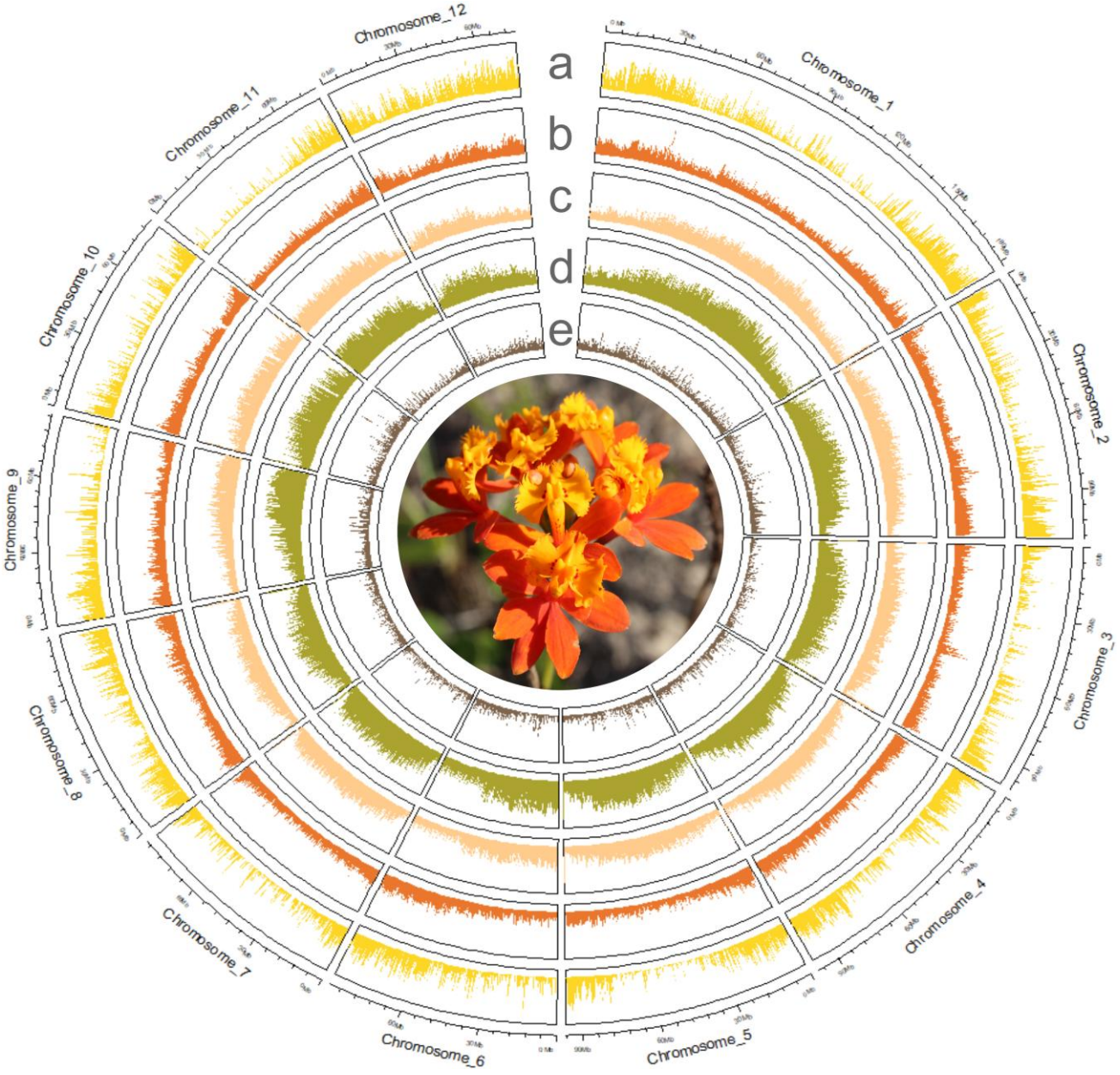

Fig. 1. Circos plot of *Epidendrum fulgens'* genomic features. a) Gene density; b) density of all mutator TEs; c) density of all Gypsy TEs; d) density of all LTR retrotransposon TEs; and e) density of all CACTA TEs.

According to this analysis, we concluded that our data did not have a strong level of contamination, and therefore, we kept all the raw reads into the assembly procedure.

Repetitive Content, Annotation, and Gene Prediction

EDTA pipeline identified 976,074,862 bp (976 Mbp) of repetitive elements, representing 76.89% of the assembled genome (Fig. 1; Table 1) for the primary assembly, with most of them being long terminal repeats (LTR) of the class Gypsy (41.82%) and unknown LTRs (14.94%). Terminal inverted repeats (TIRs) were present at a smaller proportion (Mutator at 5.87% and CACTA at 3.52%). Non-LTR elements, such as LINEs, were present at a very low abundance (0.19%) (Fig. 2; Table S1). BRAKER3 pipeline predicted 30,830 protein-coding genes (Fig. 1), with a BUSCO completeness score of 93.6%, indicating high-quality gene annotation. Functional annotation using eggNOG-mapper and InterProScan identified protein domains in 24,094 protein-coding genes, classifying them into 130 COG functional categories and 6,141 Gene Ontology (GO) terms (Fig. S4). OMArk found a

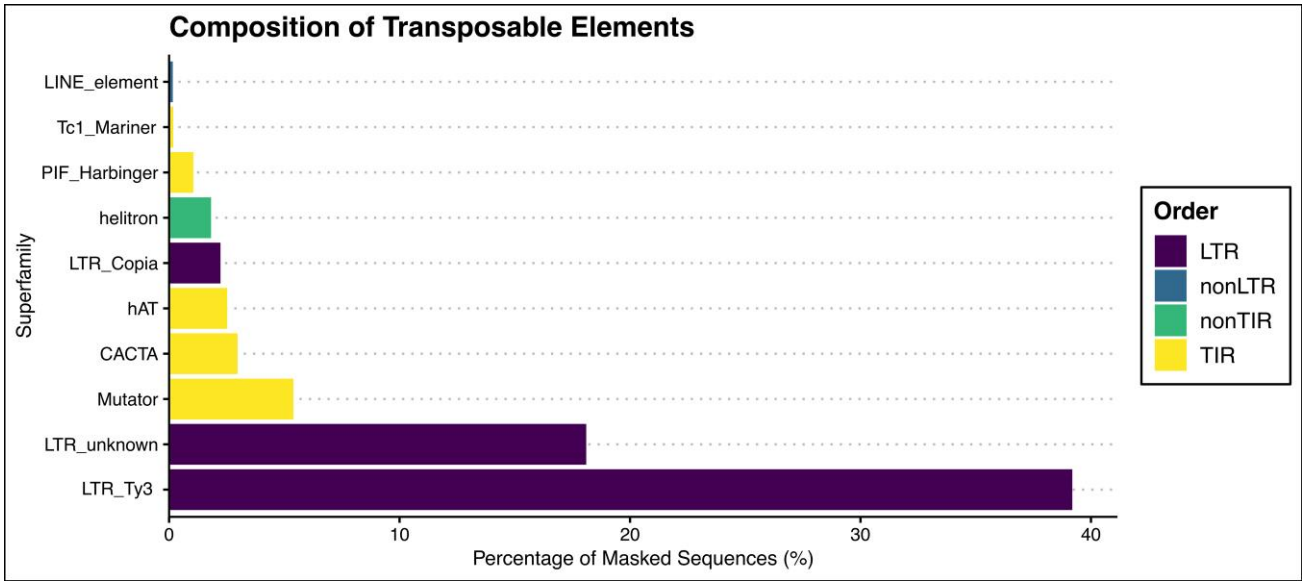

Fig. 2. Transposable elements landscape composition for *E. fulgens* genome, showing proportions of each TE class and colored by TE order (LTR, non-LTR, TIR, and non-TIR).

completeness of 85.1% when compared to the ancestral clade of Petrosaviidae (Fig. S5).

Historical Changes in Effective Population Size

Demographic reconstruction revealed a ~10 million-year evolutionary history for *E. fulgens*, characterized by two major cycles of population expansion and contraction (Fig. 3; Fig. S7). The first cycle occurred during the Late Miocene (~10 mya), followed by a second in the mid-Pleistocene (~800 kya), each marked by post-bottleneck population growth followed by a significant N_E decrease. Following these cyclical fluctuations, the population stabilized at reduced levels, maintaining a contemporary N_E of approximately 20,000, suggesting long-term impacts of historical demographic shocks on genetic diversity.

Genome Evolution

By analyzing the genomes of six plant species alongside *E. fulgens*, we generated a comparative genomics analysis including orthologous group comparisons and expansion and contraction of gene families. We were able to find relationships that indicate *E. fulgens* as being most closely related to *Phalaenopsis equestris*, and second to *Dendrobium nobile*. Orchids clustered all together, as expected, leaving *Arabidopsis thaliana* and *Oryza sativa* further out in the tree, since they represent further evolutionary relationships.

In *E. fulgens*, we identified 1,009 expanded and 1,504 contracted gene families. A total of 4,138 gene families were unique to *E. fulgens*, whereas 9,505 were shared among all species analyzed. Additionally, 565 orthogroups in *E. fulgens* were composed of single-copy orthologous genes. *E. fulgens* shared 837 orthogroups with *Dendrobium nobile* and 1,252 with *Phalaenopsis equestris*, its closest relatives in the dataset. In contrast, it shared only 211 and 373 orthogroups with the more distantly related orchids *Apostasia shenzhenica* and *Vanilla planifolia*, respectively. GO enrichment analyses revealed two significantly enriched GO terms (adjusted P-value < 0.05 with Bonferroni correction) among contracted gene families and 26 among expanded gene families (adjusted P-value < 0.05 with Bonferroni correction) (tables S1 and S2), highlighting functional categories that have likely been lost or gained in *E. fulgens* during its evolutionary history. The two GO terms that are significantly contracted relative to the other taxa in our analysis belong to the broad GO category of Cellular Component (Table S2). Conversely, among the 26 GO terms that are significantly expanded relative to outgroup taxa, 9 belong to biological process, 5 to molecular function, and 12 to cellular component (Table S3). The two GO terms in the contracted gene set were "plasma membrane" (GO:0005886) and "cell periphery" (GO:0071944). In contrast, the GO terms of significantly expanded gene families were predominantly related to cell cycle regulation, cell organization, environmental plasticity, and stress response.

Discussion

In this work, we present the first chromosome-scale genome assembly of the neotropical orchid *E. fulgens,* being also the first one for the megadiverse genus *Epidendrum*.

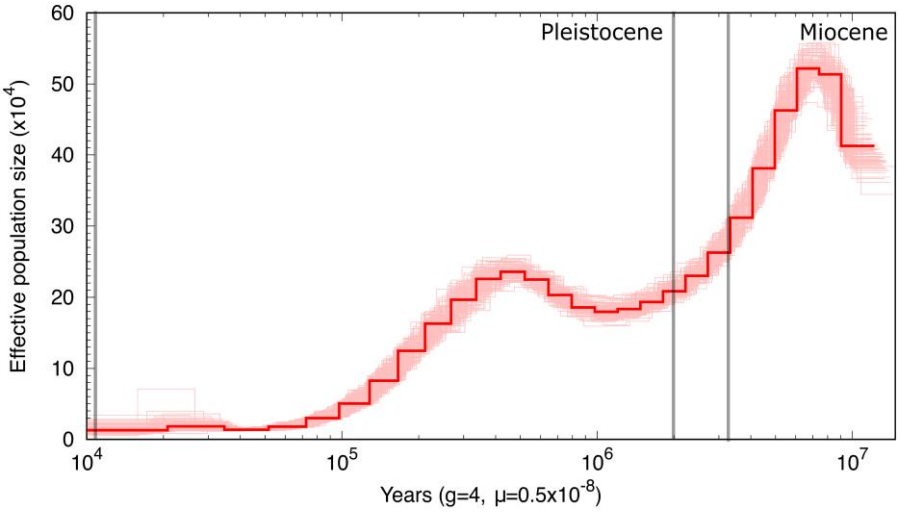

Fig. 3. Changes in the ancestral effective population size for *E. fulgens*. The dark red line shows the consensus effective population size through time, with the bootstraps represented by lighter red lines. The metrics *g* and μ represent generational time and mutation rate, respectively. The vertical gray lines show the delimitation of geological epochs (Miocene and Pleistocene).

Despite being the largest flowering plant family (Givnish et al. 2015) and containing many species with ornamental and medicinal importance, Orchidaceae has been scarcely sequenced with very few publicly available genomic resources—a total of only 11 chromosome-scale genome assemblies representing 5 genera for the entire Orchidaceae family, and none for the tribe Epidendroideae (Lindl. ex Endl., 1837) were found on NCBI (as of June, 2025). Currently sequenced orchid genomes predominantly belong to species of ornamental breeding, such as *Phalaenopsis equestris* (Cai et al. 2015), those recognized for therapeutic properties, like *Dendrobium* (Yan et al. 2015, Chen et al. 2024, Li et al. 2025), and selected genera that are famous for their species-specific interaction with pollinators, such as *Angraecum* (Jiang et al. 2025), *Gongora* (Guizar Amador et al. 2024), and *Ophrys* (Russo et al. 2024). Notably, only one other orchid genome from the neotropical region, namely *Vanilla planifolia* (Piet et al. 2022), has been sequenced to date. Here, we used this newly sequenced *Epidendrum* genome to also study the demographic history and genome evolution of this species, having found an expected pattern of population expansion and decline based on previous biogeographic research on *E. fulgens'* range and many functionally important genes that have been acquired and lost in *E. fulgens* in relation to other related orchids and monocots.

Generation of a High-quality Chromosome-scale Genome Assembly and Annotation

The past 20 years have seen a remarkable surge in high-quality chromosome-scale plant genomes, driven especially by the advent of long-read technologies like PacBio HiFi and Oxford Nanopore, along with chromatin conformation capture methods. (e.g. Hi-C and Omni-C; Kang et al. 2015, Porubsky et al. 2020, Sun et al. 2022). Those techniques provided a new and feasible way of attaining the generation of chromosome-scale genome assemblies (Marks et al. 2021, Rhie et al. 2021, Hotaling et al. 2021) that have been crucial for studies that aim to understand species origins and evolutionary history through comparative genomics (Lu et al. 2019, Hundsdoerfer et al. 2022, Sahu et al. 2023) and to investigate the basis of genetic diversity among species and populations (Hibrand Saint-Oyant et al 2018, Song et al. 2023).

We assembled a chromosome-level genome for *E. fulgens,* recovering a total of 12 pseudomolecules, which agree with the expected chromosome number for *E. fulgens (*de Assis et al. 2013, Moraes et al. 2013*)*. Chromosome lengths were also in accordance with the expected sizes for *E. fulgens* karyotype (de Assis et al. 2013, Moraes et al. 2013), Notwithstanding, contiguity parameters, including the embryophyta BUSCO of over 95%, confirm that our genome assembly was of high quality and accurate to the expectations for this species. Despite having a highly complete and contiguous assembly, we also faced some difficulties regarding genome curation due to very small contigs that presented no Hi-C signal, which made it challenging to place them inside the assembled chromosomes. Many factors can influence the assembly and scaffolding of genomes, one of them being heterozygosity (Kong et al. 2023). We tried to get around this issue by selecting an individual from

a population that is the least heterozygous in *E. fulgens* distribution, but considering the reproductive mode of this species being outcrossing, some heterozygosity is still expected, which can compromise the assembly process. In addition, some biological aspects of chromatin organization might cause fluctuations in Hi-C contact maps, which could potentially lead to more issues and challenges in the scaffolding of short contigs and estimating the distances between loci during assembly (Yamaguchi et al. 2021, Kong et al. 2023).

We found 30,830 protein-coding genes in the *E. fulgens'* genome, which compares to the expected number of genes for other members of the Orchidaceae family (Cai et al. 2015; Zhang et al. 2021; Yan et al. 2015) and confirms the expectated of number of genes for this species. BUSCO completeness for the annotation, however, was only 93.6%, evidencing a very good annotation but with room for improvement in terms of finding more single-copy orthologs in the set of predicted genes.

Demographic History and Effective Population Size Changes

The evolutionary history found for *E. fulgens* with the PSMC approach has traced its history back to around 10 mya and shows two main events of population decline and expansion, followed by a steady period reaching the population's lowest levels in more recent times. To date, there have been a few studies on orchid population size dynamics (Alexandersson and Ågren 1996; Gillman and Dodd 1998), whereas the PSMC approach has been more recently used for other orchid species in a study about the *Apostasia ramifera* genome (Zhang et al. 2021). The method successfully recovered important aspects of orchid demographic trajectories, bounded by upper and lower mutation rate estimates. Most of the studied species exhibited a period of population growth between 10 million and 1 million years ago, followed by bottlenecks within the last 100,000 years (Zhang et al. 2021). While the general demographic patterns were broadly similar across species, *E. fulgens* appears to follow a distinct evolutionary trajectory. This species may have undergone two major bottlenecks rather than one, with its peak effective population size occurring during the mid to late Miocene—a period when other orchids showed relatively low and stable population sizes that only began to increase during the Pleistocene.

Population declines found with the PSMC analysis can help understand further *E. fulgens'* evolutionary history when accounting for more ancient events that led to the generation and expansion of the forests and dune environments to which this species is now adapted, mainly in the Atlantic Forest complex. Many climatic fluctuations happened during the Quaternary period (2.6 mya), which could have shaped genetic patterns in *E. fulgens'* populations in terms of expansion of forests and contraction of grasslands after the LGM, although these are much more recent events (Alexandersson and Ågren 1996; Gillman and dodd, 1998; Bai et al. 2017). The southern Brazil coastal plains are formed mainly by marine, transitional, and continental Quaternary deposits, which were influenced by the sea-level variations during the formation of the Atlantic Forest complexes, such as the vegetation in which most populations of *E. fulgens* occur on (restingas) (Kuhn et al. 2023). The demographic patterns uncovered in this study reflect ancient events (1–10 million years ago), suggesting long-term persistence of *E. fulgens* in its current range. Until now, population-level studies of *E. fulgens* have primarily detected more recent demographic changes, such as those associated with the Last Glacial Maximum (LGM, ~20,000 years ago; Pinheiro et al. 2011; Sujii et al. 2019; de Mattos et al. 2023). For instance, Pinheiro et al. (2011) reported strong signals of population decline—indicative of bottlenecks—near the northern limit of the species' distribution, where the individual analyzed in our study was collected. Following the LGM, sea-level rise and forest expansion restricted *E. fulgens'* main habitat (restingas) to a narrow coastal corridor only 20–50 meters wide in northern areas, likely contributing to population isolation and reduced density. Notably, Pinheiro et al. (2011) also found evidence for long-term population persistence rather than recent colonization in this region, which aligns with our findings.

Genome Evolution and Gene Functional Strategies

The Orchidaceae family is estimated to have originated between 76 and 112 Mya, during the Cretaceous period (Gvinish et al. 2015; Zhang et al. 2023a; Pérez-Escobar et al. 2024a, 2024b), supporting the hypothesis of an "ancient origin" for orchids (Pérez-Escobar et al. 2024a, 2024b). Initial diversification may have occurred in Laurasia, aligning with the broader view that this region played a key role in the diversification of many tropical plant and animal lineages (Pérez-Escobar et al. 2024a, 2024b). However, alternative hypotheses suggest different timings and locations for the family's early diversification, including an origin in Australia during the mid-Cretaceous (90–120 Mya) (Givnish 2015), highlighting ongoing uncertainties in the evolutionary history of Orchidaceae. While orchids are thought to have undergone rapid diversification primarily during the Cenozoic era, higher epidendroids are estimated to have diversified around 37.9–30.8 Mya (Givnish

2015; Zhang et al. 2023a). Nevertheless, biogeographic models proposed by Pérez-Escobar et al. (2024a, 2024b), (2024b) suggest the presence of epidendroids as early as the Paleocene (55–50 Mya), consistent with fossil evidence for the subfamily Epidendroideae. *E. fulgens* belongs to the tribe Epidendreae, which comprises five subtribes: Bletiinae, Chysiinae, Laeliinae, Ponerinae, and Pleurothallidinae. Molecular phylogenetic analyses indicate that within Epidendreae, the genus *Epidendrum* is most closely related to the genus *Cattleya*, both belonging to the subtribe Laeliinae (Zhao et al. 2023).

The comparative genomic analysis presented here provides important phylogenetic context for *E. fulgens*, representing the first complete and haplotype-phased genome assembly within the genus *Epidendrum*. This contrasts with many previous studies that primarily relied on molecular phylogenies based on plastid genomic regions and low-copy nuclear genes (Cardoso-Gustavson et al. 2018; Givnish et al. 2018; Pérez-Escobar et al. 2021; Zhao et al. 2023), or more recently, transcriptomic data (Zhang et al. 2023a; Wang et al. 2024, Pérez-Escobar et al. 2024a, 2024b). Among the taxa with available complete genome assemblies included in our analysis, *E. fulgens* was most closely related to *Phalaenopsis equestris* (Cai et al. 2015), followed by *Dendrobium nobile* (Xu et al. 2022). Interestingly, *E. fulgens* exhibited a more similar pattern of gene counts and number of orthologs with *D. nobile* than with *P. equestris*. Our phylogenetic estimates suggest that *E. fulgens* diverged from *P. equestris* approximately 50 Mya, and their common ancestor diverged from *D. nobile* around 55 Mya. The shared ancestor of these three species and *Vanilla planifolia* is estimated to have diverged around 80 Mya. This pattern of divergence aligns with the established subfamilial classification: both *D. nobile* and *P. equestris* belong to the Epidendroideae subfamily, whereas *V. planifolia* is a member of Vanilloideae, and *Apostasia shenzhenica* belongs to Apostasioideae—the most basal orchid subfamily and sister to all other lineages. These results are consistent with prior studies on Orchidaceae phylogeny (Pérez-Escobar et al. 2024a, 2024b) and underscore the phylogenetic placement of *E. fulgens* among the epidendroids, contributing valuable insights into the evolutionary history of this diverse plant family. It is important to note that the phylogenetic relationship within the Orchidaceae family remains unresolved and continues to be the subject of debate (Pérez-Escobar et al. 2024a, 2024b).

Contrary to the expected phylogenetic relationship, *E. fulgens* exhibited a more similar pattern of gene counts and number of orthologs with *D. nobile* than with *P. equestris* (Fig. 4). The shared orthogroups between *E. fulgens* and *D. nobile*, despite the considerable time since their divergence (38 my approximately) (Cardoso-Gustavson et al. 2018; Serna Sanchez et al. 2021), show that these orchid species retain a large number of orthologous genes and gene functions. These similarities were also not expected since these plants belong to different orchid tribes (Malaxideae and Epidendreae, respectively) and also occur in distinct biogeographic domains (Indo-Malay/Australasia and Neotropics, respectively). Genome sizes and chromosome numbers are not significantly different for these species though, considering Orchidaceae is the most variable Angiosperm family regarding genome sizes, with values ranging 168-fold (Leitch et al. 2009). Because of the similarities in gene families among distantly related orchids, our results allow us to predict and to expect even higher genomic similarities among orchid species or genera that are even more closely related. In contrast, non-orchid angiosperms such as *Arabidopsis thaliana* and *Oryza sativa* differ markedly from orchids in their ortholog content (Fig. 4).

Multiple genomic adaptations to stress

To better understand the functional categories associated with gene family expansions and contractions in *E. fulgens*, a GO enrichment analysis was performed in comparison with other analyzed species. These patterns of gene family expansion and contraction may reflect the genetic basis of some of *E. fulgens'* distinctive traits and are likely shaped by the specific evolutionary pressures this species has experienced. As a coastal plant, *E. fulgens* faces a range of environmental stressors, including intense light exposure, sandy and nutrient-poor soils, and specialized pollination dynamics. Many of the GO terms found in the expanded gene set were found to be related to the development and stability of the plant cell wall and abiotic stressors. The cortical microtubule (as well as other related gene ontology functions such as the terms "cortical cytoskeleton," "cytoplasmic microtubule," and "cortical microtubule transverse to long axis") is most plausibly associated in *E. fulgens* with salt stress (Chun et al. 2021; Hsiao et al. 2023), as the microtubules will go through changes (Dou et al. 2018) that promote cell survival under the salinity stress conditions that *E. fulgens* must endure. The expansion of polygalacturonase (PG) metabolism genes also relates to cell-morphological modifications as a response to salt stress (He et al. 2023). Polygalacturonase hydrolyzes the pectin in cell walls to increase their flexibility and elongate cells (Hadfield et al. 1998; Babu and Bayer 2014), which may be an adaptation to salt stress.

Another broad adaptation to salt stress involves the regulation of ion transport through the cell membrane,

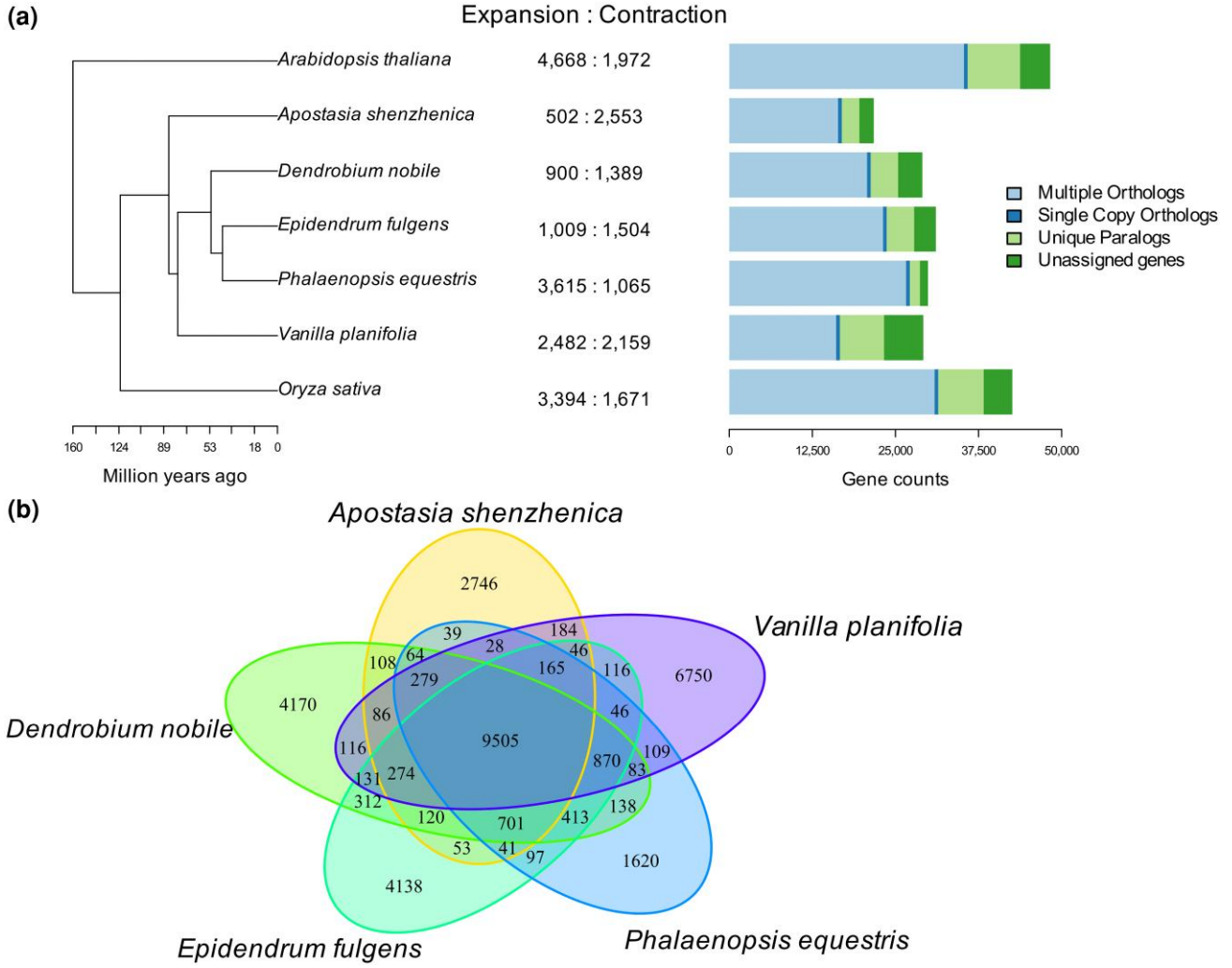

Fig. 4. Gene family evolution analysis for Epidendrum fulgens and other six species: four orchids (*D. nobile, P. equestris, A. shenzhenica,* and *V. planifolia*) and two model plants (*O. sativa* and *A. thaliana*). a) Phylogenetic tree among species showing the expansion and contraction ratio among gene families, alongside with a bar plot showing the number of protein-coding genes in each of the species. b) Venn diagram with the shared orthologous groups for the five orchid species analyzed here.

known as osmoregulation (Gunde-Cimerman et al. 2009; Bahieldin et al. 2013). We found several expanded gene families related to osmoregulation, including glycerol biosynthesis (Bahieldin et al. 2013) and alditol biosynthesis (Dichio et al. 2009), each of which acts as an intracellular compatible solute that counteracts the effect of osmotic pressure. Another pathway in this category involves the biosynthesis of proline, which also acts as an intracellular compatible solute (El Moukhtari et al. 2020), but its main adaptive role is as a nonenzymatic antioxidant—limiting the proliferation of reactive oxygen species (ROS) that arise under stressful conditions (Hayat et al 2012). Finally, the gene family regulating the transport of magnesium ions has also expanded, which is yet another adaptation to regulate the osmolarity inside and outside the salt-stressed cells.

The loss of genes in the plasma membrane functional category and the simultaneous gain in genes related to other types of environmental adaptations suggest potential mechanisms enabling this species to survive in its coastal dune habitats. Supporting this, de Lima et al. (2023, 2024) reported exceptional levels of salt tolerance in *E. fulgens* under controlled experimental conditions, likely linked to its succulent leaves and the phenotypic plasticity of several key functional traits. We propose that these patterns could reflect local adaptation and merit further investigation. Assessing the relationship between habitat type and allelic variation in these gene families could provide insights into the

genetic mechanisms underlying ecological specialization. Since *E. fulgens* is found both in salt-exposed coastal dunes and in comparatively less saline inland cliff habitats, it may exhibit population-specific adaptations in these expanded or contracted gene sets. In conclusion, the gene family expansions and contractions observed in *E. fulgens*—reflecting changes in the abundance of specific gene functions—point to genomic adaptations to the restinga and coastal environments where this species naturally occurs. These patterns highlight the diverse strategies by which the genome of *E. fulgens* may be evolving to cope with the harsh and variable conditions of its habitat.

Conclusions

This study presents the first chromosome-scale genome assembly and annotation for a species of the genus *Epidendrum*—one of the most species-rich genera within the already diverse Orchidaceae family. The genome of *E. fulgens* was fully assembled into chromosomes and completely phased into its two haplotypes using proximity-ligation sequencing data. This high-quality assembly enabled investigations into the species' demographic history and gene family evolution in comparison with other plant species. However, several analyses conducted here are constrained by the limited genomic resources currently available for Orchidaceae, underscoring the urgent need to expand sequencing efforts across more orchid taxa and other non-crop and non-model plant species. In fact, most of the orchid species used to build reference genomes were epiphytes occurring in humid tropical forests (Cai et al. 2015; Yan et al. 2015; Chen et al. 2024; Guizar Amador et al. 2024; Jiang et al. 2025; Li et al. 2025), limiting our ability to explore why orchids are able to thrive in different and more stressful habitats such as grasslands, rock outcrops, and coastal sand dunes. Importantly, this genome assembly serves as a critical resource for conservation and evolutionary research in the highly threatened Brazilian Atlantic Forest, as well as other Neotropical vegetation complexes characterized by severe habitat loss and fragmentation.

Materials and Methods

Sample Collection and Genome Sequencing

Leaf fragments from a cultivated *E. fulgens* specimen, previously gathered at Ubatuba city, in the state of São Paulo (Brazil), were collected in liquid nitrogen (Fig. 5). Frozen leaf samples were kept in dry ice and shipped to Dovetail Genomics (Cantata Bio) for Omni-C library preparation. The sample was sequenced using both PacBio SMRT and Omni-C technologies. Two flow cells of PacBio Sequel II were used (300 Gb/flow cell), and the library was prepared according to the SMRTbell HiFi protocol, which produces 10–25 kb reads. The Omni-C library was prepared using the Dovetail Omni-C Kit according to the manufacturer's protocol. The library was sequenced on an Illumina HiSeqX platform.

Chromosome-scale Genome Assembly and Manual Curation

Here, we followed the Vertebrate Genome Project pipeline v.1.6 (Rhie et al. 2020) for most of the steps of the de novo assembly. In short, we used Hifiasm v.0.16.1 software (Cheng et al. 2021) with the PacBio HiFi reads as a first pass to assemble contigs. The Hifiasm phased assembly and the Dovetail Omni-C reads were then used again in Hifiasm, generating a second assembly. We assessed the contamination of PacBio raw reads by performing analyses with Blobtools v.1.1.1 (Laetsch and Blaxter 2017), and quality checks were performed with Quast v.5.2.0 (Gurevich et al. 2013), BUSCO v.5.5.0 (Simão et al. 2015) using the embryophyta_odb10 (v.2019-11-20) and liliopsida_odb10 (v.2019-11-20) datasets, and Merqury v.1.3 (Rhie et al. 2020). Telomeres were detected by counting repeats using TIDK v. 0.1.5 (Brown et al. 2025). Haplotypic duplications in the primary assembly were identified with purge_dups v.1.2.5 (Guan et al. 2020), but for the case of our assembly, it did not substantially improve assembly quality. Scaffolding was then achieved using the Omni-C data and was carried out with YAHS v.1.1a (Zhou et al. 2023). Manual curation of the contact map was performed with the aid of PretextView v.0.2.5 (Howe et al. 2020) from the Rapid Curation Workflow to ensure that contigs were assembled in the correct order and orientation. All three genome assemblies (primary assembly and both haplophases) generated in this work were manually curated. We carefully analyzed possible structural errors, removed contaminants, and assigned the sequences into chromosomes. Collinearity analysis between the two phased genomes was performed with Jupiterplot v.1.0 (Chu 2018) in order to evaluate the assemblies and genomic synteny.

RNA Sequencing

In order to assist gene annotation, we generated RNA-seq data for leaf tissue from eight individuals across the species geographic range (de Mattos et al., unpublished results). The samples were prepared using the Illumina Truseq stranded mRNA library prep. Libraries were then normalized, pooled equimolar, and sequenced on a Miseq nano cartridge PE 100 bp for each read, according to the Illumina dilute and

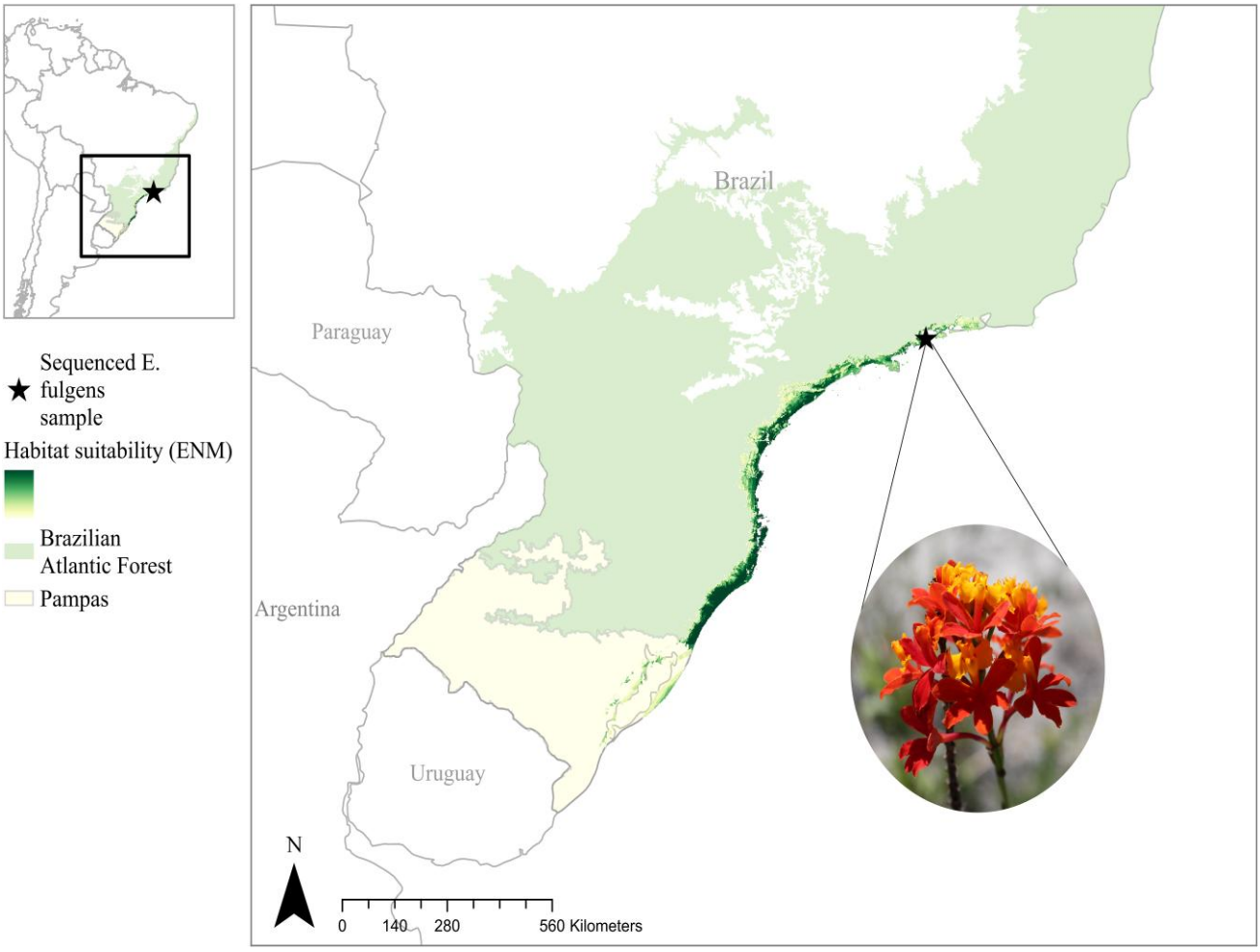

Fig. 5. Map showing the genomic sampling location (northern population; black star; –44.9582°, –23.3793°), along with the geographical distribution of *E. fulgens* (Orchidaceae) along the Brazilian coastal domain (Color scheme for biomes and habitat suitability indicated in the map legend).

denature protocol for loading. The pool was rebalanced based on the Miseq results, and the final pool was sequenced on an S4 200-cycle Novaseq cartridge PE 2 × 100, accessed via Wales Gene Park. The target read depth for the sample was 100 M reads. In addition, we also used RNA-seq data of flower tissue collected previously by our team and root tissue from four individuals from a previous study by Leal et al. (2020).

Structural and Functional Annotation

For the transposable element characterization, we used EDTA v.1.9.3 (Extensive de novo Transposable Element Annotator; Ou et al. 2019), with parameters "–sensitive 1 –anno 1 –evaluate 0 –threads 10 –species others –overwrite 0." The genome assembly was then masked with Repeatmasker (v.4.1.1; Tarailo-Graovac and Chen 2009), using the soft-masking option (-xsmall) and the gene families library generated by EDTA. RNA-seq data from the leaves, root, and flower were then mapped to the genome using STAR aligner v.2.6.1d (Dobin et al. 2013), and the gene prediction was carried out with BRAKER3 v.2.1.6 (Brůna et al. 2021) by using the soft-masked genome and the RNA-seq data to help with predictions. Functional annotation and protein function identification were performed using two different approaches: eggNOG-mapper v.2.1.6 (–target_taxa Viridiplantae; Cantalapiedra et al. 2021) to predict conserved protein sequences and domains (e.g. Gene Ontology terms) and also InterproScan v.5.63-95.0 (Jones et al. 2014) to look for protein families. Annotation quality was assessed by BUSCO (Simao et al. 2015) using the embryophyta_odb10 (v.2019-11-20) and liliopsida_odb10 (v.2019-11-20) datasets.

Historic Changes in Effective Population Size

The Pairwise Sequentially Markovian Coalescent (PSMC; v.0.6.5-r67) (Li and Durbin 2011) analysis was used to predict changes in the historic effective population size (Ne)

of *Epidendrum fulgens*. To perform the alignment used in the modeling analysis from PSMC, we used the PacBio HiFi reads and the assembled genome (collapsed primary assembly). The alignment was performed in bwa-mem v.0.7.17-r1188 (Li and Durbin 2009), and the consensus sequence and variant calling were performed using SAMtools v.0.1.19 (Li et al. 2009) and bcftools v.1.7 (Danecek and McCarthy 2017). To minimize the possibility of sequencing errors being introduced as variants, we filtered the VCF to retain only high-quality (Q = 999 in column 6 of the VCF format) variants. To run PSMC, the default settings were adopted, and the plot was generated using a generation time of 4 years (corresponding to the generation time for *E. fulgens* populations) and a mutation rate of 5×10^{-9} (corresponding to the mutation rate for Orchidaceae; Zhang et al. 2023b). Upper and lower bounds for the mutation rate within Orchidaceae were obtained from De La Torre 2017, which were used to explore possible alternative demographic histories. The data set was then bootstrapped 100 times to generate an expectation distribution around the coalescent curve.

Genome Evolution Analyses

For the genome evolution analyses, we downloaded the genomic and protein sequences of the other six angiosperm species, being representative of both Angiosperms/Monocots and the Orchiadaceae family. Many of the species with available genomes on NCBI do not provide an annotation, hence, our criterion here was to select the species with a high-quality genome assembly and available annotation. The selected species for the analyses were *Arabidopsis thaliana* (Brassicaceae), *Oryza sativa* (Poaceae), *Apostasia shenzhenica* (Orchidaceae), *Phalaenopsis equestris* (Orchidaceae), *Dendrobium nobile* (Orchidaceae), and *Vanilla planifolia* (Orchidaceae). We used a comparative genomics pipeline *compare_genomes* (Paril et al. 2023), which includes Orthofinder (Emms and Kelly 2019) and CAFE5 (Mendes et al. 2020) and utilizes genomics and phylogenetics tools to streamline the analysis and visualization of eukaryotic genome divergence, mainly identifying orthologous groups and inferring the expansion and contraction of gene families. Divergence times among species were estimated according to TimeTree.org (Kumar et al. 2017). Finally, we conducted a GO enrichment analysis using the TopGO v.3.22 package in R in order to understand which gene ontology (hereafter, GO) terms and gene functions were enriched in the gene families that were either expanded or contracted in *E. fulgens.*

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgements

Support for this work was provided by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; grant no. 2019/02150-3), FAEPEX (FUNCAMP), CBioClima (Processo FAPESP 2021/10639-5), and fellowships to FP and CPS (CNPq productivity grant no. 302849/2021-1; 305398/2019-9, respectively). JSM had a Doctorate fellowship from CNPq. We also had support from a CNPq fellowship for DMRP (Process 311558/2021-6) and a CNPq IC fellowship for FDR. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. We would also like to thank Dr. Marcela Uliano (Sanger Institute), Dr. Jonathan Wood (Sanger Institute) and Nadolina Brajuka (Vertebrates Genome Project) for helping with the genome assembly and curation, and Brian Smart (USDA) for helping with the circosplot analysis.

Conflict of Interest

There was no conflict of interest.

Data Availability

The raw reads for both PacBio Hifi and Omni-C can be found at NCBI, under BioProject PRJNA1045043. The genome assembly can be found under BioProject PRJNA1229662.

Literature Cited

Alexandersson R, Ågren J. Population size, pollinator visitation and fruit production in the deceptive orchid Calypso bulbosa. Oecologia. 1996:107:533–540. https://doi.org/10.1007/BF00333945.

Antonelli A, et al. Why plant diversity and distribution matter. New Phytol. 2023:240:1331–1336. https://doi.org/10.1111/nph.19282

Arida BL, et al. Reproductive barriers and fertility of two neotropical orchid species and their natural hybrid. Evol Ecol. 2021:35:41–64. https://doi.org/10.1007/s10682-020-10095-5

Arida BL, et al. The consequences of flower colour polymorphism on the reproductive success of a neotropical deceptive orchid. Plant Biol. 2025:27:515–527. https://doi.org/10.1111/plb.70020

Babu Y, Bayer M. Plant polygalacturonases involved in cell elongation and separation—the same but different? Plants. 2014:3:613–623. https://doi.org/10.3390/plants3040613

Bahieldin A, et al. Control of glycerol biosynthesis under high salt stress in Arabidopsis. Funct Plant Biol. 2013:41:87–95. https://doi.org/10.1071/FP13005

Bai W, et al. Demographically idiosyncratic responses to climate change and rapid Pleistocene diversification of the walnut genus Juglans (Juglandaceae) revealed by whole-genome sequences. New Phytol. 2017:217:1726–1736. https://doi.org/10.1111/nph.14917

Bevan M, Walsh S. The Arabidopsis genome: a foundation for plant research. Genome Res. 2005:15(12):1632–1642. https://doi.org/10.1101/gr.3723405.

Brown MR, et al. Tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. Bioinformatics. 2025:41:2. https://doi.org/10.1093/bioinformatics/btaf049

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP and AUGUSTUS supported by a protein database. NAR Genomic Bioinformat. 2021:3:lqaa108. https://doi.org/10.1093/nargab/lqaa108.

Cai J, et al. The genome sequence of the orchid phalaenopsis equestris. Nat Genet. 2015:47:65–72. https://doi.org/10.1038/ng.3149.

Cantalapiedra CP, Herņandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021:38:5825–5829. https://doi.org/10.1093/molbev/msab293.

Cardoso-Gustavson P, Saka MN, Pessoa EM, Palma-Silva C, Pinheiro F. 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 Biol. 2018:18:1–13. https://doi.org/10.1186/s12870-018-1398-y.

Chen B, et al. *Dendrobium thyrsiflorum* genome and its molecular insights into genes involved in important horticultural traits. Plant Biotechnol J. 2024:22:2887–2889. https://doi.org/10.1111/pbi.14411

Chu J. Jupiter plot: a circos-based tool to visualize genome assembly consistency (Version 1.0). Zenodo. 2018. http://doi.org/10.5281/zenodo.1241235.

Chun HJ, et al. Microtubule dynamics plays a vital role in plant adaptation and tolerance to salt stress. Int J Mol Sci. 2021:22:5957. https://doi.org/10.3390/ijms22115957

Corlett RT. Plant diversity in a changing world: status, trends, and conservation needs. Plant Divers. 2016:38:10–16. https://doi.org/10.1016/j.pld.2016.01.001

Cornwell WK, Pearse WD, Dalrymple RL, Zanne A. What we (don't) know about global plant diversity. Cold Spring Harbor Laboratory; 2018. https://doi.org/10.1101/404376 (Accessed June 2, 2025).

Cozzolino S, Widmer A. Orchid diversity: an evolutionary consequence of deception? Trends Ecol Evol. 2005:20:487–494. https://doi.org/10.1016/j.tree.2005.06.004.

Danecek P, Mccarthy SA. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics. 2017:33:2037–2039. https://doi.org/10.1093/bioinformatics/btx1.

de Assis FNM, et al. Karyology of the genus *Epidendrum* (Orchidaceae: Laeliinae) with emphasis on subgenus *Amphiglottium* and chromosome number variability in *Epidendrum secundum*. Bot J Linn Soc. 2013:172:329–344. https://doi.org/10.1111/boj.12045

De La Torre AR, Li Z, Van de Peer Y, Ingvarsson PK. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. Mol Biol Evol. 2017:34:1363–1377. https://doi.org/10.1093/molbev/msx069

de Lima TM, et al. Phenotypic plasticity rather than ecotypic differentiation explains the broad realized niche of a neotropical orchid species. Plant Biol. 2024:26:989–997. https://doi.org/10.1111/plb.13684

de Lima TM, da Silva SF, Ribeiro RV, Sánchez-Vilas J, Pinheiro F. Short-term salt spray reveals high salt tolerance in a neotropical orchid species. Theor Exp Plant Physiol. 2023:35:355–362. https://doi.org/10.1007/s40626-023-00291-3

de Mattos JS, et al. The relative role of climate and biotic interactions in shaping the range limits of a neotropical orchid. J Biogeogr. 2023:50:1315–1328. https://doi.org/10.1111/jbi.14629

Dichio B, et al. Changes in water status and osmolyte contents in leaves and roots of olive plants (olea europaea L.) subjected to water deficit. Trees. 2009:23:247–256. https://doi.org/10.1007/s00468-008-0272-1

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013:29:15–21. https://doi.org/10.1093/bioinformatics/bts63.

Dou L, He K, Higaki T, Wang X, Mao T. Ethylene signaling modulates cortical microtubule reassembly in response to salt stress. Plant Physiol. 2018:176:2071–2081. https://doi.org/10.1104/pp.17.01124

El Moukhtari A, Cabassa-Hourton C, Farissi M, Savouré A. How does proline treatment promote salt stress tolerance during crop plant development?. . Front Plant Sci. 2020:11:1127. https://doi.org/10.3389/fpls.2020.01127.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:20:238. https://doi.org/10.1186/s13059-019-1832-y

Fernández P, et al. A 160 Gbp fork fern genome shatters size record for eukaryotes. iScience. 2024:27:109889. https://doi.org/10.1016/j.isci.2024.109889

Gillman MP, Dodd ME. The variability of orchid population size. Bot J Linn Soc. 1998:126:65–74. https://doi.org/10.1111/j.1095-8339.1998.tb02516.x

Givnish TJ, et al. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. Am J Bot. 2018:105:1888–1910. https://doi.org/10.1002/ajb2.1178

Givnish TJ, et al. Orchid phylogenomics and multiple drivers of their extraordinary diversification. Proc Biol Sci. 2015:282:20151553. https://doi.org/10.1098/rspb.2015.1553.

Goodstein DM, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012:40:D1178–D1186. doi:10.1093/nar/gkr944.

Guan D, Mccarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020:36:2896–2898. https://doi.org/10.1093/bioinformatics/btaa.

Guizar Amador MF, et al. The *Gongora gibba* genome assembly provides new insights into the evolution of floral scent in male euglossine bee–pollinated orchids. G3 (Bethesda). 2024:14:jkae211. https://doi.org/10.1093/g3journal/jkae211

Gunde-Cimerman N, Ramos J, Plemenitas A. Halotolerant and halophilic fungi. Mycol Res. 2009:113:1231–1241. https://doi.org/10.1016/j.mycres.2009.09.002

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013:29:1072–1075. https://doi.org/10.1093/bioinformatics/btt0.

Hadfield KA, Bennett AB. Polygalacturonases: many genes in search of a function. Plant Physiol. 1998:117:337–343. https://doi.org/10.1104/pp.117.2.337

Hayat S, et al. Role of proline under changing environments: a review. Plant Signal Behav. 2012:7:1456–1466. https://doi.org/10.4161/psb.21949

He P, et al. Genome-wide identification and expression analysis of the polygalacturonase gene family in sweetpotato. BMC Plant

Biol. 2023:23:300. https://doi.org/10.1186/s12870-023-04272-1.
Hibrand Saint-Oyant L, et al. A high-quality sequence of Rosa chinensis to elucidate genome structure and ornamental traits. Cold Spring Harbor Laboratory; 2018. https://doi.org/10.1101/254102.
Hotaling S, et al. Long reads are revolutionizing 20 years of insect genome sequencing. Genome Biol Evol. 2021:13:evab138. doi: 10.1093/gbe/evab138 .
Howe K, et al. Significantly improving the quality of genome assemblies through curation. Cold Spring Harbor Laboratory; 2020. https://doi.org/10.1101/2020.08.12.247734.
Hundsdoerfer AK, et al. High-quality haploid genomes corroborate 29 chromosomes and highly conserved synteny of genes inHyleshawkmoths (Lepidoptera: Sphingidae). Cold Spring Harbor Laboratory; 2022. https://doi.org/10.1101/2022.04.08.487644.
Jiang K, et al. Genomic, transcriptomic, and metabolomic analyses reveal convergent evolution of oxime biosynthesis in Darwin's orchid. Mol Plant. 2025:18:392–415. https://doi.org/10.1016/j.molp.2024.12.010
Jones P, et al. InterProScan 5: genome-scale protein function classification. Bioinforma Oxf Engl. 2014:30:1236–1240. https://doi.org/10.1093/bioinformatics/btu0.
Kareiva P, Kareiva I. Biodiversity hotspots and conservation priorities. In: Oxford Research encyclopedia of environmental science. Oxford University Press; 2017. https://doi.org/10.1093/acrefore/9780199389414.013.95 (Accessed July 25, 2025).
Karremans AP. With great biodiversity Comes great responsibility: the underestimated diversity of *Epidendrum* (Orchidaceae). Harv Pap Bot. 2021:26:299–369. https://doi.org/10.3100/hpib.v26iss2.2021.n1
Kong W, Wang Y, Zhang S, Yu J, Zhang X. Recent advances in assembly of Complex plant genomes. Genom Proteom Bioinform. 2023:21:427–439. https://doi.org/10.1016/j.gpb.2023.04.004
Kress WJ, et al. Green plant genomes: what we know in an era of rapidly expanding opportunities. Proc Natl Acad Sci U S A. 2022:119:e2115640118. https://doi.org/10.1073/pnas.2115640118
Kuhn LA, Zonneveld KA, Souza PA, Cancelli RR. Late quaternary palaeoenvironmental evolution and sea level oscillation of Santa Catarina Island (southern Brazil). Biogeosciences. 2023:20:1843–1861. https://doi.org/10.5194/bg-20-1843-2023.
Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017:34:1812–1819. https://doi.org/10.1093/molbev/msx116
Kyriakidou M, Tai HH, Anglin NL, Ellis D, Strömvik MV. Current strategies of polyploid plant genome sequence assembly. Front Plant Sci. 2018:9:1660. https://doi.org/10.3389/fpls.2018.01660
Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. F1000Research. 2017:6:1287. https://doi.org/10.12688/f1000research.12.
Leal BSS, Ambrosano GB, Margarido GRA, Palma-Silva C, Pinheiro F. Interploidy gene flow does not prevent adaptive genetic differentiation in sympatric populations of Epidendrum fulgens and E. Puniceoluteum (Orchidaceae). Bot J Linn Soc. 2025;209:237–249. https://doi.org/10.1093/botlinnean/boaf015
Leal BSS, Brandão MM, Palma-Silva C, Pinheiro F. Differential gene expression reveals mechanisms related to habitat divergence between hybridizing orchids from the Neotropical coastal plains. BMC Plant Biol. 2020:20:554. https://doi.org/10.1186/s12870-020-02757-x
Leitch IJ, et al. Genome size diversity in orchids: consequences and evolution. Ann Bot. 2009:104:469–481. https://doi.org/10.1093/aob/mcp003
Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009:25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011:475:493–496. https://doi.org/10.1038/nature10231.
Li Y, et al. Pangeneric genome analyses reveal the evolution and diversity of the orchid genus *Dendrobium*. Nat Plants. 2025:11:421–437. https://doi.org/10.1038/s41477-024-01902-w
Loiseau O, et al. Genome skimming reveals widespread hybridization in a neotropical flowering plant radiation. Front Ecol Evol. 2021:9:668281. https://doi.org/10.3389/fevo.2021.668281.
Lu H, Liu Z, Lan S. Genome sequencing reveals the role of MADS-box gene families in the floral morphology evolution of orchids. Plant J. 2019:5:247–254. https://doi.org/10.1016/j.hpj.2019.11.005
Marks RA, Hotaling S, Frandsen PB, Vanburen R. Representation and participation across 20 years of plant genome sequencing. Nat Plants. 2021:7:1571–1578. https://doi.org/10.1038/s41477-021-01031-8.
Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2020:36:5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022
Moraes AP, Chinaglia M, Palma-Silva C, Pinheiro F. Interploidy hybridization in sympatric zones: the formation of Epidendrum fulgens x E. Puniceoluteum hybrids (epidendroideae, Orchidaceae). Ecol Evol. 2013:3:3824–3837. https://doi.org/10.1002/ece3.752
Morrell PL, Buckler ES, Ross-Ibarra J. Crop genomics: advances and applications. Nat Rev Genet. 2012:13(2):85–96. https://doi.org/10.1038/nrg3097.
Myburg A, et al. The Eucalyptus grandisGenome Project: Genome and transcriptome resources for comparative analysis of woody plant biology. BMC Proc. 2011:5(Suppl 7. https://doi.org/10.1186/1753-6561-5-S7-I20.
Nystedt B, et al. The Norway spruce genome sequence and conifer genome evolution. Nature. 2013:497:579–584. https://doi.org/10.1038/nature12211.
Ou S, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019:20:275. https://doi.org/10.1186/s13059-019-1905-y.
Paril J, Zare T, Fournier-Level A. Compare_genomes: a comparative genomics workflow to streamline the analysis of evolutionary divergence across genomes. Cold Spring Harbor Laboratory; 2023. https://doi.org/10.1101/2023.03.16.533049 (Accessed April 16, 2024).
Perez-Escobar O, et al. The origin and speciation of orchids. New Phytol. 2024a:242:700–716. https://doi.org/10.1111/nph.19580
Pérez-Escobar OA, et al. Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. Am J Bot. 2024b;108:1166–1180. https://doi.org/10.1002/ajb2.1702
Pessoa EM, et al. Integrating different tools to disentangle species complexes: a case study in *Epidendrum* (Orchidaceae). Taxon. 2012:61:721–734. https://doi.org/10.1002/tax.614002
Piet Q, et al. A chromosome-level, haplotype-phased vanilla planifolia genome highlights the challenge of partial

endoreplication for accurate whole-genome assembly. Plant Commun. 2022:3:100330. https://doi.org/10.1016/j.xplc.2022.100330
Pinheiro F, et al. Strong but permeable barriers to gene exchange between sister species of Epidendrum. Am J Bot. 2016:103:1472–1482. https://doi.org/10.3732/ajb.1600064
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, et al. Phylogeography and genetic differentiation along the distributional range of the orchid Epidendrum fulgens: a Neotropical coastal species not restricted to glacial refugia. J Biogeogr. 2011:38:1923–1935. https://doi.org/10.1111/j.1365-2699.2011.02539.x.
Pont C, et al. Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. Genome Biol. 2019:20:29. https://doi.org/10.1186/s13059-019-1627-1.
Porubsky D, et al. Fully phased human genome assembly without parental data using single-cell strand sequencing and long reads. Nat Biotechnol. 2020:39:302–308. https://doi.org/10.1038/s41587-020-0719-5
Rhie A, et al. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020:21:245. https://doi.org/10.1186/s13059-020-02134-9
Rhie A, et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 2021:592:737–746. https://doi.org/10.1038/s41586-021-03451-0
Russo A, et al. Genome of the early spider-orchid ophrys sphegodes provides insights into sexual deception and pollinator adaptation. Nat Commun. 2024:15:6308. https://doi.org/10.1038/s41467-024-50622-4
Sahu SK, et al. Chromosome-scale genomes of commercial timber trees (Ochroma pyramidale, Mesua ferrea, and Tectona grandis). Sci Data. 2023:10:512. https://doi.org/10.1038/s41597-023-02420-8
Schwacke R, Bolger ME, Usadel B. PubPlant-a continuously updated online resource for sequenced and published plant genomes. Front Plant Sci. 2025:16:1603547.
Serna-Sánchez MA, et al. Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. Sci Reports. 2021:11:6858. https://doi.org/10.1038/s41598-021-83664-5
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31:3210–3212. https://doi.org/10.1093/bioinformatics/btv3.
Song A, et al. Analyses of a chromosome-scale genome assembly reveal the origin and evolution of cultivated chrysanthemum. Nat Commun. 2023:14:2021. https://doi.org/10.1038/s41467-023-37730-3
Sujii PS, Cozzolino S, Pinheiro F. Hybridization and geographic distribution shapes the spatial genetic structure of two co-occurring orchid species. Heredity. 2019:123:458–469. https://doi.org/10.1038/s41437-019-0254-7.
Sun Y, Shang L, Zhu Q-H, Fan L, Guo L. Twenty years of plant genome sequencing: achievements and challenges. Trends Plant Sci. 2022:27:391–401. https://doi.org/10.1016/j.tplants.2021.10.006
Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Current Protocols Bioinformatics. 2009:25:4.10.11–14.10.14. https://doi.org/10.1002/0471250953.bi041.
Tsai W-C, et al. Post genomics era for orchid research. Bot Stud. 2017:58:61. https://doi.org/10.1186/s40529-017-0213-7
Wang Y, et al. Progress in systematics and biogeography of Orchidaceae. Plant Divers. 2024:46:425–434. https://doi.org/10.1016/j.pld.2024.05.002
Xu Q, et al. Chromosome-scale assembly of the *Dendrobium nobile* genome provides insights into the molecular mechanism of the biosynthesis of the medicinal active ingredient of *Dendrobium*. Front Genet. 2022:13:844622. https://doi.org/10.3389/fgene.2022.844622
Yamaguchi K, et al. Technical considerations in hi-C scaffolding and evaluation of chromosome-scale genome assemblies. Authorea, Inc; 2021. https://doi.org/10.22541/au.162183528.83085003/v1 (Accessed June 2, 2025).
Yan L, et al. The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. Mol Plant. 2015:8:922–934. https://doi.org/10.1016/j.molp.2014.12.011
Zhang D, Li W, Xia EH, Zhang QJ, Liu Y, Zhang Y. The medicinal herb Panax notoginseng genome provides insights into ginsenoside biosynthesis and genome evolution. Molec Plant. 2017:10:903–907. https://doi.org/10.1016/j.molp.2017.02.011.
Zhang G et al. Comprehensive phylogenetic analyses of Orchidaceae using nuclear genes and evolutionary insights into epiphytism. J Integr Plant Biol. 2023a:65:1204–1225. https://doi.org/10.1111/jipb.13462
Zhang W, et al. Genome sequence of *Apostasia ramifera* provides insights into the adaptive evolution in orchids. BMC Genomics. 2021:22:536. https://doi.org/10.1186/s12864-021-07852-3
Zhang Z, et al. Exploring island syndromes: variable matrix permeability in *Phalaenopsis pulcherrima* (Orchidaceae), a specialist lithophyte of tropical Asian inselbergs. Front Plant Sci. 2023b:14:1097113. https://doi.org/10.3389/fpls.2023.1097113
Zhao Z, et al. Characterization and comparative analysis of the complete plastomes of five *Epidendrum* (Epidendreae, Orchidaceae) Species. Int J Mol Sci. 2023b:24:14437. https://doi.org/10.3390/ijms241914437
Zhou C, Mccarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023:39:btac808. https://doi.org/10.1093/bioinformatics/btac.

Associate editor: Chuan Ku