Estimation of divergence time and comparative plastid genomics of Orchis species (Orchidaceae)

Botanical Journal of the Linnean Society, 2024, **XX**, 1–10
https://doi.org/10.1093/botlinnean/boae050
Advance access publication 29 August 2024
Original Article

Original Article

Estimation of divergence time and comparative plastid genomics of *Orchis* species (Orchidaceae)

Jacopo Calevo[1,2,3,*], Juan Viruel[2], Martino Adamo[1], Leif Bersweden[2,4], Roberta Gargiulo[2], Robyn S. Cowan[2], Michael F. Fay[2,5]

[1]Department of Life Sciences and Systems Biology, University of Turin, Turin 10125, Italy
[2]Royal Botanic Gardens, Kew, Richmond TW9 3DS, United Kingdom
[3]School of Molecular and Life Sciences, Curtin University, Bentley, Western Australia 6102, Australia
[4]Queen Mary University of London, London E1 4NS, United Kingdom
[5]University of Western Australia, Crawley, Western Australia 6009, Australia

[*]Corresponding author. Jodrell Laboratory, Royal Botanic Gardens, Kew , Richmond TW9 3DS, United Kingdom. E-mail: j.calevo@kew.org

ABSTRACT

Low-coverage sequencing in plants allows whole plastomes to be obtained that can be used to investigate phylogenetic relationships among groups. The genus *Orchis* (*c.* 20 species), is usually divided into *Orchis* subgenera *Orchis* and *Masculae*. These subgenera are composed of three (*Anthropophorae*, *Italicae*, and *Orchis*) and four (*Masculae*, *Provinciales*, *Pusillae*, and *Robustocalcare*) sections, respectively. In this study, we used genome-skimming data to assemble the plastid genomes of 11 species (15 accessions) of *Orchis*, representing six out of the seven sections, from which we constructed a dated phylogenetic tree. Results suggest that the divergence between the subgenera occurred *c.* 10.53 Mya, whereas the main separation of the sections is dated between 6.53 and 3.48 Mya. Furthermore, we found 206 (in *O. anthropophora*) to 230 (in *O. provincialis*) microsatellite regions in the assembled plastomes, which could be used to design specific primers for further population genetics and phylogenetic studies and, ultimately, inform conservation efforts. The plastome data here presented represent a new contribution to the molecular systematics of the genus, and they can be used to further explore infrageneric and infrascpecific molecular variation in *Orchis*.

Keywords: divergence time; orchids; phylogeny; plastomes

INTRODUCTION

During the first event of endosymbiosis between cyanobacteria and eukaryotes, believed to have occurred *c.* 1.5 billion years ago (Chan *et al.* 2011), bacterial genes were transferred to the host cell nucleus or lost (Keeling *et al.* 2004), resulting in the establishment of the primary plastid. Now the plastid genome (plastome), which is almost always circular, has a size of 100–200 kbp (120–160 kbp in land plants) and harbours 100–120 highly conserved genes with their own gene expression system (Olejniczak *et al.* 2016).

Hight-throughput sequencing technologies provide the opportunity to obtain plastome data as a byproduct of low-coverage sequencing of plant genomic DNA, not only allowing the study of biogeography and plastid evolution across groups (e.g. Yuan *et al.* 2018, Frankel *et al.* 2022), but also the investigation of phylogenetic relationships among taxa (Bedoya *et al.* 2019, Serna-Sánchez *et al.* 2021, Lee *et al.* 2022).

The genus *Orchis* Tourn. ex L., currently comprising *c.* 20 recognized species (POWO 2023), has been (and still is) a source of debate among taxonomists. *Orchis* spp. are spread among subtropical, temperate, and arctic-alpine climates, ranging from Macaronesia to Mongolia, across Europe, temperate Asia, North Africa, and the Middle East, with a centre of diversification in the Mediterranean Basin (POWO 2023). They form > 40 known hybrids, revealing low reproductive isolation.

Species are often divided into two subgenera (Kretzschmar *et al.* 2007): *Orchis* subgenus *Orchis* (which includes the 'anthropomorphic' species) and *Masculae* H.Kretzschmar, Eccarius & H.Dietr. (including the 'non-anthropomorphic' species). Together, these subgenera comprise seven sections: three in subgenus *Orchis* (*Anthropophorae* H.Kretzschmar, Eccarius & H.Dietr., *Italicae* H.Kretzschmar, Eccarius & H.Dietr., and *Orchis*) and four in subgenus *Masculae* (*Masculae* Lindl., *Provinciales* Parl., *Pusillae* Parl., and *Robustocalcare* Hautz.) (Kretzschmar *et al.* 2007). However, after the segregation of

Received 25 April 2023; revised 2 July 2024; accepted 3 July 2024

several genera since the original description (e.g. *Anacamptis* Rich., *Dactylorhiza* Neck. ex Nevski, *Gymnadenia* R.Br., and *Neotinea* Rchb.f.), the genus has been the target of taxonomic revision by several authors, with some treating subgenus *Masculae* as a separate genus *Androrchis* Tyteca & Klein (e.g. Foelsche and Jakely 2009, Gamarra *et al.* 2012, Tyteca *et al.* 2012). However, this separation has not been accepted by most orchidologists (as summarized by Bateman 2009, 2012; Scopece *et al.* 2010). In addition, the number of species varies depending on authors, with up to 36 species described in Delforge (2016). In this study we adopted the taxonomy proposed by Kretzchmar *et al.* (2007) with some modifications reported in POWO (2023), and resulting from more recent studies, particularly in section *Robustocalcare* (e.g. Calevo *et al.* 2021, 2023).We used genome-skimming data to assemble the plastid genomes of 11 species (15 accessions) representing six out of seven sections, with the aim of analysing our data in a comparative framework to investigate the evolution of the plastid genomes and divergence times of the species in the genus; also, we compared plastid phylogenetics with phylogenetic trees based on the internal transcribed spacer (ITS) of the ribosomal DNA, and described the abundance of plastid microsatellites (SSRs), which might be used as a resource by researchers interested in the molecular investigation of the species of *Orchis*.

MATERIAL AND METHODS

Plant material, DNA extraction, and sequencing

Samples of *Orchis patens* Desf. subsp. *brevicornis* (Viv.) Asch. & Graeb. (Calevo *et al.* 2023) and *O. provincialis* Balb. ex Lam. & DC. were collected in north-western Italy, and the sample of *O. canariensis* Lindl. was collected in the Canary Islands, Spain. Samples of *O. anthropophora* All., *O. mascula* (L.) L., *O. militaris* L., *O. simia* Lam., *O. purpurea* Huds., *O. anatolica* Boiss., *O. spitzelii* Saut. ex W.D.J.Koch subsp. *nitidifolia* (W.P.Teschner) Soó, and *O. quadripunctata* Cirillo ex Ten. were obtained from the Kew DNA & Tissue Bank or the Kew Herbarium. Information on collection localities and voucher specimens is provided in Table S1 in the Supporting Information. All species included are representative of the centre of speciation of the genus (Eckert *et al.* 2008) ranging from Macaronesia to the Middle East.

Total genomic DNA was extracted from silica-dried leaf tissue or herbarium specimens using a modified CTAB protocol and purified by isopropanol precipitation or silica columns (Epoch Life Science, Missouri City, TX, USA) from the aqueous supernatant after chloroform/isoamyl alcohol purification (Neubig *et al.* 2014). DNA was checked on a 1% agarose gel to assess DNA quality. DNA concentrations were measured with a Qubit fluorometer using the dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). Genomic libraries were prepared using NEBNext® UltraTM II DNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA) with AMPure XP magnetic beads for purification and size selection (300–350 bp). NEBNext® Multiplex Oligos for Illumina® (Dual Index Primer Sets I and II) was used to add barcodes for multiplexed sequencing (Viruel *et al.* 2019). Library quality was evaluated using a QuantusTM fluorometer (Promega Corp., Madison, WI, USA) and an Agilent 4200 TapeStation (Agilent Technologies, Santa Clara, CA, USA). Multiplexed libraries were then sequenced at RBG Kew on an Illumina MiSeq (Illumina, San Diego, CA, USA) lane. Read quality was checked by FastQC v.0.11.7 (Andrews 2010), and Trimmomatic v.0.35 (Bolger, Lohse and Usadel 2014) was used with the default parameters to remove low-quality reads and adapter sequences, discarding sequences with an average Phred33 score < 20.

ITS detection and alignment

All the reads were mapped-to-reference ITS sequences (*O. mascula* GenBank accession number AY351379 for subgenus *Masculae*, and *O. simia* GenBank accession number KU697369 for subgenus *Orchis*) in Geneious v.8.1.9. and CAP3 (Huang and Madan 1999) to extract the ITS of our 15 accessions from the trimmed read libraries. Sequences were then aligned in MAFFT v.7031b (Katoh and Standley 2013) with the default parameters (specifically the default strategy L-INS-i, which performs iterative refinement incorporating local pairwise alignment information; the default gap open penalty is 1.53 and the default gap extension penalty is 0.0). Phylogenetic analysis (see below) was performed in CIPRES Science Gateway (Miller *et al.* 2010). To double check the presence/absence of ITS sequence variation in introgressed (or putatively introgressed) samples, ITS mapping was performed using bbMap from the bbTools suite (BBMap—Bushnell B.: sourceforge.net/projects/bbmap/); mapped reads were assemble using SPADEs (Bankevich *et al.* 2012) and the obtained scaffolds were merged, when possible, using CAP3 (Huang and Madan 1999).

Plastid DNA assembly

A reference-guided assembly was performed on all the reads that were mapped to the plastid genome of *Habenaria radiata* (Thunb.) Spreng. [= *Pecteilis radiata* (Thunb.) Raf.] (GenBank Accession number KX1237.1), the closest relative of *Orchis* for which a fully annotated plastome was available at the time this step was executed, using Geneious v.8.1.9.

Mapped-to-reference plastomes were then used as references for a *de novo* assembly of the original reads in Geneious v.8.1.9. with default Medium-Low Sensitivity/Fast option. Gene annotation was exhaustively performed for all 15 accessions with GeSeq web version; annotations were checked, and tRNAs were further checked with tRNAscan-SE v.2.0 as implemented in GeSeq (Tillich *et al.* 2017). The diagrams for the circular genomes were obtained with the program OrganellarGenomeDRAW web version (Lohse *et al.* 2013, Greiner *et al.* 2019).

Microsatellite detection

We used Imperfect Microsatellite Extractor (Mudunuri and Nagarajaram 2007), an online server tool for microsatellite (SSR) detection from genomic sequences. Minimum thresholds for the search were set at eight for mononucleotide repeats, five for dinucleotide repeats and four for tri-, tetra-, penta-, and hexanucleotide repeats.

Phylogenetic analyses

ITS sequences retrieved from raw reads were aligned with *Orchis* sequences available from GenBank and used to build an ITS-based maximum likelihood (ML) phylogenetic tree. Sequences were first aligned using MAFFT v.7031b (Katoh and Standley 2013) with default settings, and then trimmed with TrimAl

Table 1. Information of NGS sequencing and SSRs detection from 15 *Orchis* accessions.

	Species ID	SSRs	Input read pairs	Both surviving	Forward only	Reverse only	Dropped
1	*O_anthropophora*	204	2 741 985	2 020 488 (73.69%)	709 300 (25.87%)	4550 (0.17%)	7647 (0.28%)
2	*O_militaris1*	222	2 375 559	1 525 332 (64.21%)	839 568 (35.34%)	3764 (0.16%)	6895 (0.29%)
3	*O_militaris2*	219	1 222 508	718 594 (58.78%)	497 994 (40.74%)	1738 (0.14%)	4182 (0.34%)
4	*O_simia1*	215	1 080 436	409 061 (37.86%)	667 104 (61.74%)	1046 (0.10%)	3225 (0.30%)
5	*O_simia3*	216	2 514 567	1 398 427 (55.61%)	1 104 923 (43.94%)	3408 (0.14%)	7809 (0.31%)
6	*O_purpurea1_1*	215	2 041 466	781 922 (38.30%)	1 247 124 (61.09%)	2081 (0.10%)	10 339 (0.51%)
7	*O_purpurea1_2*	220	1 731 400	908 366 (52.46%)	815 060 (47.08%)	2323 (0.13%)	5651 (0.33%)
8	*O_purpurea2_2*	210	2 933 739	2 140 468 (72.96%)	781 557 (26.64%)	4479 (0.15%)	7235 (0.25%)
9	*O_mascula*	207	1 532 285	1 037 742 (67.73%)	487 143 (31.79%)	3078 (0.20%)	4322 (0.28%)
10	*O_pauciflora*	210	2 146 960	797 857 (37.16%)	173 740 (8.09%)	2108 (0.13%)	1 173 255 (54.65%)
11	*O_canariensis*	214	1 682 867	1 007 618 (59.88%)	667 438 (39.66%)	2618 (0.16%)	5193 (0.31%)
12	*O_patens* subsp. *brevicornis*	226	2 315 342	1 477 939 (63.83%)	827 750 (35.75%)	3474 (0.15%)	6179 (0.27%)
13	*O_prisca* (*O. spitzelii* subsp. *nitidifolia*)	227	1 820 619	789 822 (43.38%)	222 441 (12.22%)	1576 (0.09%)	806 780 (44.31%)
14	*O_quadripunctata*	225	1 263 760	443 067 (35.06%)	85 388 (6.76%)	2041 (0.16%)	733 264 (58.02%)
15	*O_provincialis*	230	1 675 330	660 673 (39.44%)	1 008 145 (60.18%)	2025 (0.12%)	4487 (0.27%)

v.1.2.59 (Capella-Gutierrez *et al.* 2009) using automated selection on 'gappyout' mode. ML estimation was performed on the final dataset with randomized accelerated maximum likelihood (RAxML) v.8.2.12 (Stamatakis 2014) through 1000 bootstrap replicates using the GTR + GAMMA algorithm. Alignment, trimming and ML estimation were performed on CIPRES Science Gateway (Miller *et al.* 2010).

Plastomes from our 15 *Orchis* samples were aligned by using MAFFT v.7031b (Katoh and Standley 2013). The dataset was then cleaned by removing poorly aligned regions with TrimAl v.1.3 (Capella-Gutierrez *et al.* 2009) on the webserver Phylemon v.2.0 (Sánchez *et al.* 2011) using the option automated1, that uses a heuristic selection of the automatic method based on similarity statistics. Poorly aligned positions and divergent regions were trimmed using GBlock v.0.9.1b (Talavera and Castresana 2007). The final aligned dataset was used to fit the substitution model using JModelTest2 (Darriba *et al.* 2012), which was applied to build a ML phylogenetic tree with RAxML v.8.2.12 (Stamatakis 2014), and 5000 bootstrap replications were performed.

Estimating divergence time

We estimated divergence times among species based on Bayesian inference in BEAST2 v.2.6.6 (Bouckaert *et al.* 2019). *Habenaria radiata* (GenBank accession NC_035834), *Ophrys fusca* Link subsp. *iricolor* (Desf.) K.Richt. (GenBank accession AP018716), and *Ophrys sphegodes* Mill. (GenBank accession AP018717) plastid genomes were used as outgroups.

A TimeTree was inferred by applying the BEAST2 method on plastid sequences. Absolute age estimation analysis relied on one calibration point, a TIM + I + G substitution model, an optimized relaxed molecular clock, and a calibrated Yule speciation model. The best substitution model was selected with the BEAUti package cModelTest v.1.3.3, and the best molecular clock model was selected using a nested sampling approach (Russel *et al.* 2018) and based on marginal likelihood, using the NS package v.1.2.1, and following the suggestions of the BEAST2 community (Barido-Sottani *et al.* 2018), see Supporting Information, Table S2 for details. Priors were estimated where not available as recommended by Drummond and Bouckaert (2015). We imposed a calibration point at the *Ophrys-Orchis* MRCA node at 16.8 Mya (sigma = 3.68 Mya). Calibrations were extracted from the TimeTree database (Hedges *et al.* 2006, Hedges *et al.* 2015), using three studies as reference (Inda *et al.* 2012, Sramkó *et al.* 2014; Zhang *et al.* 2022). We conducted an MCMC analysis with 100 million generations, sampling each 1000th tree generated. Convergence of the chains and effective sample size (ESS) values were evaluated in TRACER v.1.6.0 (Rambaut *et al.* 2014), and all parameters showed ESS values > 200. A maximum clade credibility tree was generated with TreeAnnotator v.1.8, burn-in 10% of the initial states and edited using FigTree v.1.4.4.

RESULTS

Illumina paired-end sequencing produced from 1 080 436 (*O. simia1*) to 2 933 739 (*O. purpurea2_2*) raw reads. Cleaned and high-quality reads (Table 1) were used to assemble and annotate the 15 plastomes (e.g. Plastome scheme Fig. 1). Plastome sizes were comparable among all the examined taxa, ranging between 150 041 for *O. spitzelii* subsp. *nitidifolia* to 154 958 for *O. anthropophora*. IMEx analysis for the detection of microsatellites from our 15 new plastid sequences identified 204 (*O. anthropophora*) to 230 (*O. provincialis*) SSRs (Table 1).

The plastid phylogenetic tree topology (Fig. 2) and the ITS-based phylogenetic tree topology (Fig. 3), divided 'anthropomorphic' from 'non-anthropomorphic' species, namely subgenera *Orchis* and *Masculae*, with branches being supported by high bootstrap values. *Orchis canariensis* clustered, as expected, with *O. patens*. Some anthropomorphic accessions of the same species (*O. purpurea* and *O. simia*) did not form clades (some accessions of *O. purpurea* clustering closer to *O. simia* and *O. militaris*).

However, in both the ITS tree (Fig. 3) and the plastid DNA tree (Fig. 2), the specimen labelled as *O. anatolica* clustered with *O. pauciflora* Ten. in section *Masculae* (to which *O. pauciflora*

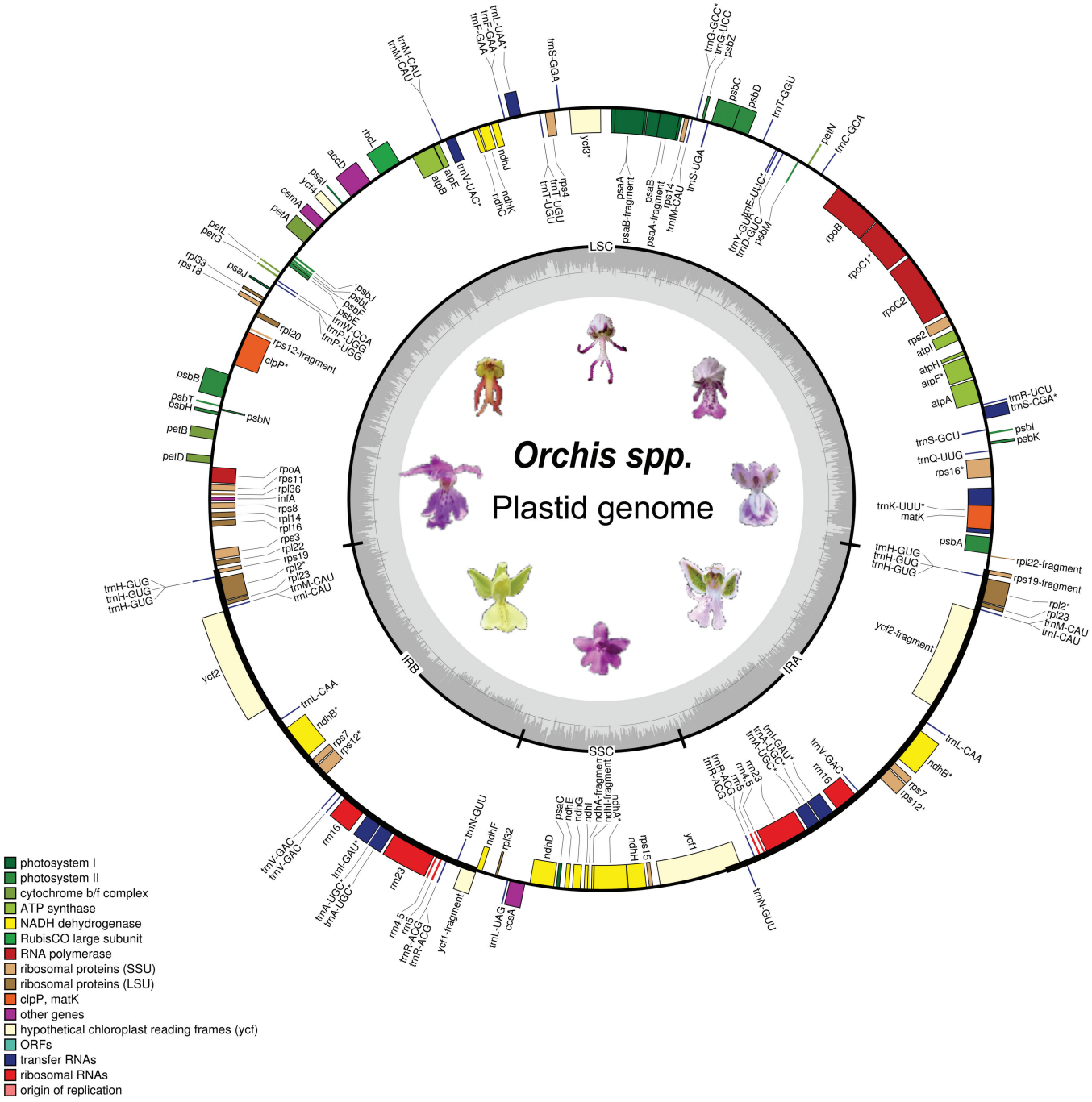

Figure 1. Gene map of the plastome of *Orchis* spp. Genes are indicated by boxes on the inside (clockwise transcription) and outside (counterclockwise transcription) of the outermost circle. The inner circle identifies the major structural components of the plastome (LSC, IR, and SSC). Genes belonging to different functional groups are colour coded. The dashed area in the inner circle indicates the GC content of the plastome.

belongs), and visual inspection of the herbarium voucher confirmed this identification. Therefore, it will be treated as *O. pauciflora* from hereon. Phylogenetic trees based on plastid DNA place *O. quadripunctata* (which belongs to section *Pusillae*) between accessions belonging to section *Robustocalcare* (Figs 1, 2).

BEAST2 analysis revealed that subgenera *Orchis* and *Masculae* diverged *c.* 10.53 Mya. The divergence between the anthropomorphic sections *Anthropophorae* (*O. anthropophora*) and *Orchis* (*O. militaris*, *O. purpurea*, and *O. simia*) occurred 5.88 Mya, and the divergence between non-anthropomorphic sections *Provinciales* (*O. provincialis*) and *Masculae* (*O. mascula*) and between *Masculae* and *Robustocalcare* (*O. canariensis*, *O. patens*, and *O. spitzelii* subsp. *nitidifolia*) occurred 6.30 and 6.53 Mya, respectively (Fig. 4).

The mVISTA-based identity plot (Fig. S1 in the Supporting Information) revealed DNA sequence and gene synteny conservation across the 15 plastomes. Generally, the number, order and orientation of genes were relatively conserved, but distinct

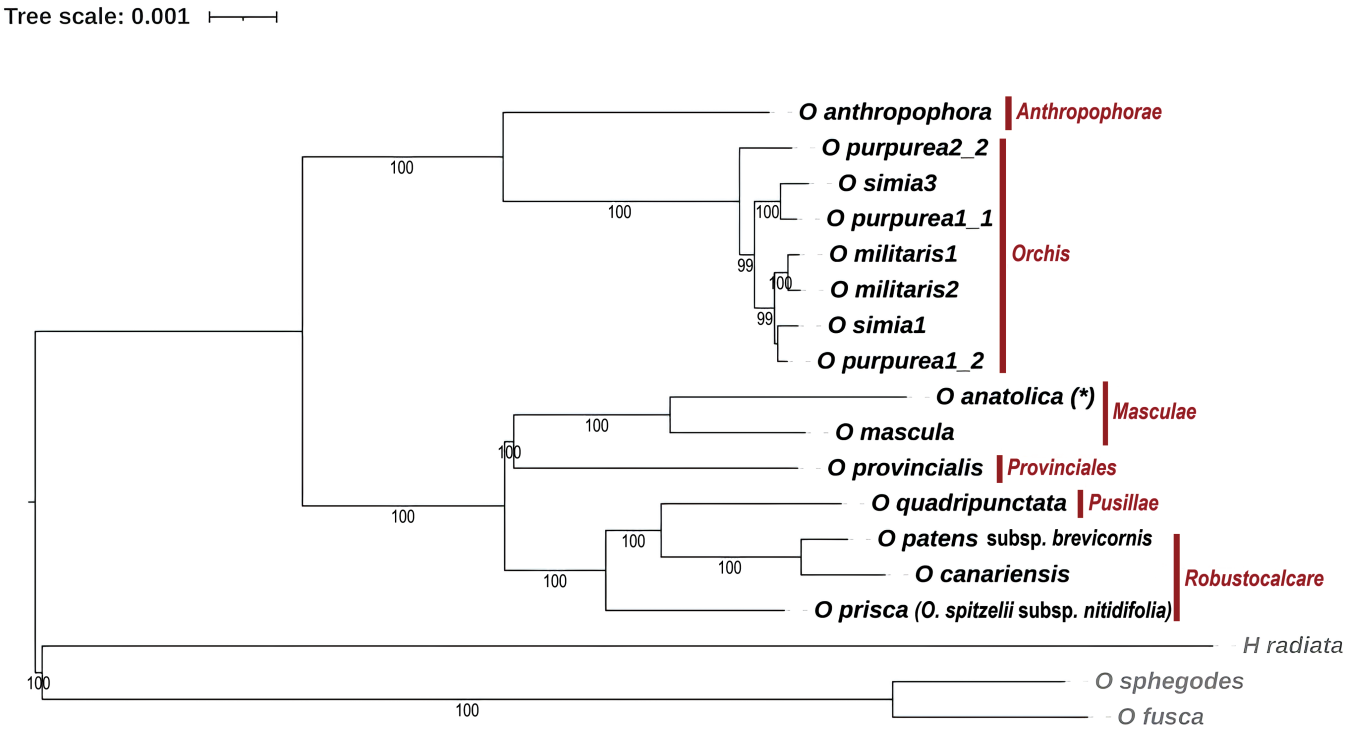

Figure 2. Phylogenetic tree based on total aligned and trimmed plastid sequences. The RAxML tree was built in CIPRES Science Gateway. The sample of *Orchis anatolica* (*) collected from Kew Herbarium clustered with *O. mascula* in section *Masculae* instead of *Pusillae*. *Habenaria radiata*, *Ophrys fusca* subsp. *iricolor*, and *O. sphegodes* were used as the outgroups.

sequence variations were recorded in several regions, e.g. *accD*, *rps3*, *rps3-rps2-2*, *rp4-ndhK*, *rps11-rps8*, *ycf2*, and *ycf2-2*.

DISCUSSION

Strongly supported phylogenetic trees are crucial for understanding evolution, phylogenetic classification, conservation, and drivers of clade diversification (Li *et al.* 2019). Advances in next-generation sequencing, providing opportunities for inferring phylogenies using numerous loci, decrease potential incongruences resulting from stochastic error (Salamin *et al.* 2005, Philippe *et al.* 2011).

A recent study by Li *et al.* (2019) indicated that, after the early rapid divergence of the first three small subfamilies in Orchidaceae between 90 and 81 Mya, there might have been a period of stasis. Then another 14 Myr passed before the divergence of the two largest subfamilies, Orchidoideae (to which *Orchis* belongs) and Epidendroideae.

In our study, we assembled and used plastid genomes for 15 samples representing 11 *Orchis* spp. in a comparative phylogenomic analysis to estimate divergence times among taxa. Plastome genome sizes for Orchidaceae published to date range from 14 015 bp in *Pogoniopsis schenckii* Cogn. (Klimpert *et al.* 2022) to 178 131 bp in *Cypripedium formosanum* Hayata (Kim *et al.* 2020), a discrepancy explained by the different life forms; heterotrophic plants like *P. schenkii* usually having lost at least some of the photosynthesis-related genes, resulting in smaller plastid genomes (Barrett *et al.* 2019). The plastome sizes of the species we studied were all > 150 000 bp, which is congruent with the size of plastomes in other photosynthetic orchids (Kim *et al.* 2020).

Phylogenetic results were consistent with phylogenetic trees published previously (e.g. Bateman *et al.* 2003) but show some ambiguities. In the ML trees (Figs 2 and 3), we obtained some mixed clusters, with, e.g. *O. purpurea* accessions clustering with *O. simia* or *O. militaris* rather than with accessions of the same species. This might be explained by hybridization and genetic introgression that often occur among these species (Bateman 2009; Jacquemyn *et al.* 2012, Bersweden *et al.* 2021). Our samples of *Orchis purpurea* (*O. purpurea1_1* and *O. purpurea1_2*), were collected as 'morphologically different' individuals in hybrid zones (Bersweden *et al.* 2021) and exhibit an identical ITS sequence (Fig. 3). To the best of our knowledge, this is the first published case in which a hybrid or introgressed orchid displays the plastid DNA sequences identical to one parental species and ITS sequences identical to the other. Even though we might have expected two different sequences of ITS in these samples, one from each of the parental species, we detected some sequence variation but were unable to obtain a second consensus ITS copy. This phenomenon can be explained with concerted evolution that has been reported to eliminate all but one copy relatively rapidly during introgression in orchids (Pillon *et al.* 2007) and in the genus *Orchis* [Sanger sequencing of ITS from introgressed individuals in *Orchis* revealed only one ITS copy, the second copy presumably having been lost by concerted evolution (Fay *et al.*, unpublished data)]. Similarly, *O. quadripunctata* in the ITS-based tree clustered correctly with other GenBank accession in section *Pusillae* (Fig. 3), but was intermingled with specimens included in the section *Robustocalcare* when considering plastome analyses (Figs 2–4). This may be explained by genetic introgression in the specimen analysed; alternatively, it could raise questions about the robustness of section *Pusillae*. The specimen of

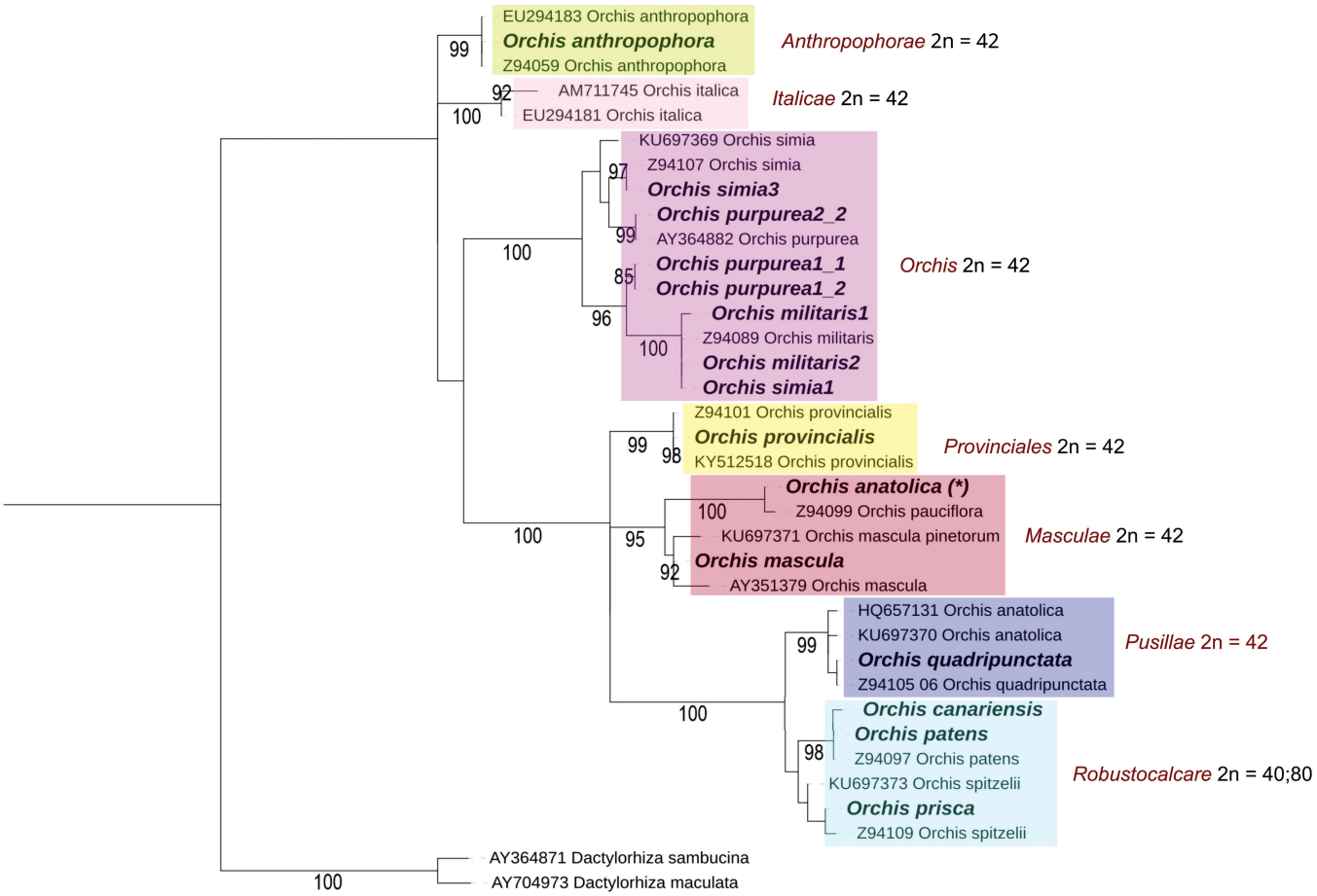

Figure 3. Phylogenetic tree based on ITS sequences retrieved from high-throughput sequencing with accessions downloaded from GenBank (accession numbers are reported). The sample of *Orchis anatolica* (*) collected from Kew Herbarium failed to cluster together with other accessions available from GenBank. The RAxML tree was built in CIPRES Science Gateway. *Dactylorhiza maculata* and *D. sambucina* were used as the outgroups. Sections and ploidy levels are reported.

O. anatolica did not cluster with other specimens of the same species in the ITS tree (Fig. 3) and it clustered with section *Masculae* instead of *Pusillae* when considering plastid sequences. However, our data and a subsequent herbarium inspection clarified the identity of this individual, which was re-identified as *O. pauciflora*, and its plastome is deposited under that name.

Orchis canariensis and *O. patens* clustered together with strong bootstrap values, in accordance with ITS-based analyses (Bateman *et al.* 2003; Fig. 3). However, plastid genome analysis provided no further evidence than nuclear microsatellites (Calevo *et al.* 2021). This might also be due to the relatively recent divergence of *O. canariensis* and *O. patens*, although Bateman *et al.* (2021) also observed a relatively poor resolution of the phylogenetic relationships based on whole-plastome data in the closely related genus *Ophrys* L. Estimation of divergence times showed that the clade of tetraploid species (specifically *O. patens* and *O. canariensis*) diverged from the other *Orchis* spp. more recently (*c.* 3.48 Mya) than the other groups, with a further diversification between *O. patens* and *O. canariensis* occurring *c.* 1.22 Mya. As previously hypothesized (Calevo *et al.* 2021), it seems that this group of taxa is still undergoing diversification and speciation, resulting in low genetic differentiation. Furthermore, an independent evolution of the two species as allotetraploids deriving from different subspecies of the same parental species, members of the *O. mascula* group and *O. spitzelii* group (to which *O. spitzelii* subsp. *nitidifolia* belongs) would also result in low differentiation.

Estimation of divergence times among species performed in BEAST2 was consistent with other previously published estimates (e.g. Inda *et al.* 2012, Sramkó *et al.* 2014, Hoffmann *et al.* 2015). The analysis indicated a recent divergence occurred predominantly during the early Pliocene to late Pleistocene (Fig. 4). In particular, after the divergence between subgenera *Orchis* and *Masculae* during the Miocene (10.53 Mya), the main divergence of what we recognize today as sections occurred between 6.53 and 3.48 Mya. This result is in agreement with the evolution of other Mediterranean plant lineages (e.g. González-Martínez *et al.* 2010, Tremetsberger *et al.* 2016, Benítez-Benítez *et al.* 2018) and the geological and climatic history of the Mediterranean Basin. The Messinian Salinity Crisis that occurred 5.96–5.30 Mya (Pliocene) was a crucial period that contributed to extinction of the subtropical flora and diversification of Mediterranean lineages adapted to aridity (Rodríguez-Sánchez *et al.* 2008, Fiz-Palacios and Valcárcel 2013). This local aridification and global cooling began in the late Miocene and culminated with the commencement of the Mediterranean climatic regime in the

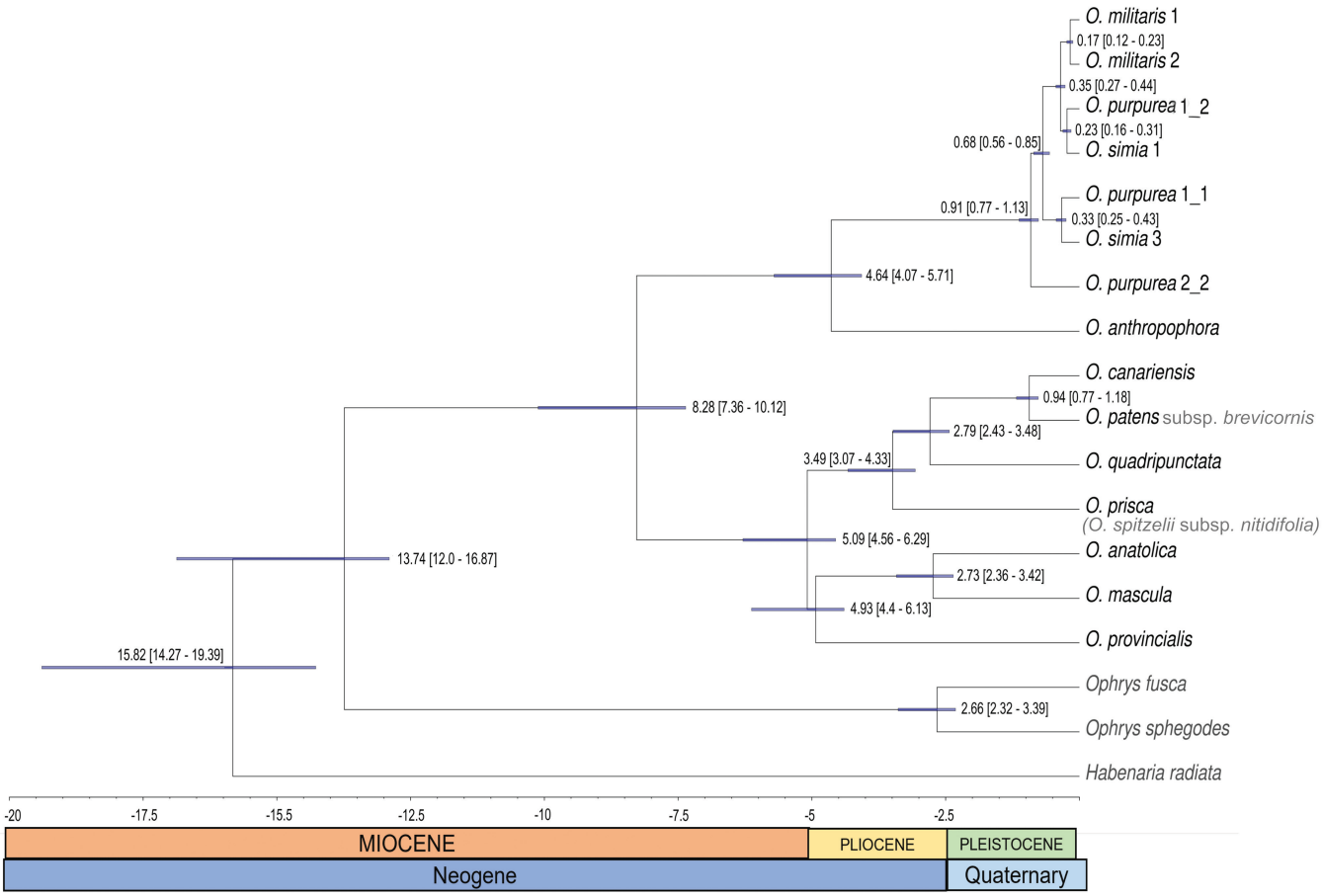

Figure 4. BEAST2 analysis of 15 *Orchis* spp. plastid genomes and divergence time estimates. *Ophrys sphegodes*, *O. fusca* subsp. *iricolor* (*O. fusca*), and *Habenaria radiata* were used as outgroups. The sample of *Orchis anatolica* (*), after molecular analysis and herbarium inspection, has been re-identified as *O. pauciflora.*

late Pliocene (3.4–2.8 Mya) and greatly modified the composition and structure of the Mediterranean flora (e.g. Milne and Abbott 2002, Thompson 2005). Later climate variability (from *c.* 2.5 Mya), characterized by alternation of cold (glacial) and warm (interglacial) periods, had a strong impact on the genetic structure of plants (e.g. Mansion *et al.* 2008, Gentili *et al.* 2015, Adamo *et al.* 2023); during this period, the Mediterranean Basin and Macaronesia played a central role as refugia for glacial relicts (Vargas 2007, Médail and Diadema 2009, Gentili *et al.* 2015, Mairal *et al.* 2017), many of which resulted in the formation of new allopatric lineages (Wiens 2004).

The Canary Islands originated from a volcanic hotspot localized at the continental–oceanic boundary west of Fuerteventura (that first emerged *c.* 22–23 Mya) and then progressed along the boundary and westwards induced by the displacement of the African plate, causing the later emergence of the other islands, Gran Canaria emerging *c.* 14.5–15.0 Mya and El Hierro 1.1–1.2 Mya (Carracedo *et al.* 1998, van den Bogaard 2013), the oldest and the youngest of the islands hosting *O. canariensis*, respectively. Although Gran Canaria, Tenerife and La Gomera emerged > 10 Mya (Carracedo *et al.* 1998, van den Bogaard 2013), catastrophic volcanic events, such as the Roque Nublo period (*c.* 5.3–3.7 Mya) for Gran Canaria, are thought to have cause the extinction, random survival on ridge-top refugia, and habitat fragmentation for existing flora (Emerson 2003, Anderson *et al.* 2009). However, the divergence date between *O. patens* and *O. canariensis* suggests that the colonization of Canary Islands happened during the late Pleistocene, and that *O. canariensis* possibly became endemic to the islands during the glaciations from a broader distribution that included western Africa (where neither of the two species can now be found), as suggested by genetic data (Calevo *et al.* 2021); this hypothesis is in line with the conception that most of the non-endemic Canary Island flora from the pre-glacial period became endemic due to the extinction of conspecific populations on the mainland (Caujapé-Castells *et al.* 2022). However, it is important to note that multiple loci should be incorporated in future analyses to reduce the effects of gene tree incongruence that can bias age estimation of branches (Carruthers *et al.* 2022).

CONCLUSIONS

Our results demonstrated a recent evolutionary history for most of the sections of *Orchis*, in agreement with the geo-climatic evolution of the Mediterranean Basin and Macaronesia, with the tetraploid species *O. patens* and *O. canariensis* diverging in the late Pleistocene. Our 15 plastid genomes are the first for *Orchis* and set the basis for further in-depth biogeographical, evolutionary, and population-level studies.

SUPPLEMENTARY DATA

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

ACKNOWLEDGEMENTS

We thank Dr Cristina González-Montelongo for support in sampling of *Orchis canariensis*, and Dr André Schuiteman and Marco Balducci for support in the Kew Herbarium. We thank Professor Andrew Leitch and Dr Jim Clarkson for useful and informative discussions in the early stages of this project, and editors and the reviewer for comments, some of which helped us to improve this paper.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

Plastid genomes sequences are available in GenBank under the accession codes:
O. anthropophora OQ459770; *O. canariensis* OQ459771; *O. mascula* OQ459772; *O. militaris1* OQ459773; *O. militaris2* OQ459774; *O. patens* OQ459775; *O. prisca* (syn. *O. spitzelii* subsp. *nitidifolia*) OQ459776; *O. provincialis* OQ459777; *O. purpurea1_1* OQ459778; *O. purpurea1_2* OQ459779; *O. purpurea2_2* OQ459780; *O. quadripunctata* OQ459781; *O. simia1* OQ459782; *O. simia3* OQ459783; *O. pauciflora* OQ836201. The sequencing data have been deposited in NCBI BioProject PRJNA1112338 with the Biosample accessions SAMN 41419049-SAMN 41419061.

REFERENCES

Adamo M, Skokanová K, Bobo-Pinilla J *et al.* Molecular evidence and environmental niche evolution at the origin of the disjunct distribution in three mountain endemic *Tephroseris* (Asteraceae) of the Mediterranean basin. *Alpine Botany* 2023;**133**:117–33. https://doi.org/10.1007/s00035-023-00300-w

Anderson CL, Channing A, Zamuner AB. Life, death and fossilization on Gran Canaria – implications for Macaronesian biogeography and molecular dating. *Journal of Biogeography* 2009;**36**:2189–201. https://doi.org/10.1111/j.1365-2699.2009.02222.x

Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (March 2018, date last accessed).

Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 2012;**19**:455–77. https://doi.org/10.1089/cmb.2012.0021

Barido-Sottani J, Bošková V, Du Plessis L *et al.* Taming the BEAST—a community teaching material resource for BEAST 2. *Systematic Biology* 2018;**67**:170–4. https://doi.org/10.1093/sysbio/syx060

Barrett CF, Sinn BT, Kennedy AH. Unprecedented parallel photosynthetic losses in a heterotrophic orchid genus. *Molecular Biology and Evolution* 2019;**36**:1884–901. https://doi.org/10.1093/molbev/msz111

Bateman RM. Evolutionary classification of European orchids: the crucial importance of maximising explicit evidence and minimising authoritarian speculation. *Journal Europäischer Orchideen* 2009;**41**:243–318.

Bateman RM. Circumscribing species in the European orchid flora: multiple datasets interpreted in the context of speciation mechanisms. *Berichte aus den Arbeitskreisen Heimische Orchideen* 2012;**29**:160–212.

Bateman RM, Hollingworth PM, Preston J *et al.* Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). *Botanical Journal of the Linnean Society* 2003;**142**:1–40.

Bateman RM, Rudall PJ, Murphy ARM *et al.* Whole plastomes are not enough: phylogenomic and morphometric exploration at multiple demographic levels of the bee orchid clade *Ophrys* sect. *Sphegodes*. *Journal of Experimental Botany* 2021;**72**:654–81. https://doi.org/10.1093/jxb/eraa467

Bedoya AM, Ruhfel BR, Philbrick CT *et al.* Plastid genomes of five species of riverweeds (Podostemaceae): structural organization and comparative analysis in Malpighiales. *Frontiers in Plant Science* 2019;**10**:1035. https://doi.org/10.3389/fpls.2019.01035

Benítez-Benítez C, Escudero M, Rodríguez-Sánchez F *et al.* Pliocene–Pleistocene ecological niche evolution shapes the phylogeography of a Mediterranean plant group. *Molecular Ecology* 2018;**27**:1696–713. https://doi.org/10.1111/mec.14567

Bersweden L, Viruel J, Schatz B *et al.* Microsatellites and petal morphology reveal new patterns of admixture in *Orchis* hybrid zones. *American Journal of Botany* 2021;**108**:1388–404. https://doi.org/10.1002/ajb2.1710

van den Bogaard P. The origin of the Canary Island Seamount Province - new ages of old seamounts. *Scientific Reports* 2013;**3**:2107. https://doi.org/10.1038/srep02107

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina Sequence Data. *Bioinformatics* 2014;**30**:2114–20. https://doi.org/10.1093/bioinformatics/btu170

Bouckaert R, Vaughan TG, Barido-Sottani J *et al.* BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 2019;**15**:e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Calevo J, Christenhusz MJM, Fay MF. A taxonomic and nomenclatural overview of *Orchis* sect. *Robustocalcare* (Orchidaceae). *Phytotaxa* 2023;**592**:157–62. https://doi.org/10.11646/phytotaxa.592.2.10

Calevo J, Gargiulo R, Bersweden L *et al.* Molecular evidence of species- and subspecies-level distinctions in the rare *Orchis patens s.l.* and implications for conservation. *Biodiversity and Conservation* 2021;**30**:1293–314. https://doi.org/10.1007/s10531-021-02142-6

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972–3. https://doi.org/10.1093/bioinformatics/btp348

Carracedo J, Day S, Guillou H *et al.* Hotspot volcanism close to a passive continental margin: the Canary Islands. *Geological Magazine* 1998;**135**:591–604. https://doi.org/10.1017/S0016756898001447

Carruthers T, Sun M, Baker WJ *et al.* The implications of incongruence between gene tree and species tree topologies for divergence time estimation. *Systematic Biology* 2022;**71**:1124–46. https://doi.org/10.1093/sysbio/syac012

Caujapé-Castells J, García-Verdugo C, Sanmartín I *et al.* The late Pleistocene endemicity increase hypothesis and the origins of diversity in the Canary Islands flora. *Journal of Biogeography* 2022;**49**:1469–80. https://doi.org/10.1111/jbi.14394

Chan CX, Gross J, Yoon HS *et al.* Plastid origin and evolution: new models provide insights into old problems. *Plant Physiology* 2011;**155**:1552–660.

Darriba D, Taboada G, Doallo R *et al.* jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 2012;**9**:772.

Delforge P. *Orchidées d'Europe, d'Afrique du Nord et du Proche-Orient. 4th edn*. Paris: Delachaux & Niestlé, 2016.

Drummond AJ, Bouckaert RR. *Bayesian Evolutionary Analysis with BEAST*. Cambridge: Cambridge University Press, 2015. https://doi.org/10.1017/CBO9781139095112

Eckert CG, Samis KE, Lougheed SC. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology* 2008;**17**:1170–88. https://doi.org/10.1111/j.1365-294X.2007.03659.x

Emerson BC. Genes, geology and biodiversity: faunal and floral diversity on the island of Gran Canaria. *Animal Biodiversity and Conservation* 2003;**26**:9–20. https://raco.cat/index.php/ABC/article/view/57471

Fiz-Palacios O, Valcárcel V. From Messinian crisis to Mediterranean climate: a temporal gap of diversification recovered from multiple plant

phylogenies. *Perspectives in Plant Ecology, Evolution and Systematics* 2013;**15**:130–7. https://doi.org/10.1016/j.ppees.2013.02.002
Foelsche W, Jakely D. *Androrchis* × *koenighoferae*, die Hybride zwischen *Androrchis ovalis* and *Androrchis pauciflora* in Istrien. *Journal Europäischer Orchideen* 2009;**41**:319–58.
Frankel L, Murúa M, Espíndola A. Biogeography and ecological drivers of evolution in the Andes: resolving the phylogenetic backbone for *Calceolaria* (Calceolariaceae). *Botanical Journal of the Linnean Society* 2022;**199**:76–92. https://doi.org/10.1093/botlinnean/boab079
Gamarra R, Ortúñez E, Galán Cela P *et al*. *Anacamptis* versus *Orchis* (Orchidaceae): seed micromorphology and its taxonomic significance. *Plant Systematics and Evolution* 2012;**298**:597–607. https://doi.org/10.1007/s00606-011-0569-1
Gentili R, Bacchetta G, Fenu G *et al*. From cold to warm-stage refugia for boreo-alpine plants in southern European and Mediterranean mountains: the last chance to survive or an opportunity for speciation? *Biodiversity* 2015;**16**:247–61. https://doi.org/10.1080/14888386.2015.1116407
González-Martínez SC, Dubreuil M, Riba M *et al*. Spatial genetic structure of *Taxus baccata* L. in the western Mediterranean Basin: past and present limits to gene movement over a broad geographic scale. *Molecular Phylogenetics &. Evolution* 2010;**55**:805–15. https://doi.org/10.1016/j.ympev.2010.03.001
Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 2019;**47**:W59–64. https://doi.org/10.1093/nar/gkz238
Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 2006;**22**:2971–2. https://doi.org/10.1093/bioinformatics/btl505
Hedges SB, Marin J, Suleski M *et al*. Tree of Life reveals clock-like speciation and diversification. *Molecular Biology and Evolution* 2015;**32**:835–45. https://doi.org/10.1093/molbev/msv037
Hoffmann V, Verboom GA, Cotterill FP. Dated plant phylogenies resolve Neogene climate and landscape evolution in the Cape Floristic Region. *PLoS ONE* 2015;**10**:e0137847. https://doi.org/10.1371/journal.pone.0137847
Huang X, Madan A. CAP3: a DNA sequence assembly program. *Genome Research* 1999;**9**:868–77. https://doi.org/10.1101/gr.9.9.868
Inda LA, Pimentel M, Chase MW. Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. *Annals of Botany* 2012;**110**:71–90. https://doi.org/10.1093/aob/mcs083
Jacquemyn H, Brys R, Honnay O *et al*. Asymmetric gene introgression in two closely related *Orchis* species: evidence from morphometric and genetic analyses. *BMC Evolutionary Biology* 2012;**12**:178. https://doi.org/10.1186/1471-2148-12-178
Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 2013;**30**:772–80. https://doi.org/10.1093/molbev/mst010
Keeling PJ. Diversity and evolutionary history of plastids and their hosts. *American Journal of Botany* 2004;**91**:1481–93. https://doi.org/10.3732/ajb.91.10.1481
Kim Y-K, Jo S, Cheon S-H *et al*. Plastome evolution and phylogeny of Orchidaceae, with 24 new sequences. *Frontiers in Plant Science* 2020;**11**:22. https://doi.org/10.3389/fpls.2020.00022
Klimpert NJ, Mayer JLS, Sarzi DS *et al*. Phylogenomics and plastome evolution of a Brazilian mycoheterotrophic orchid, *Pogoniopsis schenckii*. *American Journal of Botany* 2022;**109**:2030–50. https://doi.org/10.1002/ajb2.16084
Kretzchmar H, Eccarius W, Dietrich H. *The Orchid Genera Anacamptis, Orchis, Neotinea*. Bürgel: EchinoMedia, 2007.
Lee SY, Turjaman M, Chaveerach A *et al*. Phylogenetic relationships of *Aquilaria* and *Gyrinops* (Thymelaeaceae) revisited: evidence from complete plastid genomes. *Botanical Journal of the Linnean Society* 2022;**200**:344–59. https://doi.org/10.1093/botlinnean/boac014
Li YX, Li ZH, Schuiteman A *et al*. Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Molecular Phylogenetics and Evolution* 2019;**139**:106540. https://doi.org/10.1016/j.ympev.2019.106540
Lohse M, Drechsel O, Kahlau S *et al*. OrganellarGenomeDRAW - a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* 2013;**41**:W575–81. https://doi.org/10.1093/nar/gkt289
Mairal M, Sanmartín I, Pellissier L. Lineage-specific climatic niche drives the tempo of vicariance in the Rand Flora. *Journal of Biogeography* 2017;**44**:911–23. https://doi.org/10.1111/jbi.12930
Mansion G, Rosenbaum G, Schoenenberger N *et al*. Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean Basin by the angiosperm family Araceae. *Systematic Biology* 2008;**57**:269–85. https://doi.org/10.1080/10635150802044029
Médail F, Diadema K. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* 2009;**36**:1333–45. https://doi.org/10.1111/j.1365-2699.2008.02051.x
Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, 1–8, https://doi.org/10.1109/GCE.2010.5676129. 2010.
Milne RI, Abbott RJ. The origin and evolution of Tertiary relict floras. *Advances in Botanical Research* 2002;**38**:281–314.
Mudunuri SB, Nagarajaram HA. IMEx: imperfect microsatellite extractor. *Bioinformatics* 2007;**23**:1181–7. https://doi.org/10.1093/bioinformatics/btm097
Neubig KM, Whitten WM, Abbott JR *et al*. Variables affecting DNA preservation in archival plant specimens. In: Applequist WL, Campbell LM (eds), *Proceedings of the U.S. Workshop on DNA Banking*, 81–136. St. Louis: The William L. Brown Center at the Missouri Botanical Garden, 2014.
Olejniczak SA, Łojewska E, Kowalczyk T *et al*. Chloroplasts: state of research and practical applications of plastome sequencing. *Planta* 2016;**244**:517–27. https://doi.org/10.1007/s00425-016-2551-1
Philippe H, Brinkmann H, Lavrov DV *et al*. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology* 2011;**9**:e1000602. https://doi.org/10.1371/journal.pbio.1000602
Pillon Y, Fay MF, Hedrén M *et al*. Evolution and biogeography of European species complexes in *Dactylorhiza* (Orchidaceae). *Taxon* 2007;**56**:1185–208.
POWO. 2023. *Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew*. Published on the Internet; http://www.plantsoftheworldonline.org/. (17 February 2023, date last accessed).
Rambaut A, Suchard MA, Xie D *et al*. *Tracer v1.6*. 2014. http://beast.bio.ed.ac.uk/Tracer (October 2022, date last accessed).
Rodríguez-Sánchez F, Pérez-Barrales R, Ojeda F *et al*. The Strait of Gibraltar as a melting pot for plant biodiversity. *Quaternary Science Reviews* 2008;**27**:2100–17.
Russel PM, Brewer BJ, Klaere S *et al*. Model selection and parameter inference in phylogenetics using nested sampling. *Systematic Biology* 2018;**68**:219–33. https://doi.org/10.1093/sysbio/syy050
Salamin N, Hodkinson TR, Savolainen V. Towards building the tree of life: a simulation study for all angiosperm genera. *Systematic Biology* 2005;**54**:183–96. https://doi.org/10.1080/10635150590923254
Sánchez R, Serra F, Tárraga J *et al*. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Research* 2011;**39**:W470–4. https://doi.org/10.1093/nar/gkr408
Scopece G, Cozzolino S, Bateman RM. Just what is a genus? Comparing levels of postzygotic isolation to test alternative taxonomic hypotheses in Orchidaceae subtribe Orchidinae. *Taxon* 2010;**59**:1754–64. https://doi.org/10.1002/tax.596009

Serna-Sánchez MA, Pérez-Escobar OA, Bogarín D *et al.* Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. *Scientific Reports* 2021;**11**:6858. https://doi.org/10.1038/s41598-021-83664-5

Sramkó G, Attila MV, Hawkins JA *et al.* Molecular phylogeny and evolutionary history of the Eurasiatic orchid genus *Himantoglossum s.l.* (Orchidaceae). *Annals of Botany* 2014;**114**:1609–26. https://doi.org/10.1093/aob/mcu179

Stamatakis A. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3. https://doi.org/10.1093/bioinformatics/btu033

Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 2007;**56**:564–77. https://doi.org/10.1080/10635150701472164

Thompson JD. *Plant Evolution in the Mediterranean*. Oxford: Oxford University Press, 2005.

Tillich M, Lehwark P, Pellizzer T *et al.* GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 2017;**45**:W6–W11. https://doi.org/10.1093/nar/gkx391

Tremetsberger K, Ortiz MA, Terrab A *et al.* Phylogeography above the species level for perennial species in a composite genus. *AoB Plants* 2016;**8**:142–59.

Tyteca D, Ceinos M, Gathoye J-L *et al.* On the morphological, biological and genetic heterogeneity of the genus *Orchis* (Orchidaceae, Orchidinae). *Phytotaxa* 2012;**75**:19–32.

Vargas P. Are Macaronesian Islands refugia of relict plant lineages?: A molecular survey. In: Weiss S, Ferrand N (eds), *Phylogeography of Southern European Refugia*. Dordrecht: Springer, 2007.

Viruel J, Conejero M, Hidalgo O *et al.* A target capture-based method to estimate ploidy from herbarium specimens. *Frontiers in Plant Science* 2019;**10**:937. https://doi.org/10.3389/fpls.2019.00937

Wiens JJ. What is speciation and how should we study it? *The American Naturalist* 2004;**163**:914–23. https://doi.org/10.1086/386552

Yuan Y, Jin X, Liu J *et al.* The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nature Communications* 2018;**9**:1615. https://doi.org/10.1038/s41467-018-03423-5

Zhang Q, Ye J-F, Le C-T *et al.* New insights into the formation of biodiversity hotspots of the Kenyan flora. *Diversity and Distributions* 2022;**28**:2696–711. https://doi.org/10.1111/ddi.13624