Universiteit Leiden
The Netherlands

Exploring the potential of Angiosperms353 markers for species identifcation of eastern Mediterranean orchids

Anthoons, B.; Veltman, M.A.; Tsiftsis, S.; Gravendeel, B.; Drouzas, A.D.; Boer, H. de; Madesis, P.

Citation
Anthoons, B., Veltman, M. A., Tsiftsis, S., Gravendeel, B., Drouzas, A. D., Boer, H. de, & Madesis, P. (2025). Exploring the potential of Angiosperms353 markers for species identifcation of eastern Mediterranean orchids. *Molecular Phylogenetics And Evolution*, *209*. doi:10.1016/j.ympev.2025.108360

Version:	Publisher's Version
License:	Creative Commons CC BY 4.0 license
Downloaded from:	https://hdl.handle.net/1887/4260047

Note: To cite this publication please use the final published version (if applicable).

Molecular Phylogenetics and Evolution 209 (2025) 108360

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Exploring the potential of Angiosperms353 markers for species identification of Eastern Mediterranean orchids

Bastien Anthoons [a,b,1], Margaretha A. Veltman [c,e,1], Spyros Tsiftsis [d], Barbara Gravendeel [e,f], Andreas D. Drouzas [a,*], Hugo de Boer [c,*], Panagiotis Madesis [b,g,*]

[a] Lab. of Systematic Botany and Phytogeography, School of Biology, P.O. Box: 104, Aristotle University of Thessaloniki GR-54124 Thessaloniki, Greece
[b] Institute of Applied Biosciences, CERTH, 6th km Charilaou-Thermis Road, Thermi, GR-57001 Thessaloniki, Greece
[c] Natural History Museum, University of Oslo, Postboks 1172, Blindern, 0318 Oslo, Norway
[d] Department of Forest and Natural Environment Sciences, Democritus University of Thrace, Drama GR-66132, Greece
[e] Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, the Netherlands
[f] Radboud Institute of Environmental and Biological Sciences, Heyendaalseweg 135, 6500 GL Nijmegen, the Netherlands
[g] Laboratory of Molecular Biology of Plants, School of Agricultural Sciences, University of Thessaly GR-38446 Thessaly, Greece

ARTICLE INFO

Keywords:
Orchidinae
Phylogenetics
Angiosperms353
Species delimitation
Ophrys
Target capture

ABSTRACT

Tuberous orchids are ecologically vulnerable species, threatened by a range of environmental pressures such as overharvesting, grazing and land use change. Conservation efforts require accurate species identification, but are impeded by limited phylogenetic resolution of traditional genetic markers, which is exacerbated by widespread taxonomic conflict regarding the classification of orchids. Target enrichment holds promise to resolve both these challenges by offering a large set of nuclear loci with which to increase phylogenetic resolution and evaluate competing species models.

Here, we evaluate the effectiveness of the Angiosperms353 markers for distinguishing over 50 tuberous orchid species native to Greece and we explore the possibility of narrowing these markers to a smaller set that could function as a minimal probe set. Our methodology consists of a three-tiered approach: 1) generating a species-level phylogeny using all Angiosperms353 loci with sufficient target recovery, 2) evaluating competing species models based on "splitter" and "lumper" classifications through Bayes Factor species delimitation, and 3) ranking the potential of Angiosperms353 loci to discriminate representatives of lineages with different divergence times based on their phylogenetic informativeness. While the inferred multi-species coalescent phylogeny had overall high support, Bayes Factor delimitation revealed mixed outcomes, favouring splitting in *Serapias*, while favouring splitting in basal clades and lumping in more recently diverged clades in *Ophrys*. A molecular clock analysis of *Ophrys* confirms rapid and recent radiation in clades marked by phylogenetic uncertainty, suggesting the need for additional loci to fully resolve this genus. Finally, we found 30 loci to be highly phylogenetically informative across four epochs of *Orchidinae* evolution; we suggest these are promising candidates for future marker development. Our findings enhance the Plant Tree of Life (PAFTOL) by contributing additional phylogenomic data for species that were previously underrepresented in trees built with these markers, while shedding light on the ongoing "splitter"-vs-"lumper" debate and offering new directions for species identification of tuberous orchids, a group with distinct taxonomic and conservation challenges.

1. Introduction

The Orchidaceae family, recognised as one of the two largest families of angiosperms (Christenhusz and Byng, 2016), inhabits a wide range of terrestrial biomes. With a count of at least 25,000 species and 700 genera, orchids exhibit huge taxonomic and geographical diversity (Pérez-Escobar et al., 2021). Beyond their diversity, orchids play a significant role in global trade and usage, both sustainably and

* Corresponding authors at: Laboratory of Molecular Biology of Plants, School of Agricultural Sciences, University of Thessaly, GR-38446 Thessaly, Greece (P. Madesis).
E-mail addresses: drouzas@bio.auth.gr (A.D. Drouzas), h.de.boer@nhm.uio.no (H. de Boer), pmadesis@certh.gr (P. Madesis).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ympev.2025.108360
Received 2 October 2024; Received in revised form 26 March 2025; Accepted 22 April 2025
Available online 25 April 2025
1055-7903/© 2025 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

unsustainably, legally and illegally (Hinsley et al., 2018). Among their various uses, the category of edible orchids often remains overlooked. For instance, vanilla (*Vanilla* Plumier ex Mill.) is an orchid-derived product and the world's most popular plant-derived flavour. Additionally, edible products such as chikanda, an African snack derived from orchid tubers, and salep, a popular drink crafted from orchid tubers and widely consumed in Turkey, Greece, Iran, Iraq, and Albania (De Boer et al., 2017; Hanlidou et al., 2004), highlight the cultural and regional importance of these taxa. Wild harvesting and trade for salep production primarily involve species from the genera *Orchis, Anacamptis, Dactylorhiza, Himantoglossum* and *Ophrys* with as many as 35 different species being harvested and sold as salep across the Eastern Mediterranean (Ghorbani et al. 2017; Kasparek and Grimm 1999; Kreziou et al. 2015). These orchids, characterised by fleshy roots and tubers, are used as raw material for salep, and are part of the subtribe Orchidinae (tribe Orchideae,) within the Orchidoideae subfamily, with Orchidinae comprising around 1,800 species globally (Jin et al., 2017; Govaerts et al. 2021).

The salep trade operates on an international scale, with Turkey exporting a significant portion of its harvest to countries such as the Netherlands, Cyprus, and Germany, which is a key market for medicinal plants, including salep orchids (Kasparek and Grimm, 1999). While this trade is predominantly linked to Turkey and the Turkish diaspora in Europe, a recent analysis of e-commerce detected by a web crawler (an automated software tool for searching and indexing website content) indicates a notable presence of salep trade in Greece(Masters et al., 2022). The growing demand for traditional, organic, and alternative food products has increased the consumption and harvesting of salep in Greece, resulting in rising market prices (Kreziou et al., 2015). However, the ecological vulnerabilities of orchids—due to their limited distribution, specific habitat requirements, dependence on pollinators, and mycorrhizal symbionts—make them susceptible to overharvesting (McCormick and Jacquemyn, 2014; Tsiftsis and Tsiripidis, 2020). Protection from overexploitation is essential to meet regional sustainability and conservation goals (Roy et al., 2017; Tsiftsis, 2021; Warghat et al., 2013). Yet, despite legal protection, enforcement of bans on orchid collection, particularly orchid tubers, remains inadequate (Ghorbani et al., 2014; Kreziou et al., 2015). This situation highlights the urgent need for further characterisation of Greece's remaining orchid diversity, which includes a total of 194 taxa across 19 genera and 178 species (of which 147 species in the Orchidinae subtribe) (Tsiftsis and Antonopoulos, 2017), to better inform their protection and species management.

In the past twenty years, molecular phylogenetic studies have significantly enhanced our understanding of the subtribe Orchidinae s.s. on the basis of a limited number of nuclear and chloroplast markers (Bateman et al., 2003; Inda et al., 2010, 2012; Jin et al., 2017). Cost-effective application of DNA barcoding using such markers in plants has been aided by advances in high-throughput sequencing (HTS) (Hollingsworth et al., 2016; Lemmon et al., 2012), while simultaneous developments in phylogenomic analysis have further enabled the identification of appropriate markers for species delimitation in the Orchidinae subfamily (Jin et al., 2017). However, effective DNA barcoding is hampered by phylogenetic uncertainty, which disproportionately affects the orchid family where some lineages are characterised by rapid diversification, and conflict over specific and generic circumscriptions is widespread. An example of such a lineage is the morphologically diverse *Ophrys* genus, which has been taxonomically delineated into several species complexes, including the *O. lutea, O. fusca, O. fuciflora, O. scolopax* and *O. sphegodes* groups (Bateman, 2018). Individual differences in the taxonomic approach, sometimes referred to as "lumpers" or "splitters" (Endersby, 2009), depending on their tendency to emphasise on either shared traits or distinctions among orchids, respectively, further complicate the already uncertain phylogenetic placement of orchids and contributes to the complexity of orchid classification. While the dichotomy between "splitters" and "lumpers" might seem to unfairly reduce the complexity of species delimitation to a choice between one of two sides, we use these terms to refer to the taxonomic approach defining species more broadly (i.e. grouping many minor varieties under a single name) and to the one defining species more narrowly (i.e. recognising these varieties as distinct subspecies or even separate species) and will refer to these opposing approaches as either "lumping" or "splitting" throughout the text.

Two notable strategies have been proposed as alternatives to traditional barcoding in the development of next-generation nuclear barcodes, by increasing both coverage and resolution: low-pass shotgun sequencing or genome skimming, and target capture sequencing (Manzanilla et al., 2022; Pérez-Escobar et al., 2021; Zhang et al., 2023). Despite advances in workflows and bioinformatic pipelines for genome skimming, the low sequencing depth of genome skims often limits their analysis to primarily multicopy regions, such as plastid genomes and ribosomal DNA. These regions encompass only a limited number of independent loci, which ultimately restricts the power of this method (Manzanilla et al., 2022). In contrast, target capture sequencing is designed to yield hundreds or thousands of independent loci, offering focused multi-locus datasets that could include standard DNA barcodes from nuclear genes (Schmickl et al., 2016). A significant step up from single locus DNA barcoding, target capture enables phylogenetic analysis at both deep and shallow levels and is therefore a promising tool for analysing both among-species relationships, as well as within-species diversity (Andermann et al., 2019).

While target capture offers distinct benefits for phylogenomics, the substantial investment required for a target capture sequencing study (related to development of the necessary genomic resources, design and synthesis of baits, and computational analysis of the data) has hindered the application of this method when contrasted with traditional barcoding approaches, with only one example focusing on Orchidinae s.s. to date (Veltman et al., 2024). The release of bait kits targeting universally applicable loci, most notably Angiosperms353 (Johnson et al., 2019), as well as standardised bioinformatics pipelines, such as HybPiper (Johnson et al., 2016), show promise in lowering these barriers across flowering plants (Dodsworth et al., 2019; Johnson et al., 2019) and allow for the integration of target capture data sets over large taxonomic distances. The Angiosperms353 bait kit has been successfully applied to various flowering plant groups (Baker et al., 2022; Beck et al., 2021; Pérez-Escobar et al., 2021; Shah et al., 2021; Zuntini et al., 2024), and the orchid family is currently well represented in datasets utilising the Angiosperms353 kit (Pérez-Escobar et al., 2021). However, to date, the included orchid species do not cover all major subfamilies and tribes evenly, with tropical epiphytes from the subfamily Epidendroideae favoured at the expense of terrestrial orchids from more temperate climates, including the Mediterranean.

This study aims to adress this gap, by demonstrating the effectiveness of the Angiosperms353 markers in discriminating eight genera and –40 (61 % out of 65 lumper taxa) to 56 (38 % out of 147 splitter taxa) tuberous Mediterranean orchid species, covering 80 % of the Orchidinae genera native to Greece. Building on a phylogenomic framework of 148 samples, we evaluate competing "splitter" and "lumper" classifications in controversial clades and we identify groups where, based on current data, confident species identification is either attainable or continues to be problematic. To make identification of Orchidinae with Angiosperms353 markers more accessible in the future, we propose a subsample of loci that are able to reconstruct most species relationships and could aid cost-effective identification of orchid-derived samples for applications in conservation and wildlife forensics. To achieve these goals, we employ a three-tiered strategy: 1) we generate a species-level phylogeny for our dataset using all Angiosperms353 markers with sufficient target recovery, 2) we evaluate different species classification models using both "splitter" and "lumper" approaches through Bayes Factor species delimitation, and 3) we rank and validate the potential of Angiosperms353 markers to discriminate representatives of lineages with different divergence times based on their phylogenetic informativeness.

This study contributes valuable data to the broader understanding of Greek orchid diversity andadds to ongoing efforts to characterise the Plant Tree of Life using the same Angiosperms353 baits (Zuntini et al., 2024). Our results serve as a foundation for future phylogenomic investigations of Orchidinae s.s., with applications spanning evolutionary studies, systematics, wildlife forensics, and conservation of these Mediterranean tuberous orchids.

2. Material and methods

2.1. Data collection

2.1.1. Sampling and DNA extraction

In this study, we analysed 165 individuals belonging to 56 orchid species across eight genera within the subtribe Orchidinae s.s. (Orchidaceae; Orchidoideae) (Appendix S1). Based on the Field Guide to the Orchids of Europe and the Mediterranean (Kuhn et al., 2019), which uses broader species definitions, Greece counts a total of 10 native genera within the subtribe Orchidinae, namely *Anacamptis* (9), *Dactylorhiza* (7), *Gymnadenia* (4), *Himantoglossum* (3), *Neotinea* (4), *Ophrys* (16), *Orchis* (14), *Platanthera* (2), *Pseudorchis* (1), and *Serapias* (5), with the number in brackets indicating the respective number of species within each genus, totalling 65 species. Our sampling included 8 out of the 10 genera and 40 out of the 65 species, covering 80 % of the genera and 61 % of the species that are native to Greece. Two Epidendroideae species and one Orchidoideae species from the Cranichideae tribe were included as outgroups, as detailed in Appendix S1. The specimens were collected from various locations across Greece, chosen for their high species richness and sampled across a range of habitats (Fig. 1). The collections were conducted under the appropriate permits issued by the Greek Ministry of Environment and Energy, with reference numbers 177111/6159/19–12-2018, ΥΠΕΝ/ΔΠΔ/92136/2220/10–10-2019 and 144703/2145/4–10-2016. Under Greek legislation, all members of the Orchidaceae family are protected and permission is needed for taking samples from wild plants, especially endemic ones (Presidential Decree 67/1981 and Biodiversity Law 3937–2011). In order not to jeopardise the survival of individual plants, permits were therefore obtained for collecting leaves or flowers instead of whole organisms. Photographs with diagnostic characters were made and are available upon request. DNA extractions from leaf tissue samples were performed using a CTAB method, as described by Doyle and Doyle (1987), with minor modifications. Extracted DNA was diluted in 100 μL and quantified using a Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kit (Invitrogen, United States) (Appendix S1). DNA quality was assessed by gel electrophoresis in 1 % agarose gel.

2.1.2. Library preparation, target capture and high throughput sequencing (HTS)

Capture and sequencing of genomic DNA (gDNA) of the 165 samples were outsourced to Daicel Arbor Biosciences (Ann Arbor, MI, USA). A total of 4 μg or up to 80 % of the gDNA was sonicated and size selected to an approximate average insert size of 300nt and prepared into dual-indexed Illumina-compatible libraries. The indexed libraries were quantified with a spectrofluorimetric assay and a quantitative PCR assay for properly library-adapted molecules.

Capture pools were prepared from up to 133 ng of 9–12 libraries per reaction, pooled based on the input mass of gDNA to library preparation. Each capture pool was concentrated down to 7 μL by vacuum centrifugation. Captures were performed following the myBaits v5.02 protocol using the myBaits Expert Angiosperms353 baits with an overnight

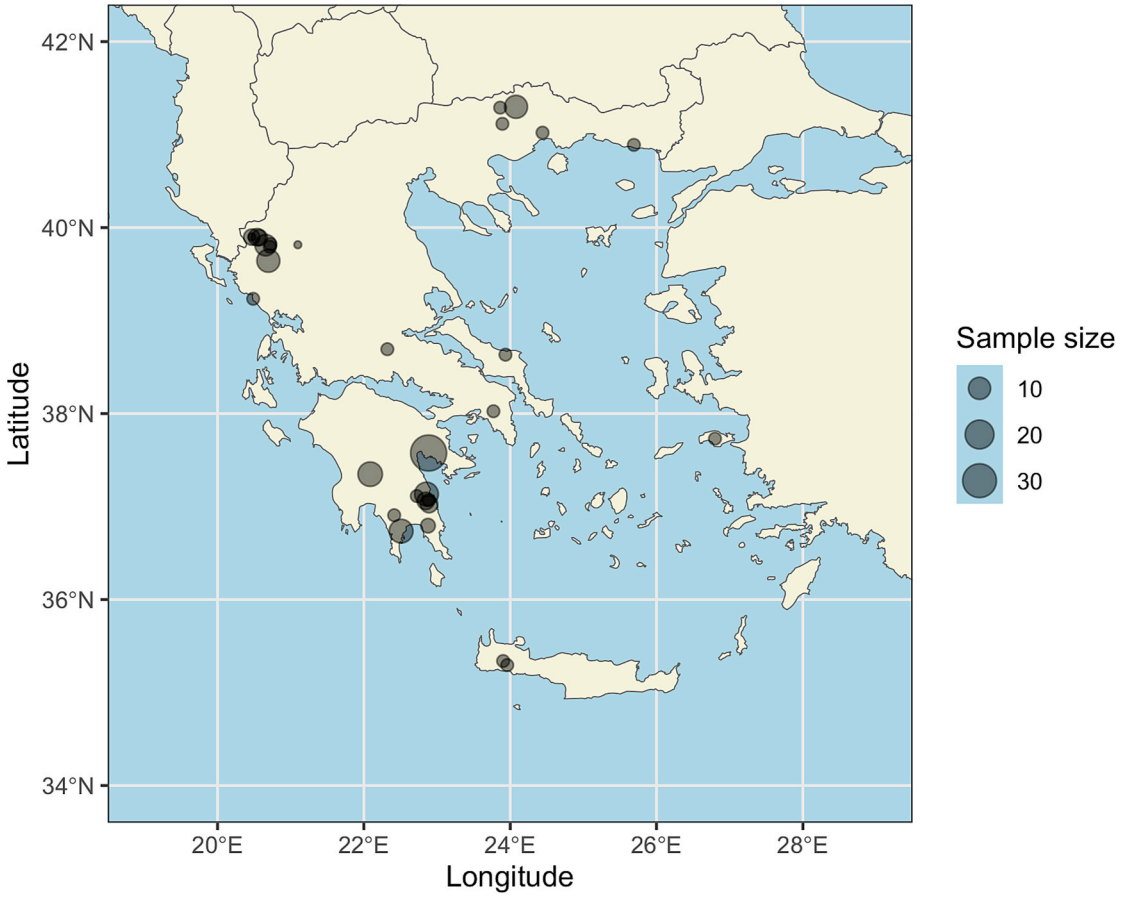

Fig. 1. Sampling locations of the orchids collected in this study, with sample count represented by the relative bubble size. Sample names, species names and respective locations are listed in Appendix S1.

hybridisation and washes at 65 °C. Post-capture, half the volume was amplified for 10 cycles and yield quantified with a spectrofluorimetric assay. For captures that did not yield sufficient material for sequencing, the second half of the capture volume was amplified for 14 cycles, subsequently quantified with a spectrofluorometric assay, and both amplified capture products were combined. The capture quality was assessed via the TapeStation 4200 (Agilent) platform with a High Sensitivity D1000 tape. The captures were pooled in equimolar ratios. Samples were sequenced on an Illumina NovaSeq 6000 platform on a partial S4 PE150 lane to approximately 1.1 Gb per library. Sequencing data has been registered in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1167703.

2.2. DNA sequence data analysis and phylogenomic inference

A detailed visualisation and overview of the bioinformatics steps in this section can be found in Appendix S2.

2.2.1. Locus assembly, alignment and trimming

Demultiplexed reads were quality trimmed with Trimmomatic (v0.39) using the settings "ILLUMINACLIP:adapter.fa:2:30:10:2:TRUE LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:40″ (Bolger et al., 2014). Target sequences were recovered with HybPiper (1.3.2), by first mapping the surviving read pairs against the target reference file using the burrows-wheeler aligner, followed by contig assembly with SPAdes (Prjibelski et al., 2020), intron–exon boundary demarcation with exonerate (Slater and Birney, 2005) and flagging of paralogs (Johnson et al., 2016). To improve target recovery, we used an expanded version of the Angiosperms353 target reference file (Mega353 target file) and filtered this to retain all orchid species included in the 1KP transcriptome data (McLay et al., 2021). Individuals with a target recovery that was more than two standard deviations from the mean exon recovery were removed from the analysis. Remaining recovered exon sequences were aligned with MAFFT (v7.453) using the –-localpair option and maximum 1000 iterations (Katoh and Standley 2013). Alignments were trimmed with trimAL (Capella-Gutiérrez et al., 2009), removing columns with > 20 % gaps as well as spurious sequences where < 75 % of the residues had a residue overlap score of at least 0.5. Alignment statistics were generated with AMAS (Borowiec, 2016).

2.2.2. Gene tree and species tree construction

A maximum likelihood (ML) species tree and ML gene trees were built with the trimmed alignments of all loci containing at least four unique sequences, using IQTREE (v2.1.2) with model selection (Kalyaanamoorthy et al., 2017), 1000 bootstrap replicates and maximum 10,000 iterations (Minh et al., 2020). Gene trees were edited with Newick Utilities (Junier and Zdobnov, 2010) to collapse nodes with a low bootstrap support value of < 30 following Baker et al. (2022), and ThreeShrink to remove implausibly long branches (which could be caused by sequencing, assembly or alignment errors, or by unflagged paralogs), defined as branches whose inclusion would increase the tree diameter by more than 5 % (Mai and Mirarab, 2018). Coalescent trees were built with ASTRAL-III using the edited gene trees, one allowing a free topology (Zhang et al. 2018) and two constraining the samples thought to belong to the same species into monophyletic clusters (Rabiee et al. 2019). Constrained trees were built to contrast competing species classifications, one splitting common taxa into more narrowly defined species – focusing on local diversity and recognising endemics such as *Dactylorhiza pythagorae* Gölz & H.R.Reinhard, *Ophrys helenae* Renz., and *Ophrys zeusii* M. Hirth (Antonopoulos and Tsiftsis, 2017; Charitonidou et al., 2021; Tsiftsis and Antonopoulos, 2017; Tsiftsis et al., 2007) – and another lumping these endemic taxa into larger species constructs (Kühn et al., 2019). Gene concordance factors were calculated for the unconstrained coalescent tree with IQTREE (v2.1.1). We labelled taxa as concordant following the procedure of Baker et al. (2022). For each tree node, two metrics were calculated for all genera and species: (i) the proportion of samples from a given genus/species descending from the node, and (ii) the proportion of descendants from the node belonging to that genus/species. These metrics were multiplied to obtain a concordance score, and the node with the highest score was considered to best represent the taxon, and labelled as concordant if > 0.5. The species assignment for each sample was as follows: (i) Confirmed, if the sample descended from a node with a score > 0.5; (ii) Rejected, if the sample didn't descend from a node with a score > 0.5; or (iii) Inconclusive, if the best-scoring node had a score < 0.5 (Baker et al. 2022). Samples falling outside of concordant clades for their respective taxa (considering the widest species definition) were removed from the analysis. Sequences of the remaining samples were realigned and the phylogenomic inference was repeated after discarding these samples. The taxon concordance metrics were re-calculated for all species (both "splitter" and "lumper" classifications) to quantify the extent of monophyly and the proportion of samples in the tree that could be unequivocally assigned to concordant species constructs. These results were used to evaluate the potential of the markers for species identification based on concordant clade membership.

2.2.3. Divergence time estimation

A full coalescent analysis was performed for the *Ophrys* genus with BEAST (v2.6.4) (Bouckaert et al., 2014), using the SNAPP add-on (Bryant et al., 2012) to conduct a Bayesian divergence time estimation following the methods of Stange et al. (2018). To increase the number of variant sites available for the analysis, a custom reference file was created with EMBOSS (v6.6.0) containing the consensus sequences of all locus alignments, setting the plurality option to 0.005. Raw reads were mapped back to this consensus with the bwa-mem algorithm of the Burrows-Wheeler Aligner (BWA v0.7.17) and the mapped reads were sorted and deduplicated with SAMtools (v1.12). Variants were called with ANGSD (v0.935) for all loci (Korneliussen et al., 2014), excluding those where more than 10 % of the individuals with recovered sequences issued paralog warnings during assembly with HybPiper. Variants with a coverage of less than 10X or more than two standard deviations above the mean (11,500X) were discarded. One or, where possible, two *Ophrys* individuals with the highest target recovery per species were retained, and among this subset of individuals, variants were filtered to keep only biallelic SNPs without missing data and with a minor allele frequency of at least 5 %. The filtered variants were pruned at a thinning distance of 100 bp, and the final Variant Call Format (VCF) was used as input for preparing the SNAPP XML file, with a chain length of 1,000,000 MCMC iterations; fixed and identical forward and reverse mutation rates of 1.0; and a strict molecular clock with a one-on-x prior for the rate. Divergence time estimation was constrained by the crown age of *Ophrys,* estimated to be around 4.9 million years ago (Mya) (Breitkopf et al., 2015). This estimate was approximated as a lognormal distribution with a mean of 1.6 and a standard deviation of 0.22. To account for the possible effects of incongruent species definitions on divergence time estimation, we ran the analysis twice: once with taxa grouped by the narrowest (splitter) species definition, and once with taxa grouped by widest (lumper) species definitions. Convergence for both runs was assessed by checking the effective sample size (ESS) of the estimated parameters after discarding the first 100,000 iterations as burn-in, considering ESS > 200 indicative of good mixing of the chain.

2.2.4. Species delimitation

To test the competing species classification models of splitters and lumpers for clades where these different approaches would lead to different species trees, Bayesian Factor Delimitation (BFD) was performed in BEAST (v2.6.4). Three different analyses were performed, one for *Serapias* using all available individuals, and two for different clades in *Ophrys*: one containing the *O. fuciflora*, *O. scolopax* and *O. sphegodes* complexes (hereafter called "non-basal *Ophrys* clade") and one containing the *O. fusca* and *O. lutea* species complexes (hereafter called "basal *Ophrys* clade"). For the *Ophrys* clades, one to two samples with the

longest cumulative exon recovery per species were selected to reduce computational load. Variant sites were extracted from the multi-fasta locus alignment and converted to the variant call format (VCF) with snp-sites (Page et al., 2016). Variants were filtered with VCFtools (v0.1.16) to retain only biallelic SNPs and subsequently pruned at a thinning distance of 100 bp (Danecek et al., 2011). The filtered variants were converted to PHYLIP format with vcf2phylip (Ortiz, 2018) and used as input to create the XML files for Bayesian Factor Delimitation with BEAUTi, which were subsequently edited to prepare it for marginal likelihood estimation (MLE) through a path sampling analysis (Leaché et al., 2014). Path-sampling was conducted with 48 steps, 1,000,000 MCMC iterations and a pre-burnin of 10,000 iterations. Convergence of the path sampling analysis was assessed by checking ESS values of the likelihood estimates in the log file after discarding the first 100,000 iterations as burn-in, and considered sufficient when the ESS of the likelihood estimate was higher than 200. The analysis was conducted for each clade and species classification model: once with a base scenario using the widest (lumper) species definition, and once with an alternative scenario using the narrowest (splitter) species definition. Bayes Factor (BF) was calculated by subtracting the MLE values of the competing models and then multiplying the difference by two. The results were evaluated using the framework of Kass and Raftery (1995), with $0 < |BF| < 2$ considered not conclusive, $2 < |BF| < 6$ as moderate support, $6 < |BF| < 10$ as strong support, and $|BF| > 10$ as decisive for the model with the highest MLE.

2.3. Marker selection

2.3.1. Phylogenetic informativeness

We evaluated the effectiveness of Angiosperms353 low-copy nuclear genes in resolving orchid relationships by assessing their Phylogenetic Informativeness (PI) and ability to discriminate species of interest. Since PI profiling requires a stable topology, we considered a selection of 42 samples whose species relationships were constant between the coalescent and concatenation approaches of species-tree inference, and well supported by both methods as indicated by bootstrap values of at least 80 % and posterior probabilities of at least 0.9. We prioritised samples with higher target recovery to maximise the amount of phylogenetic information present in the concatenated alignment. The corresponding ML species tree was calibrated in R using the “chronos” function of the ’ape’ package (Paradis and Schliep, 2019). We based our higher level taxa (subfamily and tribe) divergence time estimates on Gustafsson et al. (2010) and lower level taxa divergence time estimates (subtribe and genus) on Inda et al. (2012), additionally constraining the crown of *Ophrys* to the most recent estimate of 4.9 Mya according to Breitkopf et al. (2015). Site substitution rates were estimated for 177 loci, rooted on Epidendroideae, using HyPhy (Pond et al., 2005) followed by PI profiling according to the method of Townsend et al. (2007), both implemented in TAPIR (Faircloth et al., 2012).

Locus PI through time was determined by delimiting epochs of the evolution of the Orchidoideae, so that one epoch would be circumscribed by two well-resolved divergences. The time intervals examined include: 53–61 million years ago (Mya), marking the period from the divergence of Epidendroideae to the split between Cranichideae and Orchideae; 17–53 Mya, from the split of Cranichideae to the crown age of Orchidinae; 9–17 Mya, from the crown age of Orchidinae to the most recent split between genera; 5–9 Mya, a period of significant speciation in older genera; and 0–5 Mya, reflecting more recent speciation events in younger genera and those with contentious species concepts (Gustafsson et al., 2010; Inda et al., 2012). Cumulative informativeness was calculated by integrating over each of these epochs, estimated by taking the Area Under the Curve (AUC). To address potential confounding of ancient phylogenetic signals by more recent ones, we corrected the AUCs with a penalty following Bellot et al. (2020) by checking for each epoch whether its median age was older or younger than the time of the maximum PI of a locus. When the median age of an epoch was older, its AUC was multiplied with the ratio of the PI at the median age to the maximum PI. For each epoch, loci were then ranked based on their AUC. Loci for which the AUC for a given epoch had to be corrected were considered potentially misleading for that epoch and were removed from the set prior to ranking.

2.3.2. Locus ranking and performance assessment

We selected the ten most informative loci per epoch as candidates for a reduced set of markers, and tested their ability to reconstruct the topology of the tree generated with the full set of markers. To this end, we repeated the phylogenomic analyses with both the reduced and the full set of loci for the selected 42 species with a stable topology. Gene trees were pruned in R using the ’ape’ package (Paradis and Schliep, 2019) prior to constructing multi-species coalescent trees with ASTRAL-III (Zhang et al., 2018). The performance of the reduced set of loci was assessed by quantifying the similarity of the resulting tree to the tree based on the full set of loci with the R package ’TreeDist’ (v2.8.0), and by assigning both trees a quartet score based on the full set of pruned gene trees. To conclude whether the candidate loci performed above expectation, we compared their performance to that of 100 different sets of an identical number of randomly subsampled loci. Topological differences between the reduced and the full tree were visualised in a cophyloplot generated with the R package ’phytools’ (Revell, 2012).

3. Results

3.1. Target recovery is variable but taxonomically independent

Sequencing yielded 1250 million reads, at an average of 7.5 million reads (~1.1 Gb) per sample. Target enrichment and recovery statistics are reported in Table 1. Enrichment success was variable, with 10 % of the total number of reads, ranging from 1–44 % per sample, mapping on target. Median exon recovery was 112 kb, amounting to about 42 % of the total target length of 266 kb spanning all reference loci. On average, samples had a target recovery of at least 50 % for 120 loci. This number dropped to 58 loci when considering a coverage threshold of 75 % but rose to 200 loci when considering a threshold of 25 %, and 250 when considering sequences of any length.

Relative target recovery per locus was variable among samples, but not taxonomically biased (Appendix S3). Expansion of the target reference file with the 1KP orchid transcriptome sequences (Mega353 target file) yielded a total target length that was only slightly (~ 5 kb) higher than the original (Table 1). Nonetheless, this strategy resulted in a considerable improvement in target recovery compared to the original Angiosperms353 reference sequences, where the median recovery was only 79 kb (or 30 % of the total of 261 kb), hence raising the amount of available data for phylogenetic analysis by more than 40 %.

3.2. Nuclear phylogenomics of Orchidinae s.s.

In total 17 samples were discarded due to poor target recovery or taxonomic discordance. The remaining 148 samples were used for phylogenomic analyses. Alignments were made for 348 loci, which after quality trimming had an average length of 458 bp and 17 % missing data (Appendix S4). Only loci with at least four sequences, 335 in total, were considered. The number of retained sequences per alignment ranged from 4 to 147, averaging 83. The average GC content was 45 %, and 29 % of all sites were parsimony informative.

The normalised quartet score of 0.82 derived from coalescent inference with ASTRAL-III suggests that the majority of gene tree quartets are in agreement with the species tree (Fig. 2). However, gene concordance factors indicated that on average, less than half of the gene trees (153/335) were informative for each branch, and roughly 10 % (34) were concordant. Posterior Probability (PP) values for the species tree were high for the majority of branches, with 81 out of 145 nodes displaying PP > 80, supporting relationships between all genera. Species

Table 1
Target recovery statistics of 165 orchid DNA samples enriched with the Angiosperms353 probe kit. Raw sequencing reads were processed with HybPiper (v1.3.2) and mapped against two different reference files, the original Angiosperms353 targets, and these same targets but supplemented with additional orchid sequences from the Mega353 targets (McLay et al., 2021).

Reference file	Angiosperms353 targets (260,802 bp)					Mega353 targets (266,084 bp)				
Statistic	Reads mapped on target	% mapped on target	Total target recovery (bp)	% target recovery	Genes > 50 %	Reads mapped on target	% mapped on target	Total target recovery (bp)	% target recovery	Genes > 50 %
Median	491,436	7.39 %	79,488	30.48 %	77	552,828	8.13 %	112,632	42.33 %	130
Mean	774,505	9.28 %	72,851	27.93 %	76	825,471	9.95 %	104,126	39.13 %	120
Minimum	51,485	1.26 %	0	0.00 %	0	55,822	1.36 %	621	0.23 %	0
25th percentile	285,702	5.10 %	53,880	20.66 %	45	311,401	5.61 %	73,824	27.74 %	66
75th percentile	778,582	10.28 %	95,376	36.7 %	110	840,966	11.48 %	137,220	51.57 %	172
Maximum	9,151,209	43.95 %	210,549	80.73 %	294	9,317,006	44.42 %	211,818	79.61 %	296

relationships are well-resolved in the genera *Neotinea*, *Orchis*, *Gymnadenia*, *Dactylorhiza*, and *Anacamptis*, but are marked by more uncertainty in *Himantoglossum*, *Serapias* and *Ophrys.* These genera also display more gene tree-species tree conflict on internal nodes (Fig. 2). Specifically, *Himantoglossum jankae* Somlyay, Kreutz & Óvári and *H.* x *samariense* C. Alibertis & A. Alibertis form a paraphyletic clade, with *H. galilaeum* Shifman nested within it, and *H. robertianum* (Loisel) P. Delforge identified as a sister group to these species (Fig. 1). Within *Serapias*, *S. lingua* L. and *S. parviflora* Parl. form a sister clade to a group comprising *S. bergonii* E.G.Camus*, S. vomeracea* (Burm.f.) Briq. and *S. orientalis* (Greuter) H. Baumann & Künkele, whose relationships are ambiguous.

The genus *Ophrys* is divided into three primary clades. The first includes *O. speculum* Link*, O. tenthredinifera* Willd., and the O. *fusca* s.l. group, which encompasses *O. calocaerina* Devillers-Tersch. & Devillers and *O. cinereophila* Paulus & Gack, along with the *lutea* group comprising *O. lutea* Cav.*, O. melena* (Renz) Paulus & Gack, and *O. sicula* Tinea. The second clade is represented solely by *O. insectifera* L., indicating a distinct lineage within the genus. The third clade comprises *O. attica* (Boiss. & Orph.) Soó and the *O. fuciflora* group (which includes *O. holosericea* subsp. *graeca* B. Baumann & H. Baumann and *O. candica* Greuter, Matthäs & Risse); along with the *O. scolopax* group (*Ophrys heldreichii* (Soó) Devillers-Tersch. & Devillers, and *O. oestrifera* M. Bieb.), *O. ferrum-equinum* Desf.; the *O. sphegodes* group (which includes *O. grammica* (B. Willing & E. Willing) Devillers-Tersch. & Devillers, *O. mammosa* Desf., *O. zeusii*, *O. leucophthalma* Devillers-Tersch. & Devillers, *O. aesculapii* Renz, *O. spruneri* Nyman, and *O. helenae* Renz); *O. reinholdii* Spruner ex Fleischm.; and *O. argolica* H. Fleischm. The larger species complexes within *Ophrys,* being the *lutea* and *fusca* complexes, *fuciflora* and *scolopax* complexes, as well as the more recently diverged species within the *O. sphegodes* complex, display low internal resolution and remain unresolved with the current dataset. Divergence time estimations for the *Ophrys* genus with SNAPP indicate that the large species diversity within this group is predominantly the result of rapid and recent diversification of a few lineages. The phylogenetic relationships among these species complexes are notably uncertain, with divergence events dating back to only a few hundred thousand years (Appendix S5).

Telling species apart based on their DNA often assumes monophyly of the species in question. This would correspond to a taxon concordance score of 1 (where the most recent common ancestor of the species has no descendants belonging to other species). Based on this criterion of monophyly, the proportion of samples assigned to the right (concordant) taxon can give us an estimate of how easy it is to identify this species with the current marker set. A total of 29 species were found to be monophyletic, most of them supported by the majority of quartets in gene trees and posterior probabilities > 0.9 – including all species in *Orchis, Neotinea, Dactylorhiza, Gymnadenia* and *Anacamptis* (excluding three species for which only one sample was available and monophyly could not be assessed). The species that were paraphyletic, or had low support and/or widespread gene tree conflict, belonged to *Himantoglossum, Serapias* or *Ophrys*. The average taxon concordance was 0.82 when using a "splitter" classification, and 0.88 when using a "lumper" classification, showing that the splitting of larger groups into smaller units introduces more species that do not necessarily have higher scores. Consistent with this, 129 samples (87 %) could be assigned to corresponding concordant taxa in the "lumper" classification versus 105 samples (71 %) in the "splitter" classification. Some of these could theoretically be assigned multiple species identifications in the case of nested clades, such as *Serapias orientalis, Himantoglossum galilaeum*, and some *Ophrys* species, precluding definitive identification. With more than twice as many discordant taxa (concordance < 0.5), splitting results in 28 samples that belong to species without a concordant clade, as opposed to 13. Whereas only one sample falls outside of its concordant clade in the "lumper" classification (008_*Serapias orientalis*), seven do so in the "splitter" classification, belonging to multiple *Ophrys* and *Serapias* species. This shows that, based on a criterion of monophyly or minimal genetic distance to conspecific samples, identification of Orchidinae with the Angiosperms353 markers is still facing significant challenges especially in sections of *Ophrys*, *Serapias* and *Himantoglossum*, but could be relatively straightforward in the represented species from other genera.

3.3. Mixed evidence for splitter and lumper approaches

To shed light on the conflict between splitter and lumper approaches in classifying orchid taxa, we examined the available evidence in support of each, by conducting a Bayesian species delimitation analysis using SNAPP. This analysis targeted selected problematic clusters where taxonomic consensus is absent; the *Serapias* clade comprising of *S. vomeracea* and *S. bergonii*, as well as two groups within the *Ophrys* genus: the *O. lutea, O. fusca* complexes, on the one hand, and the *O. sphegodes, O. fuciflora, O. scolopax* complexes on the other. All path sampling analyses exhibited good mixing with effective sampling sizes (ESS) > 200. Comparison of marginal likelihood estimates (MLE) by means of Bayes Factor (BF) strongly support a classification where *Serapias bergonii* (which is sometimes considered a subspecies of *S. vomeracea*), is recognised as a distinct species (Table 2; Appendix S6A). Similarly, we found more support for a narrow classification of the *O. fusca* and *O. lutea* complexes, recognising *O. calocaerina, and O. cinereophila* (*fusca* s.l.) on one hand, and *O. lutea* s.s., *O. sicula,* and *O. melena* (*lutea* s.l.) on the other hand as separate lineages (Table 2; Appendix S6B). In contrast, we found less support for splitting up the *O. sphegodes, O. fuciflora, O. scolopax* complexes, with the evidence strongly in favour of broader species definitions (Table 2; Appendix S6C). The maximum clade credibility trees of each path sampling analysis (Appendix S6D-F) are in broad agreement with the constrained

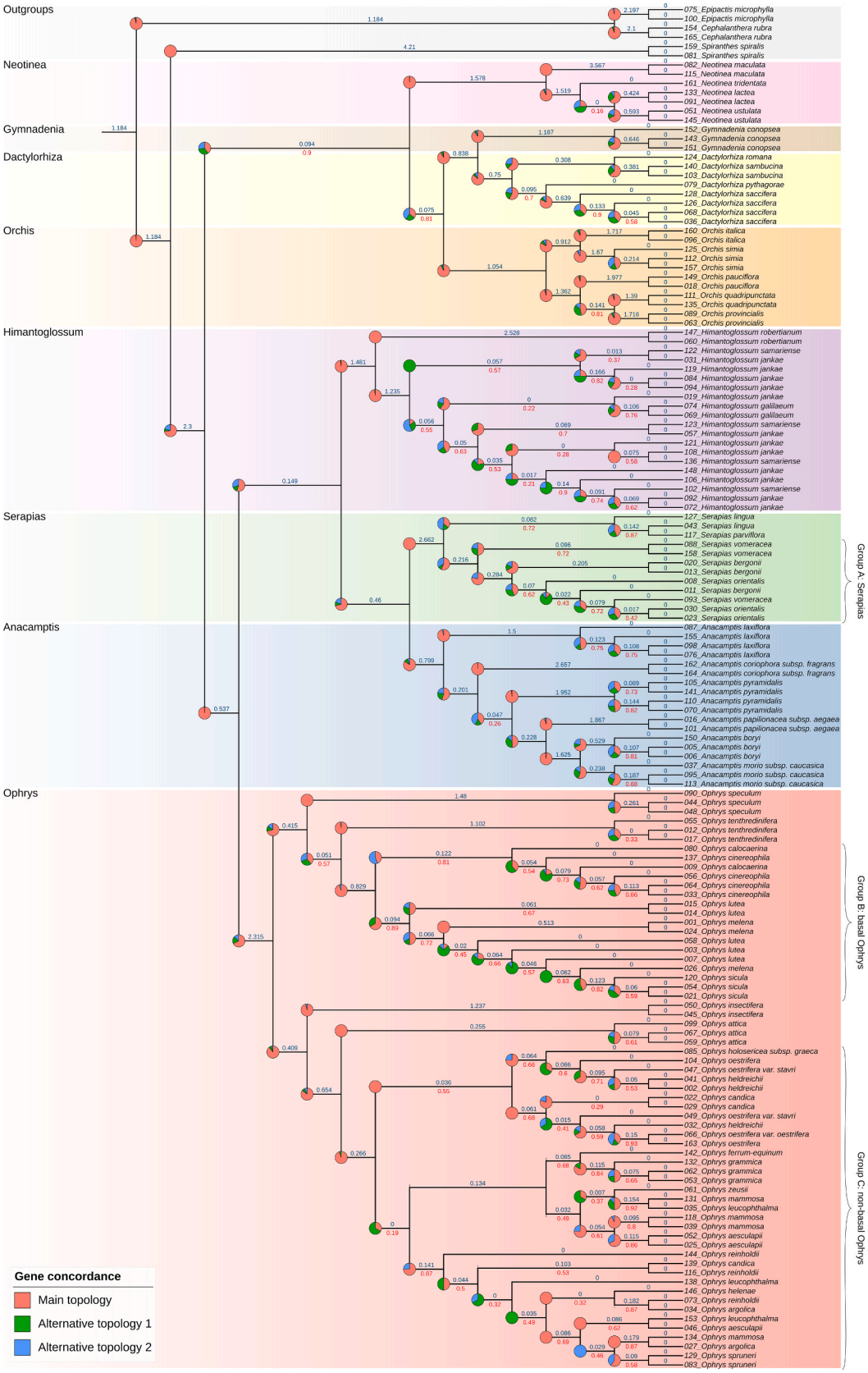

Fig. 2. Multispecies coalescent phylogeny of 148 orchid samples included in this study, constructed on the basis of 335 gene trees with ASTRAL-III. Posterior Probabilities < 0.9 and branch lengths for each internal branch are indicated with numbers at mid-branch length. The three clades for which competing species concepts were evaluated with Bayes Factor Delimitation (BFD) are indicated with curly brackets. Dashed lines highlight the clades in the tree that belong to these three clades. Pie charts display the proportion of quartets found in gene trees that support either the main topology in the tree (concordant), or one of two alternative topologies.

Table 2
Bayesian Factor Delimitation (BFD) between splitter and lumper classifications of samples with competing species concepts. Relative strength of support for either classification was assessed by calculating Bayes Factor (BF) from their Marginal Likelihood Estimates (MLE), and evaluating BF based on the framework of Kass and Raftery (1995).

Species model: Clade	Splitter MLE	ESS	Lumper MLE	ESS	Evaluation BF	Preferred	Support
Serapias	−1979.6	424.7	−1993.5	413.4	−27.8	Splitter	Decisive
Ophrys (basal)	−1778.6	392.6	−1782.0	446.6	−6.6	Splitter	Strong
Ophrys (non-basal)	−1439.7	283.8	−1432.4	407.3	14.6	Lumper	Decisive

multispecies coalescent trees generated with ASTRAL-III (Fig. 3), with the exception of *O. ferrum-equinum*; which was placed as sister to *O. sphegodes* by SNAPP (Appendix S6F), but as sister to a larger clade comprising *O. sphegodes*, *O. reinholdii* and *O. argolica* (all considered by some to belong to the *O. sphegodes* macrospecies, see Bateman et al., 2018; Betrand et al., 2021) by ASTRAL-III (Fig. 3).

3.4. Selected markers can reproduce most species relationships that are robust between methods

To evaluate the power of our dataset to resolve phylogenetic relationships across different timescales and taxonomic levels, we analysed the phylogenetic informativeness (PI) of 177 nuclear loci based on a time-calibrated maximum likelihood (ML) tree of species whose relationships were congruent between our different phylogenetic analyses (Fig. 4A). By integrating the PI between the boundaries of different time periods, we identified the highest-ranking loci as candidate loci with the largest potential for resolving problematic nodes falling within these time intervals (Fig. 4B). The number of loci excluded from consideration in each epoch because a more recent peak in PI could confound the signal, was relatively low in the first three epochs and rose steeply for the fourth epoch (Appendix S7). In the fifth and oldest epoch, no candidate loci were identified. For the remaining four epochs, we ranked the top 10 loci with the highest area under the curve (AUC) after excluding misleading loci, proposing a total set of 30 loci that, out of all genes enriched by the Angiosperms353 baits, are most likely to be informative for characterising the species diversity of Mediterranean tuberous orchids (Fig. 4C; Appendix S8). These loci are characterised by longer alignment lengths, more variable sites and more parsimony informative sites than those that are not included in the top 10 (Fig. 4D). While the total AUC per locus correlated strongly with alignment length ($r = 0.87$) and the absolute number of variable ($r = 0.93$) and parsimony informative sites ($r = 0.90$), the correlation with the number of taxa and percentage of missing data was lower, and there was no correlation at all with base composition (Appendix S9).

To demonstrate the effectiveness of these loci in recovering relationships among 42 selected Orchidinae species, we calculated quartet scores and normalised Robinson-Foulds (RF) distances (Smith 2020) of the set of 30 candidate loci and 100 other sets of 30 randomly chosen loci, relative to the species relationships in the tree based on the full set of genes. The candidate loci ranked in the 2nd or 1st percentile, respectively, for both methods, with a quartet score approaching that of the full tree (0.82) and an RF distance of 0.10. Topological comparisons between the 30-locus tree and the 327-locus tree (where 8 out of 335 loci were removed because they contained insufficient taxa) revealed minor differences, switching the positions of *Neotinea ustulata* and *N. tridentata* and those of *Ophrys tenthredinifera* and *O. speculum*, and placing *Ophrys mammosa* as a sister of *O. scolopax* instead of *O. zeusii* (Appendix S10). In these cases, despite displaying a stable topology between the original MSC and ML analyses, the samples all descend from nodes that are characterised by marked gene-tree discordance, hinting that, even if the species tree is largely approximated by the candidate loci, the specific

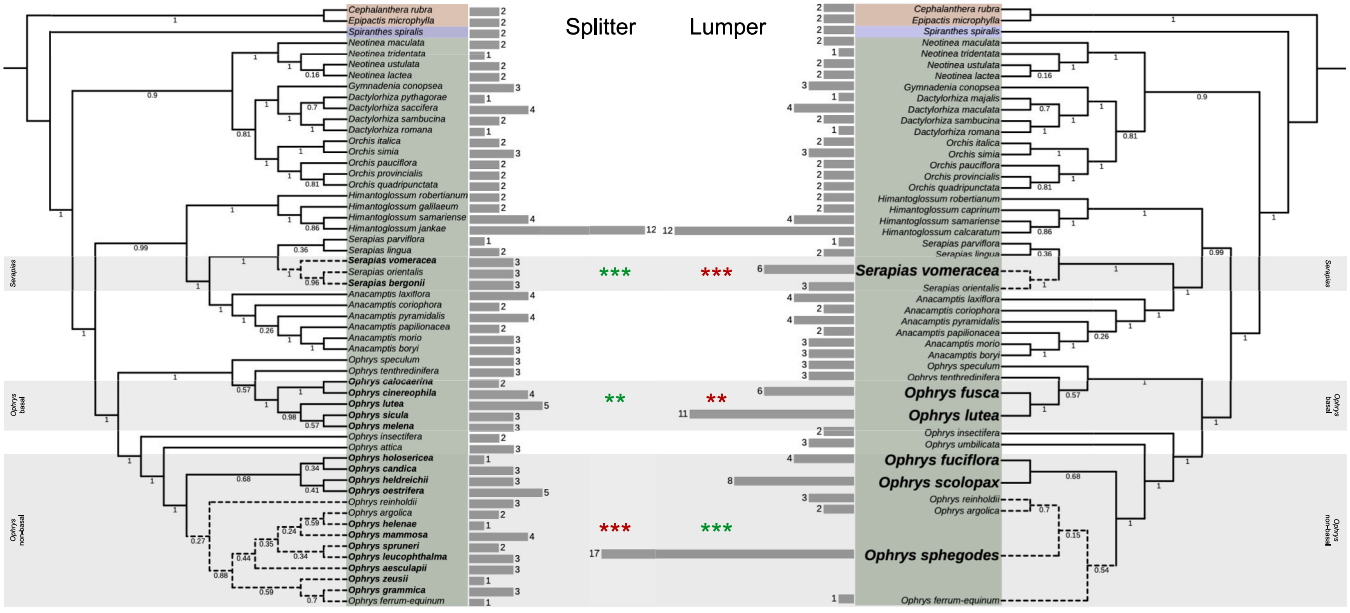

Fig. 3. Species tree constrained by splitter (left) and lumper (right) sample to species assignments. Generated from 148 samples and 335 gene trees under the multispecies coalescent with ASTRAL-III. Species in bold are those with conflicting splitter and lumper taxonomies. Lumper species are indicated with an increased font size and are aligned with the splitter species of which they are composed (left). Coloured asterisks indicate which scenario was preferred based on the Maximum Likelihood Estimate (MLE), green being the preferred model with a higher MLE and red the rejected model with a lower MLE value. The number of asterisks refers to the strength of support for the best scoring model based on the Bayes factor: one for moderate support, two for strong support and three for decisive support, as described by Kass and Raftery, (1995). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

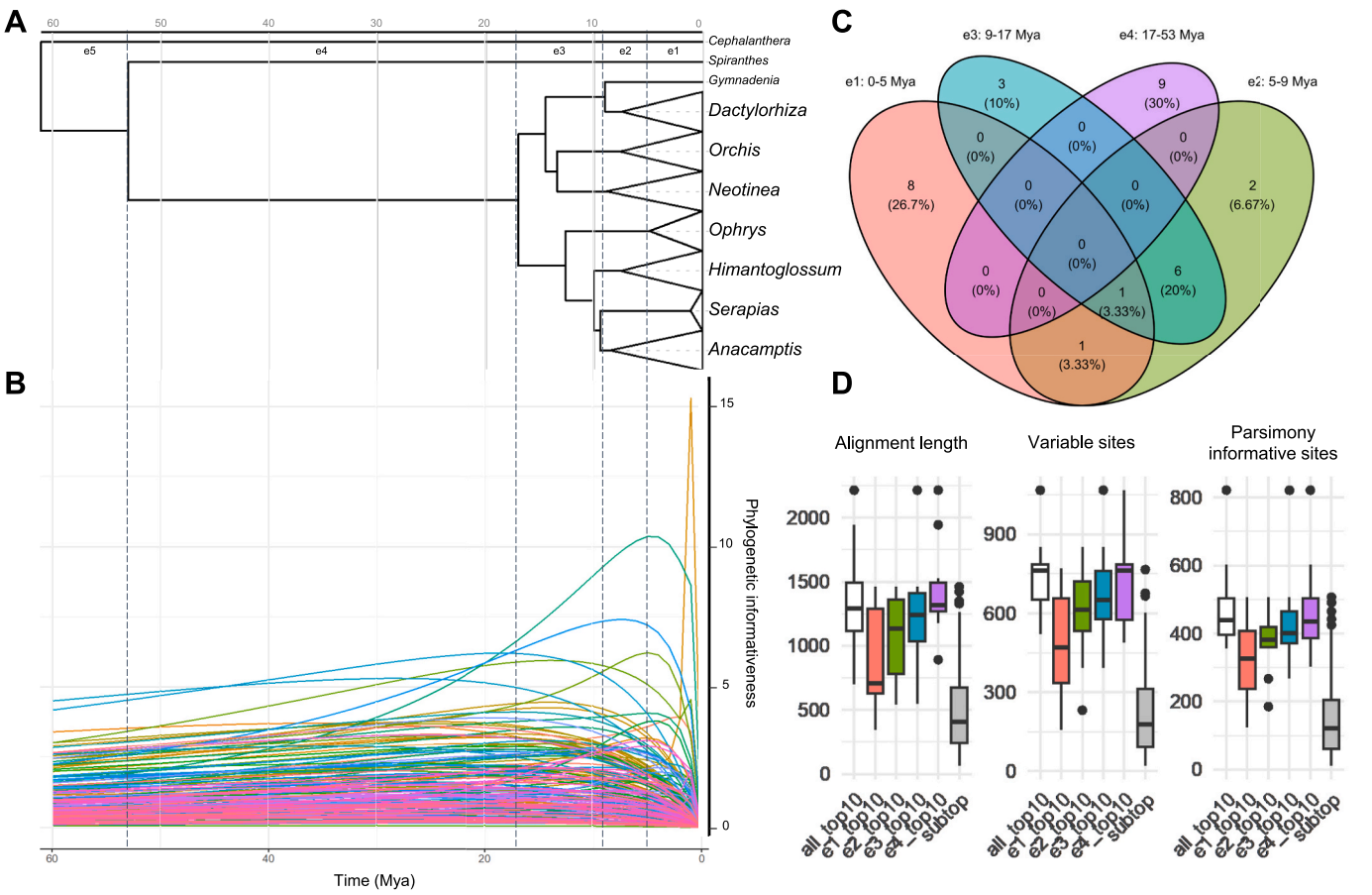

Fig. 4. **A.** Calibrated maximum likelihood tree of 42 samples. Nodes are collapsed at the genus level. **B.** Phylogenetic Informativeness (PI) profiles for 177 loci with at least four taxa, rooted on Epidendroideae. Vertical dashed lines indicate the boundaries between different epochs of orchid evolution. **C.** The top 10 most informative loci for each epoch and their overlap. Loci were ranked by Area Under the Curve (AUC) between epoch boundaries; to minimise the risk of homoplasy obscuring the signal of informativeness in older epochs, loci with a maximum PI younger than the median age of the epoch were excluded for that epoch, reducing overlap between the sets and increasing the total set of candidate loci from 20 to 30. **D.** Box plots illustrating the distribution of alignment length, number of variable sites, and number of parsimony informative sites across the top 10 ranking loci for all epochs collectively, for each epoch individually and for loci that did not rank among the top 10.

markers chosen might influence inferred species relationships in areas of the phylogeny where gene tree conflict is widespread (Fig. 2).

4. Discussion

The Angiosperms353 probe set has opened up new possibilities to chart existing diversity within and relationships among orchid taxa. This study has applied this probe set to a wide variety of genera within the orchid family subtribe Orchidinae s.s. native to Greece, based on a dense taxon sampling including multiple representatives within species complexes. In doing so, this study has generated valuable data for previously under sampled genera (Pérez-Escobar et al., 2021, 2024), offering several new insights into Mediterranean terrestrial orchid relationships and proposing a reduced set of markers for species identification.

4.1. Target recovery

At a sequencing depth of ~ 1 Gb on average per sample (recommended by Daicel Arbor Biosciences as the standard coverage for libraries enriched with Angiosperms353 baits), our target recovery falls within the expected range of the targeted 266 kb total gene space and is consistent with Angiosperms353 locus recovery for orchids reported in other studies (Pérez-Escobar et al., 2021, 2024). An increase in sequencing depth is therefore unlikely to significantly enhance gene recovery, without modifying either the hybridising probes or the reference file with less divergent sequences. Notably, our average target recovery does exceed the gene lengths documented for orchid species by Baker et al. (2022). This discrepancy can most likely be attributed to the incorporation of taxonomically pertinent reference sequences derived from the 1KP orchid transcriptomes. The shorter evolutionary distances of the species in our dataset to these reference sequences as opposed to the original Angiosperms353 reference file, render them more effective during target assembly, despite only a minimal increase in target length. We focused our phylogenetic analyses on exon sequences and excluded the non-coding sequences. Although introns are beneficial for exploring recent evolutionary dynamics within species and among closely related species, their utility is limited in broader phylogenomic studies because of the complexities and potential for misinterpretation when aligning and analysing these sequences across evolutionarily distant taxa. Nevertheless, the intron data accumulated in our study holds significant potential for future investigations that could zoom in on recently diverged lineages within the orchid family.

4.2. Genus-level relationships

Our phylogenetic analyses (Fig. 2) are characterised by their high level of node support for genus-level relationships compared to previous studies (Bateman et al., 2003; Inda et al., 2012; Jin et al., 2017), synthesising earlier findings that were sometimes contradictory. The contested placement of *Orchis* and *Neotinea*, two genera that diverged from other lineages relatively early in the evolution of the tribe, remains uncertain; with multi-species coalescent analysis grouping *Orchis* and

Neotinea as successive sister clades to *Dactylorhiza* and *Gymnadenia*, while maximum likelihood analysis groups them as sisters to each other, forming a clade that is itself a sister to *Gymnadenia* and *Dactylorhiza*. The multi-species coalescent tree lends additional credibility to the notion that *Orchis*, *Dactylorhiza*, and *Gymnadenia* collectively form a clade, as described by Bateman et al. (2003) based on internal transcribed spacer (ITS) nuclear DNA. However, the ASTRAL-III tree based on the custom Orchidinae-205 kit positioned *Orchis* as a sister clade to a group containing *Gymnadenia*, *Dactylorhiza*, and *Neotinea*, underscoring the unresolved placements of *Orchis* and *Neotinea* (Veltman et al. 2024). It further highlights that *Ophrys* diverged from the *Anacamptis/Serapias* lineage before *Himantoglossum* did, echoing findings from the multi-marker studies of Inda et al. (2012) and Jin et al. (2017), but contradicting the ITS phylogeny of Bateman et al. (2003) and more recent results generated with the Angiosperms353 markers by Baker et al. (2021). The latter offers only weak support for the alternative where *Ophrys* is a sister to *Anacamptis* and *Serapias*, instead of *Himantoglossum,* a relationship also corroborated by Veltman et al. (2024). Our MSC tree also places *Neotinea* as an outgroup to this clade, corroborating results from Bateman et al. (2003), Inda et al. (2012), and Baker et al. (2021), while conflicting our own ML topology and the multi-marker phylogeny of Jin et al. (2017). Combined, our analyses suggest the existence of two broad clades: one composed of *Anacamptis*, *Serapias*, *Ophrys* and *Himantoglossum*, and one composed of *Dactylorhiza*, *Gymnadenia*, *Orchis* and *Neotinea*. In contrast, Veltman et al. (2024) proposed three distinct clades, agreeing on the first clade but separating *Dactylorhiza, Gymnadenia* and *Neotinea* into a distinct group while placing *Orchis* at the base of the ingroup with maximum support.

4.3. Species-level relationships

Within the *Himantoglossum* clade, *H. robertianum* is consistently a sister group to the other *Himantoglossum* species, as found in previous barcoding studies (Bateman et al. 2003; Inda et al., 2012; Jin et al., 2017) as well as in a plastid phylogeny of *Himantoglossum* (Sramkó et al., 2014) and nuclear target capture data (Baker et al., 2022). Based on our tree topology, it appears that *H. jankae* is not monophyletic, with *H.* x *samariense* and *H. galilaeum* embedded within it (Fig. 2). Literature suggests that *H.* x *samariense* may be a hybrid with *H. galilaeum* (synonym of *H. caprinum* (M.Bieb.) Spreng.) and *H. jankae* (synonym of *H. calcaratum* subsp. *rumelicum* (H.Baumann & R.Lorenz) Niketić & Djordjevic) as parent species (Alibertis and Alibertis, 1989), explaining that the three taxa together form a clade. It is worth noting that from northern to southern Greece there is a gradual transition from individuals of *H. jankae* with long and spotted lip (typical *H. jankae* individuals) to individuals with shorter and unspotted lips (rather identical to those described as *H.* x *samariense*) (Antonopoulos and Tsiftsis, 2017; Tsiftsis and Antonopoulos, 2017; Tsiftsis, 2016). Moreover, blurred boundaries between *H.* x *samariense* and *H. jankae* are supported by Sramkó et al. (2014), who also noted the ambiguous relationship between the two species and the tendency of hybridisation among their close relatives. The polyphyly of *H.* x *samariense* and *H. jankae*, with *H. galilaeum* nested within this clade, along with the significant continuous morphological variability observed in *Himantoglossum* populations in Greece, suggests that they should best be treated as a single entity (Tsiftsis, 2016).

Gymnadenia and *Dactylorhiza* are also documented to hybridise both intergenerically and interspecifically (Bateman and Rudall, 2018; Bateman et al., 2018). Unfortunately, our dataset did not contain an adequate number of species samples from these genera to investigate the extent of hybridisation. Within *Serapias* we observed a similar pattern as observed by Inda et al. (2012) and Veltman et al. (2024), with two distinctive clades: one including *S. lingua* and *S. parviflora*, and one including *S. orientalis* and *S. vomeracea*, where *S. orientalis* is nested within the *S. vomeracea/S. bergonii* complex. The paraphyletic relationships between *S. vomeracea, S. bergonii* and *S. orientalis* could be explained by the extremely recent radiation of the genus (<1 Mya) and consequently high levels of incomplete lineage sorting.

4.4. Rapid radiation in Ophrys

Despite widespread uncertainty afflicting many species relationships within *Ophrys,* our findings broadly align with the taxonomic delineation of various *Ophrys* "macrospecies" or flagship species as defined by Bateman (2018), Bateman et al. (2021) and Bertrand et al. (2021), including the *O. lutea*, *O. fusca*, *O. fuciflora*, *O. scolopax* and *O. sphegodes* complexes, consistent with the findings of Veltman et al. (2024). Among these macrospecies, our topology rejects those proposed by several earlier barcoding studies (which often lacked sufficient phylogenetic resolution to confidently identify macrospecies relationships), but is in broad agreement with that of a recently published RAD-seq study (Bateman et al., 2018). Specifically, our results confirm *O. speculum* as a sister group to *O. tenthredinifera* and *O. fusca* (we did not sample *O. bombilyflora*), which was also observed by Breitkopf et al. (2015), and places *O. insectifera* as a sister species to the rest. While previous barcoding studies (Devey et al., 2008) suggested that *O. insectifera* is the earliest diverging lineage within the *Ophrys* genus, this widely held assumption was called into question by upscaling the amount of sequencing data by RAD-seq (Bateman et al., 2018) and whole plastome sequencing (Bertrand et al., 2021). Similar to another target capture study based on loci tailored to the Orchidinae tribe (Veltman et al., 2024), our results also support the inner placement of *O. insectifera* between other "basal" and "non-basal" *Ophrys* species. Additionally, our data support a sister relationship between the macrospecies *O. fuciflora* (sometimes considered to include the *O. scolopax* macrospecies) and *O. sphegodes* (in the definition of Bateman (2018) and Bateman et al. (2021) considered to include *O. argolica*, *O. reinholdii,* and *O. ferrum-equinum*], with *O. umbilicata* Desf. as their next closest relative, although we remain unable to distinguish some of the meso- and microspecies in our defined lumper and splitter classifications. In contrast, Veltman et al. (2024) found *O. argolica, O. reinholdii* and *O. ferrum-equinum* to form a distinct clade as sister to *O. sphegodes*, but as these results exhibited low support, evidence for relationships among the non-basal lineages of *Ophrys* remains inconclusive.

The lack of resolution within *Ophrys* macrospecies complexes, especially among the splitter species of the *O. fuciflora* and *O. sphegodes* groups, can be partly explained by their recent divergence and associated patterns of incomplete lineage sorting as well as limited sequence variation. The rapid diversification observed within the genus is thought to be driven by co-evolution with an adaptation to different pollinator species (Baguette et al., 2020). Given the explosive burst in speciation, the systematics and taxonomy of *Ophrys* remain subject of ongoing debate and have yet to reach a consensus (Bertrand et al., 2021). The complexities of unravelling this radiation are compounded by extensive interspecies hybridisation, which is facilitated by shared pollinators and overlapping flowering times (Stökl et al., 2009). It has been suggested that rather than being adaptively radiated, these readily hybridising groups of species correspond to "metapopulations", which cannot escape gene flow long enough to result in true speciation (Bateman, 2018; Bateman et al., 2021). Yet, genomic regions that contain clusters of genes involved in reproductive isolation and adaptation, or "speciation islands", could drive speciation despite ongoing gene flow (Wolf and Ellegren, 2017). Understanding the role of such speciation islands could thus help to delineate species and detect incipient speciation. In this respect, the genetic basis of floral morphological traits and scents that influence plant-pollinator interactions can offer key insights into this process (e.g. Sedeek et al., 2014; Russo et al., 2024) since these traits and scents are critical in facilitating floral isolation, which is often necessary for speciation in outcrossing plants relying on pollinators like *Ophrys*.

4.5. Species as hypotheses

Although especially apparent in the circumscription of *Ophrys* species, the criteria for assigning a taxon to the rank of species is a persistent area of debate within orchid systematics and taxonomy at large, with implications for species identification in applied domains such as conservation and wildlife trade. As described above, the delineation of species boundaries is a complex issue that hinges on the speciation processes considered and the species concepts adopted. The discord between splitters, who favour finer distinctions between species, and lumpers, who advocate for broader species definitions, complicates this taxonomic issue (Endersby, 2009). Genetic evidence casts new light on these distinctions and, as more data becomes available, refines our understanding of which morphologically distinct groups are genetically coherent. It is possible that the addition of new markers or even genome-wide data might tip the analysis in favour of splitting up taxa which are currently better lumped. However, our species delimitation results suggest that advocating for either a splitter or a lumper strategy as universally superior to the other, is neither desirable nor supported by statistical evidence. The effectiveness of a classification approach can vary significantly depending on the taxa under consideration and the data available. Our findings thus advocate for treating splitter and lumper classifications as distinct hypotheses for each taxon, to be examined on a case-by-case basis, and subject to repeated scrutiny as more data is added. Although the resource demands of Bayesian analysis may limit its application in large-scale phylogenomic studies, we recommend a clade-specific approach to comparing and evaluating competing classification models. This method allows for the independent assessment of species complexes, ensuring fundamentally that taxonomic decisions are informed by the most relevant and statistically robust evidence available, and ultimately that conservation and wildlife trade policies will rely on the most salient and up to date understanding of species identities.

The species concepts favoured by species delimitation are, however, not necessarily more concordant in the phylogeny containing all 148 samples, showing that – despite being considered more likely based on a full Bayesian analysis of the underlying sequence variation of selected lineages – multi-species coalescent and likelihood approaches based on gene trees or concatenated alignments can lead to different conclusions on which species are supported by the data. Discordant taxa (with low monophyly) tended to be confined to clades with low posterior probabilities and widespread gene tree conflict on internal nodes, and were encountered more frequently in the splitter classification. This could be remedied by increasing sampling and the amount of available sequencing data, potentially bringing taxon concordance scores in closer agreement with the species delimitation results. Nonetheless, some paraphyletic species will likely continue to defy clear demarcation and continued taxonomic work will be needed to revise the nomenclature of these orchids and clarify their species boundaries before confident species identification is possible. Until then, the identification for these species might only be possible at the level of genera or sections.

4.6. Markers for species identification

The phylogenetic informativeness of markers used to distinguish between species can vary due to different substitution rates, with some markers (in some lineages) evolving faster than others. Among the Angiosperms353 loci, we found that integrated PI over the age of the Orchideae tribe correlates strongly with alignment length, and the absolute number of variable and parsimony informative sites. This suggests that, rather than evolutionary rates of these genes, phylogenetic informativeness is indicative of their relative target recovery and that a higher coverage of some of the less informative markers could enhance their informativeness. Surprisingly, cumulative informativeness did not correlate strongly with the proportion of missing data. Our results therefore caution against excessive alignment trimming based on missingness, favouring the retention of informative sites over discarding those with high degrees of missing data. While the breadth of target recovery is inherently constrained by the length of sequences used as a reference and for probe design, there is conceivably still space for improvement within these limits, as higher and more consistent enrichment success could increase the amount of available data. In this respect, the universal probe design for the Angiosperms353 markers, where the identity between the sequences targeted for amplification and the reference sequences used for probe design can sometimes be as low as 70 % (Baker et al., 2021), can give a biased view of the informativeness of these markers, which could in theory be higher if they were amplified with less divergent baits. This bias should, however, affect all loci in the design equally and is not expected to influence our ranking. In fact, their relative ranking in terms of phylogenetic similarity to the full dataset relative to other, randomly subsampled loci, confirms their status as suitable candidates for reconstructing species relationships within the subtribe and potential to identify unknown samples. This potential could be maximised by designing custom primers tailored to the target species Orchidinae s.s., although this might not be able to compensate for the – sometimes large – phylogenetic uncertainty characterising some species relationships in the current dataset.

Identification of a species is only as good as the reference data to which it is compared. In that sense, missing taxa pose a problem for accurate species identification, since these may influence inferred species boundaries and relationships. Two genera and 25 species of the Orchidinae subtribe occurring in Greece were not included in the present phylogeny, limiting our ability to draw conclusions for all species. However, very few species of the unsampled genera (*Pseudorchis* and *Platanthera*) are native in Greece (1 and 2 species, respectively), and most of the unsampled species from other genera are narrow endemics. Even though the results presented here show that species assignment of the majority of sampled species is possible and that phylogenetic inference is remarkably robust to downsizing of the number of loci used for phylogenetic analyses, a more comprehensive sampling design will be useful to test the universality of these findings across clades and regional contexts, and at different geographic scales.

4.7. Universal vs custom probe sets

Universal and custom probe sets have both their advantages and disadvantages (Ufimov et al., 2021). Designed to capture a standardised set of loci across a wide array of angiosperms, the Angiosperms353 probe kit provides significant benefits for broad phylogenetic studies compared to custom bait kits such as cost-effectiveness and compatibility across diverse taxa, generating of extensive, comparable datasets that support large-scale evolutionary research (Johnson et al., 2019; Pérez-Escobar et al., 2021). The Angiosperms353 markers have effectively resolved phylogenies of rapidly evolving groups (Larridon et al., 2019), ranging from deep to shallow scales (Johnson et al., 2019) and even within-species relationships (Ufimov et al., 2021; Van Andel et al., 2019). Nevertheless, the unique phylogenetic challenges posed by orchids, particularly those in the *Ophrys* genus with high ecological diversity and extensive adaptive radiation, may not all be met by a universal probe kit. In cases where speciation histories are not captured by the Angiosperms353 loci, custom probe kits tailored to specific orchid groups, such as Orchidinae205, could yield deeper insights. These custom probes often have better enrichment success, resulting in broader coverage and less missing data (Veltman et al., 2024).

We applied the universal Angiosperms353 probe kit on specific orchid clades and encountered substantial challenges including variable and at times low target recovery and significant data gaps, highlighting the limitations of employing a universal probe kit for specialised taxa. Despite these hurdles, we successfully resolved previously contentious genus and species relationships, casting new light on prevailing taxonomic consensus and debates. We also identified 30 loci within the Angiosperms353 kit that are highly phylogenetically informative for this

group of Greek orchids. These loci have the potential to improve resolution in Eastern-Mediterranean orchid diversity, and could form the basis for tailored marker development where capture and target recovery of a subset of promising Angiosperms353 loci is optimised by custom baits with higher hybridisation efficiency. By providing better data quality at reduced costs, custom hybridisation and enrichment of these candidate loci could help improve species identification, species delimitation and conservation efforts of Mediterranean tuberous orchids. These baits could be merged with other custom baits that exist for the tribe (Veltman et al. 2024), to further enhance the resolution of species-level relationships in problematic clades such as *Ophrys*.

5. Conclusion and future directions

In this study, we present an updated phylogenetic tree for Mediterranean tuberous orchids using Angiosperms353 target capture data. Our analysis provides novel insights, notably into the *Himantoglossum*, *Serapias*, and *Ophrys* genera. Relationships between many *Ophrys* species are characterised by ambiguity due to rapid and recent diversification, as indicated by our divergence time estimation. Our results suggest that additional genetic loci beyond those examined might be necessary to resolve these clades, particularly within the *O. sphegodes* group. Our species delimitation analyses further address the ongoing debate between "splitters" and "lumpers", suggesting that neither approach is universally superior. Instead, we propose treating species classifications as distinct hypotheses to be evaluated on a case-by-case basis. Additionally, we identified 30 loci that are highly phylogenetically informative across four epochs of orchid evolution that are suitable for characterising the species diversity of Mediterranean tuberous orchids. Combined, our findings provide valuable phylogenomic data for Mediterranean tuberous orchids, a group that has been underrepresented in previous orchid phylogenomic studies using the Angiosperms353 probes. This contributes to the broader efforts of the Plant Tree of Life (Zuntini et al., 2024) By applying these probes to over 50 native orchid species in Greece, our research not only enhances the existing phylogenetic framework but also encourages continued development of tailored markers for routine identification and conservation of these vulnerable species.

Our study made a first attempt to use Angiosperms353 markers in the identification of Eastern Mediterranean orchid species and at settling some of the uncertainty arising from competing lumper and splitter classifications of Mediterranean orchids. However, our sampling design prohibited comprehensive assessments of intraspecific genetic variation across the species range, and a detailed assessment of species boundaries. We expect population level sampling of Orchidinae to substantially aid the delimitation of species in clades where these boundaries remain unclear or could not be assessed, especially in potential hybrid zones. However, collecting representatives from all as yet unsampled species might be unattainable, given current permitting restrictions and low accessibility of some narrow-ranged species in the field. Where these challenges persist, samples of the rarer species may be obtained from herbarium specimens to fill these gaps.

Future research should also investigate the role of interspecific gene flow in clades currently characterised by large phylogenetic uncertainty, an understudied aspect of orchid evolution that is at the basis of some hybrid lineages and could explain the widespread gene tree conflict observed in certain clades (Brandrud et al., 2020). Analyses of population level diversity and of gene flow would aid our understanding of orchid systematics and greatly benefit from upscaling the number of (neutrally evolving) markers, highlighting the potential of intron sequences for more fine-grained evolutionary studies of the genetic variation found in individual genera or species complexes. On the other hand, routine identification of orchids in conservation and trade requires a broadly applicable phylogenomic framework that maximises accuracy while minimising costs. Future studies could combine data from multiple phylogenomic studies and bait sets to further narrow down the most promising markers for species identification of Mediterranean orchids, and determine the minimum required number for robust inference of their species relationships. This would accelerate the use of molecular identification for promoting orchid conservation and combating illegal orchid trade.

Funding information

This research is part of the H2020 MSCA-ITN-ETN Plant.ID network, and has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement No. 765000.

Data availability statement

Raw sequencing reads have been deposited in the NCBI Sequencing Read Archive (SRA) under BioProject ID PRJNA1167703. All other data (exon sequences, alignments and gene trees) are available on Dryad (https://doi.org/10.5061/dryad.qrfj6q5sb).

CRediT authorship contribution statement

Bastien Anthoons: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Margaretha A. Veltman:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Spyros Tsiftsis:** Writing – review & editing, Validation, Supervision, Conceptualization. **Barbara Gravendeel:** Writing – review & editing, Visualization, Validation, Supervision, Conceptualization. **Andreas D. Drouzas:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Conceptualization. **Hugo de Boer:** Writing – review & editing, Visualization, Validation, Supervision, Project administration. **Panagiotis Madesis:** Writing – review & editing, Validation, Supervision, Project administration, Conceptualization.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2025.108360.

Data availability

All sequencing reads are deposited in the sequencing read archive under BioProject ID number PRJNA1167703

References

Alibertis, C., Alibertis, A., 1989. La Crète n'a pas fini de livrer ses secrets aux Orchidophiles (2ème partie). L'orchidophile 20, 109–112.

Andermann, T., Torres Jiménez, M.F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J.L., Gustafsson, A.L.S., Kistler, L., Liberal, I.M., Oxelman, B., Bacon, C.D., Antonelli, A., 2019. A guide to carrying out a phylogenomic target sequence capture project. Front. Genet. 10, 1407. https://doi.org/10.3389/fgene.2019.01407.

Antonopoulos, Z., Tsiftsis, S., 2017. Atlas of the Greek Orchids, Vol II. Mediterraneo Editions.

Baguette, M., Bertrand, J.A.M., Stevens, V.M., Schatz, B., 2020. Why are there so many bee-orchid species? Adaptive radiation by intra-specific competition for mnesic pollinators. Biol. Rev. Camb. Philos. Soc. 95, 1630–1663. https://doi.org/10.1111/brv.12633.

Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers, T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B., Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., Forest, F., 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. Syst. Biol. 71, 301–319. https://doi.org/10.1093/sysbio/syab035.

Bateman, R.M., Hollingsworth, P.M., Preston, J., Yi-Bo, L.U.O., Pridgeon, A.M., Chase, M. W., 2003. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). Botan. J. Linn. Soc. 142, 1–40. https://doi.org/10.1046/j.1095-8339.2003.00157.x.

Bateman, R.M., Murphy, A.R.M., Hollingsworth, P.M., Hart, M.L., Denholm, I., Rudall, P. J., 2018a. Molecular and morphological phylogenetics of the digitate-tubered clade within subtribe Orchidinae s.s. (Orchidaceae: Orchideae). Kew Bull. 73, 54. https://doi.org/10.1007/s12225-018-9782-1.

Bateman, R.M., Rudall, P.J., Murphy, A.R.M., Cowan, R.S., Devey, D.S., Peréz-Escobar, O.A., 2021. Whole plastomes are not enough: phylogenomic and morphometric exploration at multiple demographic levels of the bee orchid clade Ophrys sect. Sphegodes. J. Exp. Bot. 72, 654–681. https://doi.org/10.1093/jxb/eraa467.

Bateman, R.M., Rudall, P.J., 2018. Clarified relationship between Dactylorhiza viridis and Dactylorhiza iberica renders obsolete the former genus Coeloglossum (Orchidaceae: Orchidinae). Kew Bull. 73, 4. https://doi.org/10.1007/s12225-017-9728-z.

Bateman, R.M., Sramkó, G., Paun, O., 2018b. Integrating restriction site-associated DNA sequencing (RAD-seq) with morphological cladistic analysis clarifies evolutionary relationships among major species groups of bee orchids. Ann. Bot. 121, 85–105. https://doi.org/10.1093/aob/mcx129.

Bateman, R.M., 2018. Two bees or not two bees? An overview of Ophrys systematics.

Beck, J.B., Markley, M.L., Zielke, M.G., Thomas, J.R., Hale, H.J., Williams, L.D., Johnson, M.G., 2021. Are Palmer's Elm-Leaf Goldenrod and the Smooth Elm-Leaf Goldenrod Real? The Angiosperms353 Kit Provides Within-Species Signal in *Solidago ulmifolia* s. l. Syst. Bot 46, 1107–1113. https://doi.org/10.1600/036364421X16370109698740.

Bellot, S., Mitchell, T.C., Schaefer, H., 2020. Phylogenetic informativeness analyses to clarify past diversification processes in Cucurbitaceae. Sci. Rep. 10, 488. https://doi.org/10.1038/s41598-019-57249-2.

Bertrand, J., Baguette, M., Joffard, N., Schatz, B., 2021. Challenges Inherent in the Systematics and Taxonomy of Genera that have Recently Experienced Explosive Radiation: The Case of Orchids of the Genus *Ophrys*, in: Grandcolas, P., Maurel, M. (Eds.), Systematics and the Exploration of Life. Wiley, pp. 113–134. Doi: 10.1002/9781119476870.ch6.

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

Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660.

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

Brandrud, M.K., Baar, J., Lorenzo, M.T., Athanasiadis, A., Bateman, R.M., Chase, M.W., Hedrén, M., Paun, O., 2020. Phylogenomic relationships of diploids and the origins of allotetraploids in dactylorhiza (orchidaceae). Syst. Biol. 69, 91–109. https://doi.org/10.1093/sysbio/syz035.

Breitkopf, H., Onstein, R.E., Cafasso, D., Schlüter, P.M., Cozzolino, S., 2015. Multiple shifts to different pollinators fuelled rapid diversification in sexually deceptive Ophrys orchids. New Phytol. 207, 377–389. https://doi.org/10.1111/nph.13219.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932. https://doi.org/10.1093/molbev/mss086.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Charitonidou, M., Kougioumoutzis, K., Halley, J.M., 2021. An Orchid in Retrograde: Climate-Driven Range Shift Patterns of Ophrys helenae in Greece. Plants 10. https://doi.org/10.3390/plants10030470.

Christenhusz, M.J.M., Byng, J.W., 2016. The number of known plants species in the world and its annual increase. Phytotaxa 261, 201. https://doi.org/10.11646/phytotaxa.261.3.1.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Devey, D.S., Bateman, R.M., Fay, M.F., Hawkins, J.A., 2008. Friends or relatives? Phylogenetics and species delimitation in the controversial European orchid genus Ophrys. Ann. Bot. 101, 385–402. https://doi.org/10.1093/aob/mcm299.

De Boer, H., Veldman, S., Otieno, J.N., Bytebier, B., Van Andel, T.R., Gravendeel, B., 2017. DNA barcoding approaches highlight conservation challenges in chikanda orchids in southern and southeastern Africa. Genome 60, 925–926.

Dodsworth, S., Pokorny, L., Johnson, M.G., Kim, J.T., Maurin, O., Wickett, N.J., Forest, F., Baker, W.J., 2019. Hyb-Seq for flowering plant systematics. Trends Plant Sci. 24, 887–891. https://doi.org/10.1016/j.tplants.2019.07.011.

Endersby, J., 2009. Lumpers and splitters: Darwin, Hooker, and the search for order. Science 326, 1496–1499. https://doi.org/10.1126/science.1165915.

Faircloth, B.C., Chang, J., Alfaro, M.E., 2012. TAPIR enables high-throughput estimation and comparison of phylogenetic informativeness using locus-specific substitution models. arXiv.

Ghorbani, A., Gravendeel, B., Selliah, S., Zarré, S., de Boer, H.J., 2017. DNA barcoding of tuberous Orchidoideae: a resource for identification of orchids used in Salep. Mol. Ecol. Resour. 17, 342–352.

Ghorbani, A., Zarre, S., Gravendeel, B., de Boer, H.J., 2014. Illegal wild collection and international trade of CITES-listed terrestrial orchid tubers in Iran. Traffic Bull. 26, 52–58.

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., Paton, A., 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. Sci. Data 8, 215. Doi: 10.1038/s41597-021-00997-6.

Gustafsson, A.L.S., Verola, C.F., Antonelli, A., 2010. Reassessing the temporal evolution of orchids with new fossils and a Bayesian relaxed clock, with implications for the diversification of the rare South American genus Hoffmannseggella (Orchidaceae: Epidendroideae). BMC Evol. Biol. 10, 177. Doi: 10.1186/1471-2148-10-177.

Hanlidou, E., Karousou, R., Kleftoyanni, V., Kokkini, S., 2004. The herbal market of Thessaloniki (N Greece) and its relation to the ethnobotanical tradition. J. Ethnopharmacol. 91, 281–299.

Hinsley, A., de Boer, H.J., Fay, M.F., Gale, S.W., Gardiner, L.M., Gunasekara, R.S., Kumar, P., Masters, S., Metusala, D., Roberts, D.L., Veldman, S., Wong, S., Phelps, J., 2018. A review of the trade in orchids and its implications for conservation. Botan. J. Linn. Soc. 186, 435–455. https://doi.org/10.1093/botlinnean/box083.

Hollingsworth, P.M., Li, D.-Z., van der Bank, M., Twyford, A.D., 2016. Telling plant species apart with DNA: from barcodes to genomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 371. https://doi.org/10.1098/rstb.2015.0338.

Inda, L.A., Pimentel, M., Chase, M.W., 2012. Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. Ann. Bot. 110, 71–90. https://doi.org/10.1093/aob/mcs083.

Inda, L.A., Pimentel, M., Chase, M.W., 2010. Contribution of mitochondrial *cox1* intron sequences to the phylogenetics of tribe Orchideae (Orchidaceae): Do the distribution and sequence of this intron in orchids also tell us something about its evolution? Taxon 59, 1053–1064. https://doi.org/10.1002/tax.594006.

Jin, W.-T., Schuiteman, A., Chase, M.W., Li, J.-W., Chung, S.-W., Hsu, T.-C., Jin, X.-H., 2017. Phylogenetics of subtribe Orchidinae s.l. (Orchidaceae; Orchidoideae) based on seven markers (plastid matK, psaB, rbcL, trnL-F, trnH-psba, and nuclear nrITS, Xdh): implications for generic delimitation. BMC Plant Biol. 17, 222. https://doi.org/10.1186/s12870-017-1160-x.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. C., Wickett, N.J., 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. 4. https://doi.org/10.3732/apps.1600016.

Johnson, M.G., Pokorny, L., Dodsworth, S., Botigué, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., Forest, F., Kim, J.T., Leebens-Mack, J.H., Leitch, I. J., Maurin, O., Soltis, D.E., Soltis, P.S., Wong, G.-K.-S., Baker, W.J., Wickett, N.J., 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. Syst. Biol. 68, 594–606. https://doi.org/10.1093/sysbio/syy086.

Junier, T., Zdobnov, E.M., 2010. The Newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics 26, 1669–1670. https://doi.org/10.1093/bioinformatics/btq243.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.

Kasparek, M., Grimm, U., 1999. European trade in Turkish Salep with special reference to Germany. Econ Bot 53, 396–406.

Kass, R.E., Raftery, A.E., 1995. Bayes Factors. J. Am. Stat. Assoc. 90, 773–795. https://doi.org/10.1080/01621459.1995.10476572.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinf. 15, 356. https://doi.org/10.1186/s12859-014-0356-4.

Kreziou, A., de Boer, H., Gravendeel, B., 2015. Harvesting of salep orchids in north-western Greece continues to threaten natural populations. Oryx 1–4.

Kühn, R., Pedersen, H.Æ., Cribb, P., 2019. Field guide to the orchids of Europe and the Mediterranean. Royal Botanic Gardens, Kew.

Larridon, I., Villaverde, T., Zuntini, A.R., Pokorny, L., Brewer, G.E., Epitawalage, N., Fairlie, I., Hahn, M., Kim, J., Maguilla, E., Maurin, O., Xanthos, M., Hipp, A.L., Forest, F., Baker, W.J., 2019. Tackling rapid radiations with targeted sequencing. Front. Plant Sci. 10, 1655. https://doi.org/10.3389/fpls.2019.01655.

Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63, 534–542. https://doi.org/10.1093/sysbio/syu018.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61, 727–744. https://doi.org/10.1093/sysbio/sys049.

Mai, U., Mirarab, S., 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genomics 19, 272. https://doi.org/10.1186/s12864-018-4620-2.

Manzanilla, V., Teixidor-Toneu, I., Martin, G.J., Hollingsworth, P.M., de Boer, H.J., Kool, A., 2022. Using target capture to address conservation challenges: population-level tracking of a globally-traded herbal medicine. Mol. Ecol. Resour 22, 212–224. https://doi.org/10.1111/1755-0998.13472.

Masters, S., Anthoons, B., Madesis, P.G., Saroja, S., Schermer, M., Gerritsen, W., Karahan, A., Verdoes, R., Schwallier van Andel, T., de, R., Boer, H., Gravendeel, B., 2022. Quantifying an online wildlife trade using a web crawler. Biodivers. Conserv. https://doi.org/10.1007/s10531-022-02367-z.

McCormick, M.K., Jacquemyn, H., 2014. What constrains the distribution of orchid populations? New Phytol. 202, 392–400. https://doi.org/10.1111/nph.12639.

McLay, T.G.B., Birch, J.L., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Joyce, E.M., Simpson, L., Schmidt-Lebuhn, A.N., Baker, W.J., Forest, F., Jackson, C.J., 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. Appl. Plant Sci. 9. https://doi.org/10.1002/aps3.11420.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Ortiz, E.M., 2018. Vcf2Phylip V1.5: Convert A Vcf Matrix Into Several Matrix Formats For Phylogenetic Analysis. Zenodo. https://doi.org/10.5281/zenodo.1257058.
Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb. Genom. 2, e000056. https://doi.org/10.1099/mgen.0.000056.
Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633.
Pérez-Escobar, O.A., Bogarín, D., Przelomska, N.A.S., Ackerman, J.D., Balbuena, J.A., Bellot, S., Bühlmann, R.P., Cabrera, B., Cano, J.A., Charitonidou, M., Chomicki, G., Clements, M.A., Cribb, P., Fernández, M., Flanagan, N.S., Gravendeel, B., Hágsater, E., Halley, J.M., Hu, A.-Q., Jaramillo, C., Antonelli, A., 2024. The origin and speciation of orchids. New Phytol. 242, 700–716. https://doi.org/10.1111/nph.19580.
Pérez-Escobar, O.A., Dodsworth, S., Bogarín, D., Bellot, S., Balbuena, J.A., Schley, R., Kikuchi, I., Morris, S.K., Epitawalage, N., Cowan, R., Maurin, O., Zuntini, A., Arias, T., Serna, A., Gravendeel, B., Torres, M.F., Nargar, K., Chomicki, G., Chase, M. W., Leitch, I.J., Baker, W.J., 2021. Hundreds of nuclear and plastid loci yield insights into orchid relationships. BioRxiv. https://doi.org/10.1101/2020.11.17.386508.
Pond, S.L.K., Frost, S.D.W., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679. https://doi.org/10.1093/bioinformatics/bti079.
Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., Korobeynikov, A., 2020. Using SPAdes de novo assembler. Curr. Protoc. Bioinformatics 70, e102.
Rabiee, M., Sayyari, E., Mirarab, S., 2019. Multi-allele species reconstruction using ASTRAL. Mol. Phylogenet. Evol. 130, 286–296. https://doi.org/10.1016/j.ympev.2018.10.033.
Revell, L.J., 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.
Roy, S.C., Moitra, K., De Sarker, D., 2017. Assessment of genetic diversity among four orchids based on ddRAD sequencing data for conservation purposes. Physiol. Mol. Biol. Plants 23, 169–183. https://doi.org/10.1007/s12298-016-0401-z.
Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S.C.K., Cronn, R.C., Dreyer, L.L., Suda, J., 2016. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African Oxalis (Oxalidaceae). Mol. Ecol. Resour 16, 1124–1135. https://doi.org/10.1111/1755-0998.12487.
Sedeek, K.E.M., Scopece, G., Staedler, Y.M., Schönenberger, J., Cozzolino, S., Schiestl, F. P., Schlüter, P.M., 2014. Genic rather than genome-wide differences between sexually deceptive Ophrys orchids with different pollinators. Mol. Ecol. 23, 6192–6205. https://doi.org/10.1111/mec.12992.
Shah, T., Schneider, J.V., Zizka, G., Maurin, O., Baker, W., Forest, F., Brewer, G.E., Savolainen, V., Darbyshire, I., Larridon, I., 2021. Joining forces in Ochnaceae phylogenomics: a tale of two targeted sequencing probe kits. Am. J. Bot. 108, 1201–1216. https://doi.org/10.1002/ajb2.1682.
Slater, G.S.C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinf. 6, 31. https://doi.org/10.1186/1471-2105-6-31.
Smith, M.R., 2020. Information theoretic generalized Robinson-Foulds metrics for comparing phylogenetic trees. Bioinformatics 36, 5007–5013. https://doi.org/10.1093/bioinformatics/btaa614.
Sramkó, G., Attila, M.V., Hawkins, J.A., Bateman, R.M., 2014. Molecular phylogeny and evolutionary history of the Eurasiatic orchid genus Himantoglossum s.l. (Orchidaceae). Ann. Bot. 114, 1609–1626. https://doi.org/10.1093/aob/mcu179.
Stange, M., Sánchez-Villagra, M.R., Salzburger, W., Matschiner, M., 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of Sea Catfishes (Ariidae) supports miocene closure of the panamanian isthmus. Syst. Biol. 67, 681–699. https://doi.org/10.1093/sysbio/syy006.
Stökl, J., Schlüter, P.M., Stuessy, T.F., Paulus, H.F., Fraberger, R., Erdmann, D., Schulz, C., Francke, W., Assum, G., Ayasse, M., 2009. Speciation in sexually deceptive orchids: pollinator-driven selection maintains discrete odour phenotypes in hybridizing species. Biol. J. Linn. Soc. 98, 439–451. https://doi.org/10.1111/j.1095-8312.2009.01279.x.
Tsiftsis, S., Antonopoulos, Z., 2017. Atlas of the Greek Orchids, Vol I. Mediterraneo Editions.
Tsiftsis, S., Karagiannakidou, V., Tsiripidis, I., 2007. The orchid flora of East Macedonia (NE Greece).
Tsiftsis, S., Tsiripidis, I., 2020. Temporal and spatial patterns of orchid species distribution in Greece: implications for conservation. Biodivers. Conserv. 29, 3461–3489. https://doi.org/10.1007/s10531-020-02035-0.
Tsiftsis, S., 2021. The role of Natura 2000 network in protecting the orchid flora of East Macedonia (NE Greece). Eur. J. Environ. Sci. 11, 71–78. https://doi.org/10.14712/23361964.2021.8.
Tsiftsis, S., 2016. Morphological variability of Himantoglossum s.s. (Orchidaceae) in Greece. Phytotaxa 245, 17. https://doi.org/10.11646/phytotaxa.245.1.2.
Ufimov, R., Zeisek, V., Píšová, S., Baker, W.J., Fér, T., van Loo, M., Dobeš, C., Schmickl, R., 2021. Relative performance of customized and universal probe sets in target enrichment: A case study in subtribe Malinae. Appl. Plant Sci. 9, e11442. https://doi.org/10.1002/aps3.11442.
Van Andel, T., Veltman, M.A., Bertin, A., Maat, H., Polime, T., Hille Ris Lambers, D., Tjoe Awie, J., De Boer, H., Manzanilla, V., 2019. Hidden rice diversity in the guianas. Front. Plant Sci. 10, 1161. Doi: 10.3389/fpls.2019.01161.
Veltman, M.A., Anthoons, B., Schrøder-Nielsen, A., Gravendeel, B., de Boer, H.J., 2024. Orchidinae-205: A new genome-wide custom bait set for studying the evolution, systematics, and trade of terrestrial orchids. Mol. Ecol. Resour e13986. https://doi.org/10.1111/1755-0998.13986.
Warghat, A.R., Bajpai, P.K., Srivastava, R.B., Chaurasia, O.P., Sood, H., 2013. Population genetic structure and conservation of small fragmented locations of Dactylorhiza hatagirea in Ladakh region of India. Sci. Hortic. 164, 448–454. https://doi.org/10.1016/j.scienta.2013.09.044.
Wolf, J.B.W., Ellegren, H., 2017. Making sense of genomic islands of differentiation in light of speciation. Nat. Rev. Genet. 18, 87–100. https://doi.org/10.1038/nrg.2016.133.
Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19, 153. https://doi.org/10.1186/s12859-018-2129-y.
Zhang, L., Huang, Y.-W., Huang, J.-L., Ya, J.-D., Zhe, M.-Q., Zeng, C.-X., Zhang, Z.-R., Zhang, S.-B., Li, D.-Z., Li, H.-T., Yang, J.-B., 2023. DNA barcoding of Cymbidium by genome skimming: Call for next-generation nuclear barcodes. Mol. Ecol. Resour 23, 424–439. https://doi.org/10.1111/1755-0998.13719.
Zuntini, A.R., Carruthers, T., Maurin, O., Bailey, P.C., Leempoel, K., Brewer, G.E., Epitawalage, N., Françoso, E., Gallego-Paramo, B., McGinnie, C., Negrão, R., Roy, S. R., Simpson, L., Toledo Romero, E., Barber, V.M.A., Botigué, L., Clarkson, J.J., Cowan, R.S., Dodsworth, S., Johnson, M.G., et al., 2024. Phylogenomics and the rise of the angiosperms. Nature 629, 843–850. https://doi.org/10.1038/s41586-024-07324-0.