Received: 15 June 2023 | Revised: 16 May 2024 | Accepted: 30 May 2024

DOI: 10.1111/1755-0998.13986

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Orchidinae-205: A new genome-wide custom bait set for studying the evolution, systematics, and trade of terrestrial orchids

Margaretha A. Veltman[1,2] | Bastien Anthoons[3] | Audun Schrøder-Nielsen[1] | Barbara Gravendeel[2,4] | Hugo J. de Boer[1]

[1]Natural History Museum, Oslo, Norway

[2]Naturalis Biodiversity Center, Leiden, Netherlands

[3]Center for Research and Technology Hellas, Thessaloniki, Greece

[4]Radboud Institute for Biological and Environmental Sciences, Radboud University, Nijmegen, Netherlands

Correspondence
Margaretha A. Veltman, Natural History Museum, Oslo, Norway.
Email: margret.veltman@nhm.uio.no

Funding information
HORIZON EUROPE Marie Sklodowska-Curie Actions, Grant/Award Number: 765000

Handling Editor: Alison Nazareno

Abstract

Terrestrial orchids are a group of genetically understudied, yet culturally and economically important plants. The Orchidinae tribe contains many species that produce edible tubers that are used for the production of traditional delicacies collectively called 'salep'. Overexploitation of wild orchids in the Eastern Mediterranean and Western Asia threatens to drive many of these species to extinction, but cost-effective tools for monitoring their trade are currently lacking. Here we present a custom bait kit for target enrichment and sequencing of 205 novel genetic markers that are tailored to phylogenomic applications in Orchidinae s.l. A subset of 31 markers capture genes putatively involved in the production of glucomannan, a water-soluble polysaccharide that gives salep its distinctive properties. We tested the kit on 73 taxa native to the area, demonstrating universally high locus recovery irrespective of species identity, that exceeds the total sequence length obtained with alternative kits currently available. Phylogenetic inference with concatenation and coalescent approaches was robust and showed high levels of support for most clades, including some which were previously unresolved. Resolution for hybridizing and recently radiated lineages remains difficult, but could be further improved by analysing multiple haplotypes and the non-exonic sequences captured by our kit, with the promise to shed new light on the evolution of enigmatic taxa with a complex speciation history. Offering a step-up from traditional barcoding and universal markers, the genome-wide custom loci targeted by Orchidinae-205 are a valuable new resource to study the evolution, systematics and trade of terrestrial orchids.

KEYWORDS
Orchidinae, phylogenomics, species identification, target capture, wildlife trade

1 | INTRODUCTION

Overharvesting is one of the main threats to native plant diversity (Maxwell et al., 2016). Global markets and increasing population size are fuelling the demand for plant-based products, increasing harmful levels of exploitation and trade of wild plants around the world. Despite conservation policies and treaties such as the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), which aim to control the trade of goods derived from vulnerable species, illegal trade of plants often goes undetected and unchecked (Margulies et al., 2019). Without adequate monitoring and enforcement, the market for wild plants thus presents a growing problem for wildlife conservation (Jahanbanifard et al., 2022).

With an estimated 25,000 species, orchids (Orchidaceae) are one of the largest plant families (Chase et al., 2015) and the only plant family with all its species listed in one of the three CITES appendices, representing the vast majority of species that are protected under this convention. Uses of wild-harvested orchids vary from ornamental to culinary and medicinal, and target species from all major subfamilies (Hinsley et al., 2017). Among these uses, edible orchids are an important, but often overlooked group. In addition to the fruits (e.g., *Vanilla* Plum. ex Mill.) and leaves (e.g., *Jumellea* Schltr.) of some species, the starchy tubers of a wide range of terrestrial orchids are harvested on multiple continents, chief among them salep, a popular delicacy in the Eastern Mediterranean region (Bulpitt, 2005).

Salep is made by boiling and drying the tubers, which are subsequently ground up to be used in powdered form in warm drinks and ice cream. It is sold either as strings of tubers or in pre-packaged powder form (Kasparek & Grimm, 1999). Previous studies have reported as many as 35 different species being harvested and sold as salep in Greece, Turkey and Iran (Ghorbani et al., 2017; Kasparek & Grimm, 1999; Kreziou et al., 2016), spanning multiple genera in the Orchidoideae subfamily's Orchideae tribe. This tribe consists of ~1500 species, with about 100 occurring in the region (Pridgeon et al., 2001, 2003). Increasing volumes of tubers being harvested and sold have been reported in Iran, amounting to dozens of tons (equivalent to millions of individual orchids), much of which is destined for export (Ghorbani et al., 2014).

Current salep harvesting pressure and practices are unsustainable (Kreziou et al., 2016), but monitoring and controlling its trade is hampered by morphological similarities between tubers and uncertainty regarding their species identity. Targeted conservation efforts, including species management plans, designated protected areas, and alternative production methods, would benefit from knowing which species are preferentially harvested and sold. Molecular methods of plant identification, such as DNA barcoding, have been proposed as an instrument to monitor trade, but this technique relies on a small set of markers that do not always carry enough phylogenetic resolution to tell apart closely related or recently diverged taxa (Hollingsworth et al., 2011). Especially for rapidly evolving lineages, including many tuberous orchids, broader genomic coverage is therefore needed (Hollingsworth et al., 2016).

Target capture is an increasingly popular method to obtain large amounts of DNA sequence information from hundreds or even thousands of markers that (depending on the chosen markers) can be applied across a wide taxonomic range, allowing phylogenetic resolution at both deep and shallow scales (Andermann et al., 2019). Despite the clear advantages it presents for phylogenomic studies, developing the baits needed for enrichment of the selected markers requires a substantial investment in terms of genomic resources, bioinformatic analyses and bait synthesis, limiting its application versus traditional barcoding. The release and reuse of bait kits targeting 'universal' loci such as Angiosperms-353 promises to offset this challenge (Dodsworth et al., 2019) and has been successfully applied in many groups of flowering plants (Baker et al., 2021). However, their enrichment efficiency may be lower and they are likely to harbour less sequence variation than bait kits that are tailored to the taxonomic group of interest (Kadlec et al., 2017; Yardeni et al., 2022).

In addition to being optimized for a specific taxonomic group, custom bait kits offer the opportunity to include loci that are relevant for certain biochemical pathways, phenotypes and other traits of interest (Jones & Good, 2016). Analysing functional variation in selected candidate loci may be useful for understanding the evolution of certain traits that dictate consumer preference (e.g., presence of bioactive compounds) or that are important from an ecological (e.g., floral scent) or conservation (e.g., drought resistance) perspective. The purported beneficial effects of salep are directly linked to its concentration of glucomannan, a complex polysaccharide that serves as a thickening agent and brings a gelatinous texture to drinks and foods (Kurt, 2021), melting stability to ice cream (Tekinşen & Güner, 2010), and a feeling of satiety to its consumers (Ece Tamer et al., 2006). Knowing which genes underlie a high glucomannan concentration and how they vary will therefore be useful to understand which species are preferentially harvested for salep and why.

To facilitate phylogenomic and functional genomic analysis of salep orchids and their relatives, we developed Orchidinae-205, a custom bait kit tailored to all members of the subtribe Orchidinae (s.l.). The markers were selected with 14 de novo assembled transcriptomes covering all major clades in the subtribe. In addition to 174 low-copy nuclear genes, the bait kit targets 31 candidate genes putatively involved in the glucomannan biosynthesis pathway, allowing for genetic comparisons with the one other plant species outside the Orchidaceae that offers a naturally high concentration of glucomannan in its underground tubers, *Amorphophallus konjac* K. Koch (Araceae).

To explore the efficacy of the kit, we tested it on a selection of 79 species of Orchidinae s.l. occurring in the Eastern Mediterranean region, representing 12 genera, including multiple species complexes with disputed phylogenetic placements and species boundaries, and many orchids that are a potential source for salep. We demonstrate the added value of the Orchidinae-205 loci in three ways. Firstly, we compare the target recovery of this kit with two alternative kits available at the time of this study (Angiosperms-353 and Orchidaceae-963), using existing data, and make an in-depth comparison of topological support and phylogenetic

informativeness afforded by the universal Angiosperms-353 and custom Orchidinae-205 kits. Secondly, we generate a comprehensive phylogeny of Mediterranean Orchidinae using both concatenation and coalescent-based approaches, yielding new insights into species relationships and sources of phylogenetic discordance. Thirdly, we test the hypothesis that genes putatively involved in the glucomannan pathway are under selection in (some of) our target species, by conducting a branch-site test of episodic diversifying selection.

In absence of published reference genomes for most of our focal species (but see Li et al., 2022; Russo et al., 2023; Wolfe et al., 2023 for recent additions), the set of tailored genome-wide markers presented here will facilitate comparative genomic studies in this group of terrestrial orchids, enabling the use of low input and degraded DNA such as found in historical collections and derived plants products. We therefore anticipate that Orchidinae-205 will be an important resource for future population genomic and phylogenomic studies in the Orchidinae, with diverse applications ranging from evolution and systematics, to wildlife forensics and conservation.

2 | MATERIALS AND METHODS

2.1 | Orchidinae-205 development

Baits were designed with 14 transcriptomes representing 8 genera, selected from different subclades within the Orchidinae s.l. for which publicly available data was available. RNA-seq data of 23 Orchidoideae species were downloaded from the NCBI Sequence Read Archive (Table S1). Raw reads were trimmed with Trimmomatic v0.39 (Bolger et al., 2014) prior to assembly with Trinity v2.10.0 (Grabherr et al., 2011). The assembled transcripts were then filtered to optimize orthology inference following the procedure of Yang and Smith (2014), with updated scripts by Morales-Briones et al. (2021). Orthogroups were detected with OrthoFinder v2.5.1 (Emms & Kelly, 2019) for the subfamily (Orchidoideae) and tribe (Orchidinae). The sequences of orthogroups that were strictly single copy and represented in at least one species per genus were mapped against a draft genome of *Ophrys sphegodes* Mill. (Osph-v1.1) of *Ophrys sphegodes* Mill. (Russo et al., 2023) with GMAP (Wu et al., 2016). Orthogroups with *O. sphegodes* transcripts that mapped exactly once, with a coverage of 100%, zero indels and a minimum length of 750 bp were selected as potential targets. Coding sequences of these orthogroups were aligned with the MAFFT v7.470 L-INS-i algorithm (Katoh & Standley, 2013) and *p*-distances were calculated with FastME v2.1.6.1 (Lefort et al., 2015). Orthogroups where the pairwise distance between any pair of target species did not exceed 0.1 were selected for probe development (single-copy targets) alongside several orthogroups with putative functions in glucomannan synthesis (glucomannan targets).

Glucomannan targets were identified based on reported candidate genes in the literature and their homology with the *Oryza sativa* subsp. *japonica* Nipponbare (IRGSP-1.0) genome (Kawahara et al., 2013). Orthology inference was repeated with these homologues included. The resulting orchid orthologues were also mapped against the *Ophrys sphegodes* genome and selected if all *Ophrys* transcripts mapped to exactly the same region, with a minimum coverage of 70%, pairwise identity of at least 90% and a minimum alignment length of 300 bp. Because these orthogroups were not all strictly single copy, glucomannan target alignments were split into groups of sequences that clustered together with an average pairwise distance of <0.1. Groups that contained more than six sequences representing at least three different genera were selected as final targets. The coding sequences of the selected single-copy targets and glucomannan targets were submitted to Daicel Arbor Biosciences (Ann Arbor, MI, USA) for bait development.

Baits were designed with a length of 70 bp and 3× tiling density, to optimize enrichment of degraded DNA. Baits with that would potentially enrich spurious sequences were identified by mapping against four orchid genomes, the chloroplast genomes of three target species and two mitochondrial genomes of non-orchid species in the Asparagales (for details, see Data S1). Baits with additional hits to either of these genomes or with >25% repeat masking were discarded. Surviving baits for single-copy targets were filtered with stringent BLAST settings to maximize specificity, and surviving baits for glucomannan targets were filtered with relaxed BLAST settings to maximize coverage. The number of single-copy targets was subsequently reduced by removing those where <90% of the baits survived and where less than 10 out of 14 taxa remained. Remaining baits were collapsed with a minimum of 83% overlap and >95% sequence identity, following a randomisation step. The final bait set (hereafter Orchidinae-205) was synthesized by Daicel Arbor Biosciences (Ann Arbor, MI, USA).

Detailed methodological choices and considerations regarding all steps of transcriptome assembly and filtering, target selection and bait development are available in the Text S1–S3, Figures S1 and S2.

2.2 | Sample collection, library preparation and sequencing

The baits were tested on target species belonging to the Orchidinae tribe occurring in the Eastern Mediterranean region. A list was drafted of all tuberous orchids occurring in Greece, Turkey and Iran (countries where the consumption and trade of these orchids is widespread), according to the World Checklist of Selected Plant families (Govaerts, 2019); hybrids and subspecies were excluded. This list was triangulated with the Field guide to the Orchids of Europe and the Mediterranean (Kühn et al., 2019), to obtain the most up-to-date species names and more detailed range maps, resulting in a list of 101 target species accepted by the WCSP versus 80 target species accepted by Kühn et al. (2019). The nomenclature of the latter was used to prioritize the final selection of species.

DNA samples were sourced from existing vouchered collections and extracted with a variety of protocols. Given the heterogeneity of the obtained DNA samples in terms of quantity and quality, all DNA concentrations, purities and integrities were quantified using Nanodrop One (Thermo Scientific, MA, USA), Qubit 2.0 (Life Technologies, CA, USA) and Fragment Analyzer (Agilent Technologies, CA, USA) or gel electrophoresis, respectively. Libraries for 79 samples were prepared with the Swift Accel-NGS 2S Hyb DNA Library Kit (Swift Biosciences, MI, USA; Cat. No. 23023, 2021) using unique dual indexing. DNA of these samples was sonicated using a Covaris E220 focused ultrasonicator (Covaris, MA, USA) to 400 bp fragments, making sure the input quantities were within the library protocol specifications (10 pg–1 μg), and amplified using 9 indexing PCR cycles. For target enrichment, samples were pooled in 12 equimolar groups of 8 samples each. Each pool with 100–600 ng total DNA was subsequently concentrated using Ampure XP (Beckman Coulter, CA, USA) using an elution volume of 10 μL. The RNA probes were hybridized at 62°C for 24 h, and 10 amplification cycles were carried out after enrichment, following the MyBaits V5 manual. The enriched libraries were sequenced at 150 bp paired-end on an Illumina NovaSeq SP flow cell.

Libraries for 11 additional samples were prepared using the Swift Turbo v2 DNA Library Kit (Swift Biosciences, MI, USA; Cat. No. 44096, 2021) using unique dual indexing. 100–200 ng of gDNA was sheared with an optimized enzymatic fragmentation step using 4 μL of the Swift Enzyme K3 for 40 s. Indexing PCR was performed using 5 cycles. For target enrichment, the 11 samples were split into two pools of 2 and 4 samples, respectively, with a normalized quantity of 400–450 ng per sample, and one pool of 5 samples ranging from 20 to 100 ng total input per sample, before being enriched with the RNA probes as described above. Following enrichment, these samples were sequenced at 150 bp paired-end on an Illumina NextSeq mid output flow cell.

2.3 | Phylogenomic analyses

Raw sequencing reads were trimmed with Trimmomatic v0.39 (Bolger et al., 2014) with the settings '`ILLUMINACLIP:"TruSeq3-PE.fa":2:30:10:2:TRUE LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:40`'. Surviving read pairs were assembled into contigs representing the target regions with HybPiper v14 (Johnson et al., 2016), and the exon sequences that were thus retrieved for each sample were concatenated to construct a sample-specific reference sequence. Samples with a target recovery of <100 kb were discarded. For five genes where more than 10% of samples issued a paralogue warning, paralog sequences were retrieved and aligned with MACSE v2.06 (Ranwez et al., 2018), and an approximate maximum likelihood (ML) tree was constructed with FastTree v2.1.11 (Price et al., 2010). The paralog trees were visually inspected to ascertain that there were no obvious paralogues among the primary copies. Nucleotide alignments were exported by replacing internal stop codons with Ns ('NNN') and frameshifts with dashes, and were trimmed based on a masked version of the amino acid alignment. Because the benefits of automated alignment trimming for phylogenomics is disputed, two trimming strategies were employed: one trimming poorly aligning amino acid segments using HmmCleaner (Di Franco et al., 2019); and one doing the same, but followed by removal of gappy columns using trimAl with the `-automated` option. HmmCleaner was run with the `--specificity` option to reduce false positives. Columns that were flagged for removal by trimAl were masked with MSA_trimmer (Kremer, 2017). All masked residues at the amino acid level were trimmed at the codon level with MACSE followed by removal of sequences with <100 nucleotides and sites with <5% taxon occupancy.

For both sets of trimmed exon alignments, maximum likelihood (ML) gene trees were created with IQ-TREE v2.1.2 (Minh et al., 2020) including model selection, 1000 bootstrap replicates and a maximum of 1000 iterations, as well as the additional options `--nstop 500` and `--allnni` for a more thorough tree search. The best tree out of ten independent runs was selected. Gene trees were edited by collapsing branches with ultra-fast bootstrap values of <30 with Newick utilities (Junier & Zdobnov, 2010), followed by removing implausibly long branches with TreeShrink (Mai & Mirarab, 2018). The edited gene trees were used to build both unconstrained species trees (allowing free placement of each individual sample) and constrained species trees (forcing samples that belong to the same species to be monophyletic) under the multispecies coalescent model using ASTRAL-III (Zhang et al., 2018).

Both sets of trimmed exon alignments were also used to infer an ML species tree with IQ-TREE v2.1.2, using model selection, 1000 bootstrap replicates and a maximum of 1000 iterations. Gene and site concordance factors for both sets of species trees were calculated with IQ-TREE v2.1.2. A polytomy test, which calculates the probability that the observed branch lengths are the result of a polytomy, was carried out with ASTRAL-III (Sayyari & Mirarab, 2018). Robinson-Foulds distances between trees were calculated with the R package 'dendextend' (Galili, 2015). Tanglegrams were generated with the 'dendextend' package following calibration of the trees with the `chronos` function of the R package 'ape' under a relaxed clock model (Paradis & Schliep, 2019). All other trees were visualized with ITOL v6 (Letunic & Bork, 2021).

2.4 | Comparison with other bait kits

Locus overlap was assessed between the Orchidinae-205 kit and two alternative kits, one for enrichment of low-copy nuclear genes in orchids (Orchidaceae-963) developed by Eserman et al. (2021), and one for flowering plants (Angiosperms-353) developed by Johnson et al. (2019). This was done by performing an all-by-all blastn search of the target files used for probe design with BLAST+ v2.9.0 (Camacho et al., 2009), reporting only hits with an e-value of <1e-6 and >70% sequence identity. Due to the sometimes large

taxonomic distances and patchy coverage between the species used for designing the different kits, the matches were validated (where sequence availability would allow) by blasting the sequences of the *Platanthera clavellata* (Michx.) Luer transcriptome (Angiosperms-353 and Orchidinae-205 baits) against each other with a minimum sequence identity of 100%, and against the *Platanthera blephariglottis* (Willd.) Lindl. recovered exons (Orchidaceae-963) with a minimum sequence identity of 90%. The results were visualized with the R package 'eulerr' (Larsson, 2022).

To compare the relative performance of the three kits for our species of interest, target recovery information was obtained for species from the same tribe (Angiosperms-353 baits) or subfamily (Orchidaceae-963 baits) and visualized in R. The potential of individual gene alignments to yield sufficient phylogenetic information for species tree inference was assessed by doing an in-depth comparison of gene alignment and gene tree statistics between the Angiosperms-353, Orchidaceae-963 and Orchidinae-205 generated sequences. Exon sequences were retrieved from Orchidaceae-963 alignments (available on https://github.com/laeserman/Orchidaceae963) for all available Orchidoideae species (*Spiranthes* spp., *Anoectochilus chapaensis* and *Platanthera blephariglottisi*) by blasting them against the available reference file used for probe design (Eserman et al., 2021). Exons for Angiosperms-353 loci of three closely related species (*Spiranthes australis*, *Goodyera umbrosa* and *Platanthera bifolia*) with a similar phylogenetic distance were retrieved from the Kew Tree of Life Explorer (release 2.0). The same was done for nine target species belonging to different genera that are included in this study and were also sequenced by Baker et al. (2022). This allowed for a subtribe-specific phylogenetic comparison between the Angiosperms-353 and Orchidinae-205 markers, for which we used *Habenaria arenaria* Lindl. and *Habenaria delavayi* Finet as outgroups.

All four sets of Orchidoideae sequences were aligned with MACSE v2.06 as described above. Columns in the alignment extremities were trimmed if they consisted of more than 50% gaps in sliding windows with a half window size of 3 bp. Alignment statistics for all four datasets were generated with AMAS (Borowiec, 2016) and gene trees and species trees were generated for both the two 9-species datasets separately as described above. Species trees were calibrated with the `chronos` function of the 'ape' package in R, with a root age of 22 Mya following (Inda et al., 2012), under a relaxed clock model. Thylogenetic informativeness (PI) of individual loci was inferred using PhyDesign (López-Giráldez & Townsend, 2011), and visualized in R. For statistical comparison, the area under the curve (AUC) for each PI profile was calculated with the R package 'DescTools'. Because phylogenetic informativeness is sensitive to tree topology, we took the two species trees that were most dissimilar, namely the ML trees generated from the different marker sets, and profiled the loci alongside the chronograms of both.

2.5 | Positive selection tests

To test the hypothesis that genes involved in the glucomannan biosynthesis pathway are under selection in (some of) our target species, we conducted a branch-site test with aBSREL (Smith et al., 2015), which tests whether selection has occurred on a proportion of sites along each branch in a tree. To construct the species tree, we selected the sample with the highest target recovery for each species that had multiple samples available. Exon sequences of the same samples were extracted for each candidate gene and aligned and trimmed as above with one difference: to account for the possible effect of gap-rich columns on inferences of positive selection, we applied two different gap thresholds to all alignments instead of the automated trimming heuristic, and removed all columns which consisted of more than 25% or more than 50% gaps, respectively. aBSREL was run on both versions of each gene alignment with HyPhy v2.5 (Kosakovsky Pond et al., 2020) from the command-line, and *p*-values for each branch were corrected for multiple testing with the Holm-Bonferroni correction.

3 | RESULTS

3.1 | Transcriptome quality and completeness

Trinity assembled anywhere from 52 to 369K contigs per assembly (Table S2). The proportion of reads mapping back to the assemblies as proper pairs ranged from 68% to 94%, with read representation falling within or exceeding the expected range of 70%–80% for all but one assembly (*Caladenia plicata* Fitzg.). TransRate scores were well above 0.1, with pre-filtering scores averaging between 0.2 and 0.4 and post-filtering scores between 0.3 and 0.6, removing on average 20% of contigs due to low confidence (Table S2).

ExN50 peaked below Ex90 in most cases, but rarely below Ex80, indicating that read sampling was adequate but not fully saturated and that transcriptome completeness could be improved by deeper sequencing (Table S3). The expression level at which N50 was optimized (varying between 70 and 90% of the expression data) contained between 9 and 16K contigs per assembly, with E90N50 varying from 407 to 2079 bp (Table S3). Despite using proteomes of several closely related and well-annotated reference genomes within the Orchidaceae, only around 1% of transcripts were found to be chimeric in all assemblies, and chimera detection was not improved by adding the proteomes of *Asparagus officinalis* L. (Asparagales) and *Oryza sativa* (Poales) to the set of orchid proteomes (Table S4). After retaining only the largest transcripts in clusters identified by Salmon, 32–156K contigs remained per assembly. Of these, between 16 and 50K were found to contain good candidates for open reading frames (Table S5).

Two transcriptomes belonging to the same study (*Dactylorhiza incarnata* and *D. fuchsii*) were removed because of their poor expression profiles (E90N50 <500) and BUSCO completeness scores (~10%). The remaining assemblies were on average 95% complete for BUSCOs in the eukaryote lineage, 82% complete for BUSCOs in the embryophyte lineage, and 78% complete for BUSCOs in the monocot lineage, with the least complete transcriptome containing about half of monocot and embryophyte genes (Table S6).

3.2 | Orthogroup inference and marker selection

In total, 89.5% of all Orchidoideae genes and 89.0% of all Orchidinae genes were assigned to 33,788 and 31,332 orthogroups, respectively, with 83%–95% of genes assigned to orthogroups per assembly (Table S7). The number of genes in species-specific orthogroups (orthogroups not shared by any other species) ranged from <1% to 8%. As expected, the number of orthogroups that contained sequences from all species was lower in the subfamily than for the tribe (5036 vs. 6365). This difference was even more dramatic for single-copy orthogroups (351 vs. 1295), a difference that was alleviated when considering orthogroups containing at least one representative per genus instead of species (1398 vs. 1910). *Ophrys sphegodes* had a copy for about 86–87% of the latter in both analyses, with 1049 transcripts occurring in both sets and 764 transcripts unique to either one of them. Of the total of 1813 *O. sphegodes* transcripts, 1381 mapped exactly once against the reference genome; 1053 of these had zero indels; 881 mapped with 100% coverage; and 542 were at least 750 bp in length. The majority of these (481) derived from the single-copy Orchidinae orthogroups. The average length of the single-copy genes was similar: 583 amino acids (Orchideae set) and 599 amino acids (Orchidoideae set), respectively. Based on these results, the 481 genes from the Orchidineae set were selected for alignment and filtering. After removing alignments where the average pairwise distance between any pair of ingroup species exceeded 0.1, a total of 308 loci were left and submitted for bait design.

A literature search yielded 19 gene families coding for enzymes putatively involved in the biosynthesis of glucomannan (Table S8), containing 52 candidate loci in model organism *Arabidopsis thaliana* (L.) Heynh., of which 47 loci were listed in the Rice Genome Annotation Project (RGAP) database (release 7) distributed over 27 orthologous groups (Kawahara et al., 2013). In two of these, no orthologous genes in *Oryza sativa* were identified. The remaining 25 orthologous groups contained 43 *O. sativa* homologues (Table S9). Orthogroup inference revealed that the RGAP homologues clustered in 23 orthogroups with the Orchidinae transcriptomes, including 34–51 unique transcripts per orchid species and corresponding to 25 orthogroups when *O. sativa* was excluded. After discarding highly divergent sequences and separating the orthogroups into clusters with high similarity, 31 alignments remained which were added to the 308 single-copy targets.

The final single-copy target loci had an average sequence similarity of 0.935–0.977, an average target length of 1779 bp, <10% alignment gaps and a GC content of 45.4%. The final glucomannan target loci had an average sequence similarity of 0.946, an average target length of 1976 bp, <20% alignment gaps and a GC content of 48.2%. For this selection of loci, initially 100k baits were designed, covering the entire target space of 524 kb. The final kit size was reduced to 60k baits, covering 205 loci and an average target space of 306 kb (Supplementary methods, Figure S2).

3.3 | Target recovery and alignment

Total sequencing output of the Illumina NovaSeq run was 1.5 billion fragments for 79 samples. Some species that were not included or had low sequencing output were included in the Illumina NextSeq run, which produced 268 million fragments for 11 samples. Seven samples in total were discarded due to insufficient read count (<1 M fragments), leaving 83 for phylogenomic analysis (Table S10). Adapter trimming removed on average 6%–11% of the read pairs before assembly. HybPiper assembled two-thirds of these reads, with an average of 66% mapping on target per sample. Median exon recovery was 315 kb; one additional sample with <20% of the median recovery was discarded. The remaining samples had >80% of the median exon recovery, ranging from 254 to 330 kb (Figure 1). In addition to exon recovery, each locus produced non-exonic sequences (by HybPiper referred to as 'introns') that were 1000–13,000 bp long, with a median intron recovery >1000 kb per sample, exceeding the intron length recovered by both the Orchidinae-963 and the Angiosperms-353 kits.

A BLAST search showed that six loci are shared by all three kits (Angiosperms-353, Orchidaceae-963 and Orchidinae-205). No other Orchidinae-205 loci overlapped with the Angiosperms-353 (Table S11), and an additional 13 loci were found to overlap with Orchidaceae-963 (Table S12), showing most Orchidinae-205 loci are unique targets (Figure 2). Only 961 target sequences were available for the Orchidaceae-963 kit, of which 248 in total had BLAST hits with the Angiosperms-353 kit, as opposed to the 254 reported by Eserman et al. (2021). Of the matching loci, target sequences and gene recovery in the Orchidinae-205 were higher than in the competing kits. Comparison of nine identical (Angiosperms-353 and Orchidiniae-205) and three related (Orchidaceae-963) orchid species shows that while the number of genes with recovered sequences is higher for the alternative kits, their total target recovery in base pairs is lower. Relative to the total target space, target recovery of Orchidinae-205 exceeds that of the two alternative kits, both in terms of the number of loci and number of base pairs recovered. While per-sample Orchidaceae-963 exon recovery was close in absolute terms to Orchidinae-205 exon recovery, only 289 loci had sequence information for all Orchidoideae, vastly reducing the amount of overlapping target space suitable for multi-species alignment.

For the selection of three Orchidoideae species, alignment statistics were comparable among the Angiosperms-353 and Orchidaceae-963, with no detectable difference in alignment length, slightly more missing data for the Orchidaceae-963 loci and slightly fewer variable sites for the Angiosperms-353 loci (Figure S3). In contrast, for the selection of nine Orchidinae species, the differences in alignment statistics between the Angiosperms-353 and the Orchidinae loci were more pronounced at a higher level of significance (Wilcoxon rank sum test). Orchidinae-205 produces (on average) longer alignments with less missing data, but more variable and phylogenetically informative sites than the Angiosperms-353 kit (Figure 3). The resulting gene trees also had higher average bootstrap support values and were more resolved.

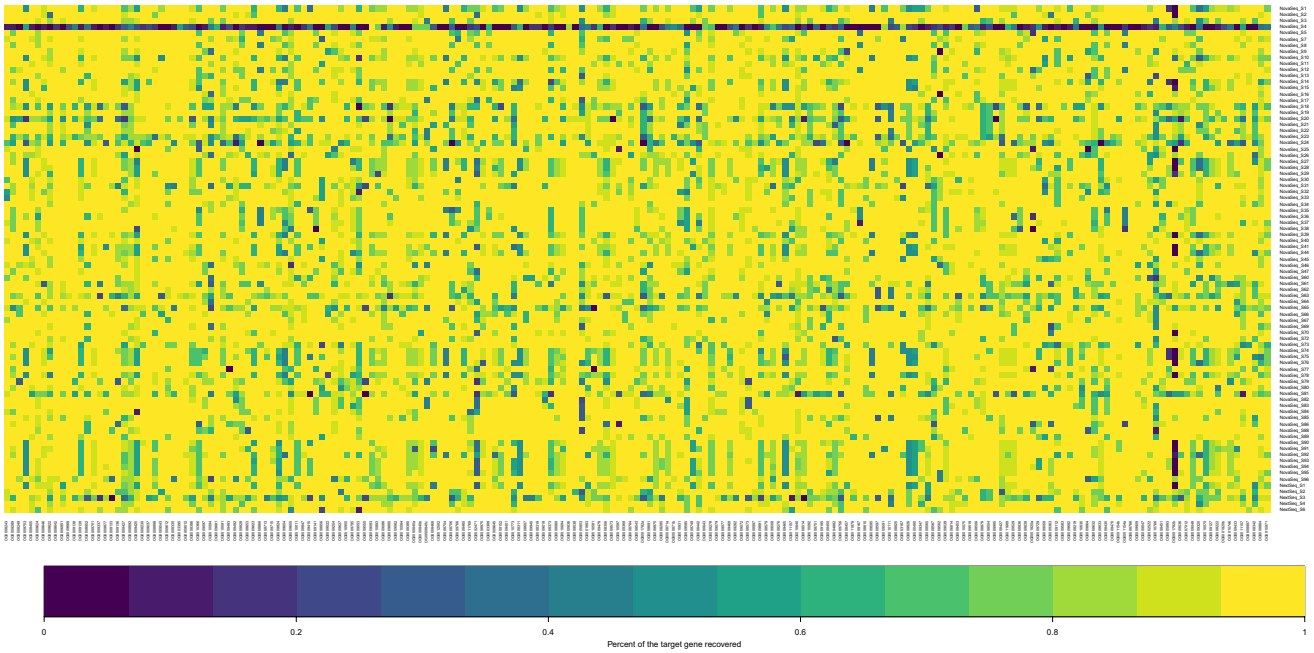

FIGURE 1 Relative exon recovery of samples across the target space. Rows indicate samples, corresponding to 83 terrestrial Orchidinae species analysed, and columns indicate loci targeted by the Orchidinae-205 baits.

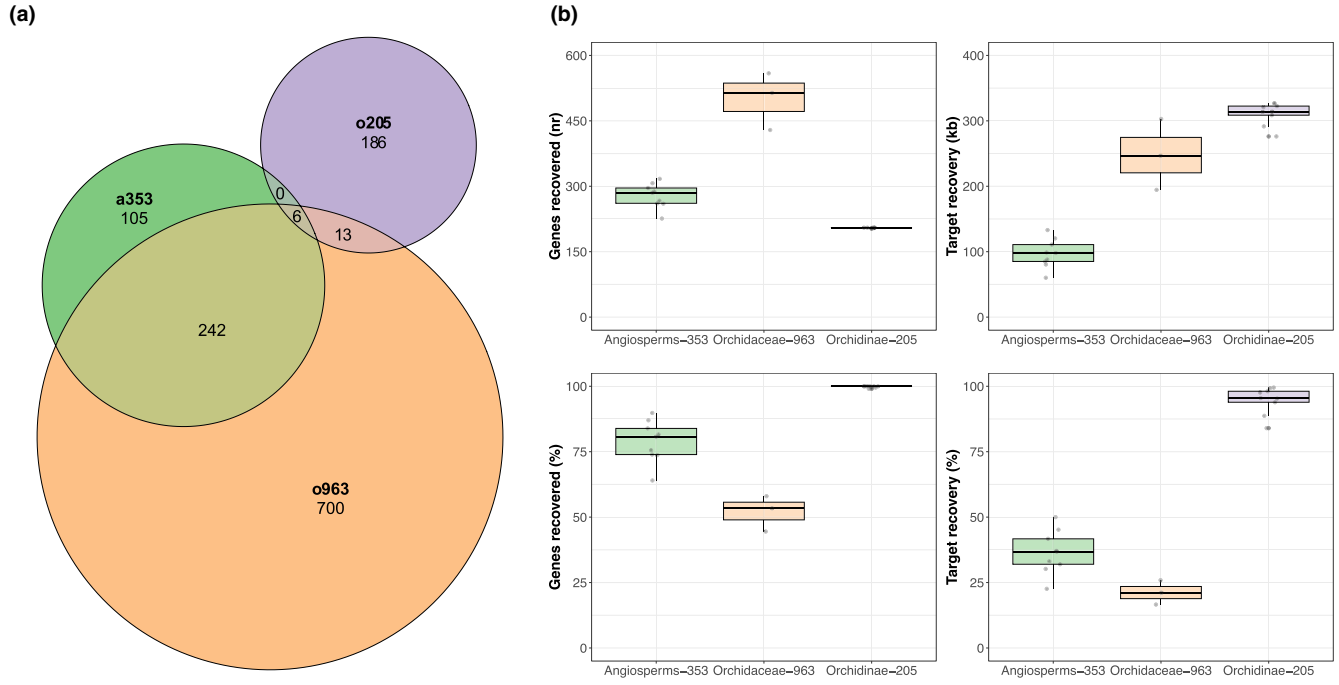

FIGURE 2 Comparison of Orchidinae-205 markers with two alternative bait sets. (a) Overlap in target loci between Orchidinae-205, Orchidaceae-963 and Angiosperms-353. (b) Target recovery in number (upper panels) and percentage (lower panels) of genes recovered (left) and base pairs recovered (right), for nine identical target species enriched with Angiosperms-353 (Baker et al., 2022) and Orchidinae-205 (this study) and 3 Orchidoideae species enriched with Orchidaceae-963 (Eserman et al., 2021). Target recovery length is based on assembled exons for the Orchidinae-205 loci (this study), exons published on the Kew Tree of Life Explorer (release 2.0) for the Angiosperms-353 loci and exons extracted from alignments published by Eserman et al. (2021) for the Orchidaceae-963 loci.

The phylogenetic informativeness (Figure 4) of the Orchidinae-205 loci was higher irrespective of topological variation in the inferred species tree (Wilcoxon rank sum test estimated difference in location: 70–76, $p < .001$), and the choice of tree did not significantly impact the ranking of the loci based on their AUC (Kendall's Tau rank correlation coefficient: .89–.91, $p < .001$). The species trees generated by the

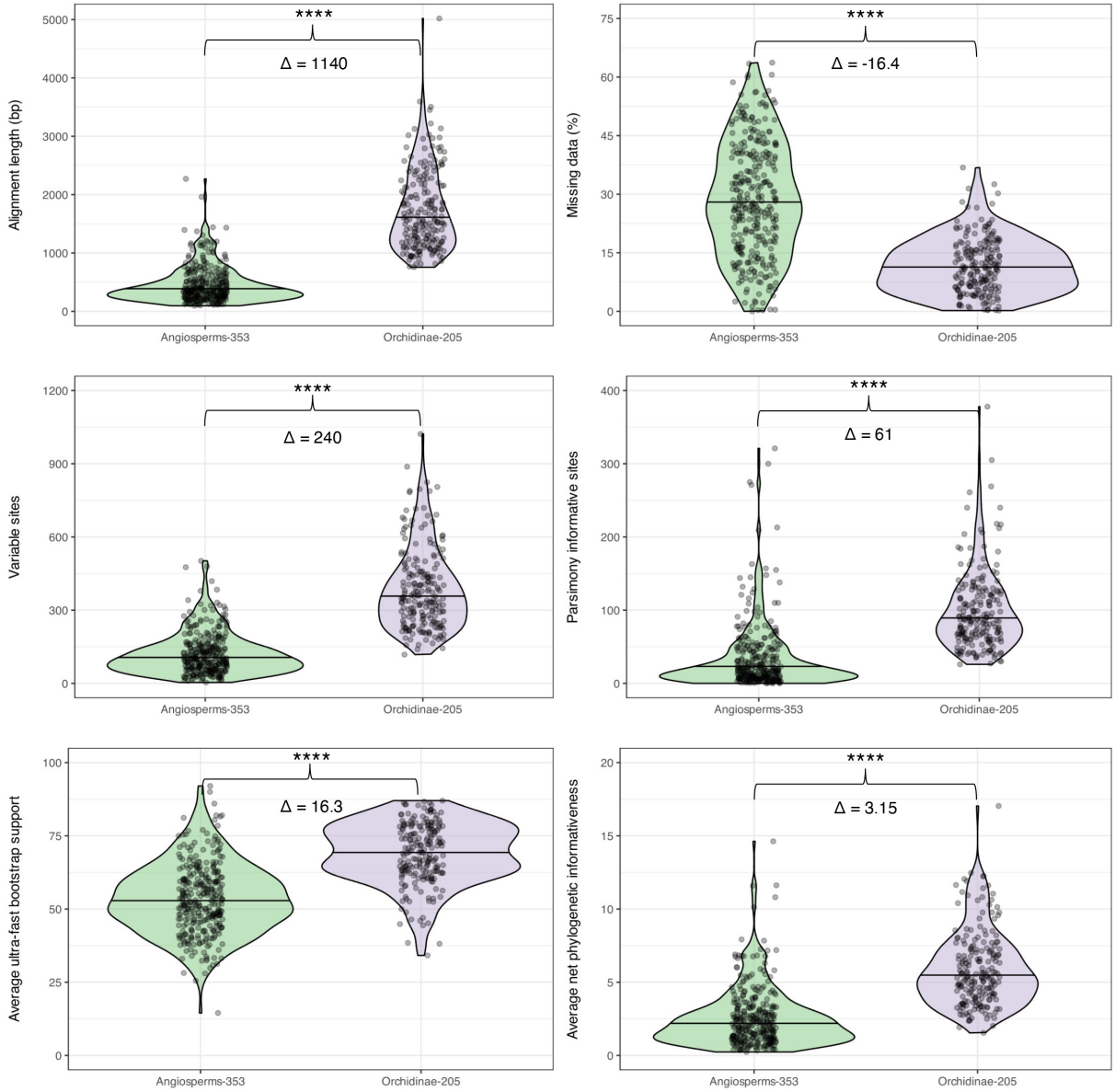

FIGURE 3 Statistical distribution of six different measures of phylogenetic information contained in the Angiosperms-353 loci and the Orchidine-205 loci. Values are based on alignments made with the same nine species for both marker sets. Each point represents a single locus. The midline of each violin plot represents the median value. Significance values and location shifts between the medians were estimated with a Wilcoxon rank sum test.

Orchidinae-205 alignments also had higher support and were more consistent than those generated by the Angiosperms-353 alignment, producing generic relationships that matched those found in the species tree (see below).

3.4 | Species tree reconstruction

For phylogenomic analyses, the 82 de novo assembled references were supplemented with three outgroup species (*Habenaria delavayi*, *H. pantlingiana* Kraenzl. and *Hemipilia forrestii* Rolfe) and two ingroup species (*Dactylorhiza hatagirea* (D.Don) Soó and *Gymnadenia densiflora* (Wahlenb.) A.Dietr.) whose sequences were used for probe design, but that were not among the regional target species. Total alignment length ranged from 759 to 7686 bp, with an average of 20% missing data. Trimming with HmmCleaner (option 1) reduced the average alignment length by 15%, and trimming with HmmCleaner + trimAl (option 2) by 21%, while the average amount of missing data per alignment was reduced to 10% and 8%, respectively. Trimming option 2 nearly doubled the loss of parsimony informative sites (17% of sites versus 9% of sites), but the proportion of parsimony informative sites per alignment remained relatively similar, at 27–28%. Taxon occupancy for each locus was high with 84–87 taxa (barring one outlier with more than 20 missing taxa) but was slightly lower for trimming option 2 where more sequences fell below the 100 bp threshold length.

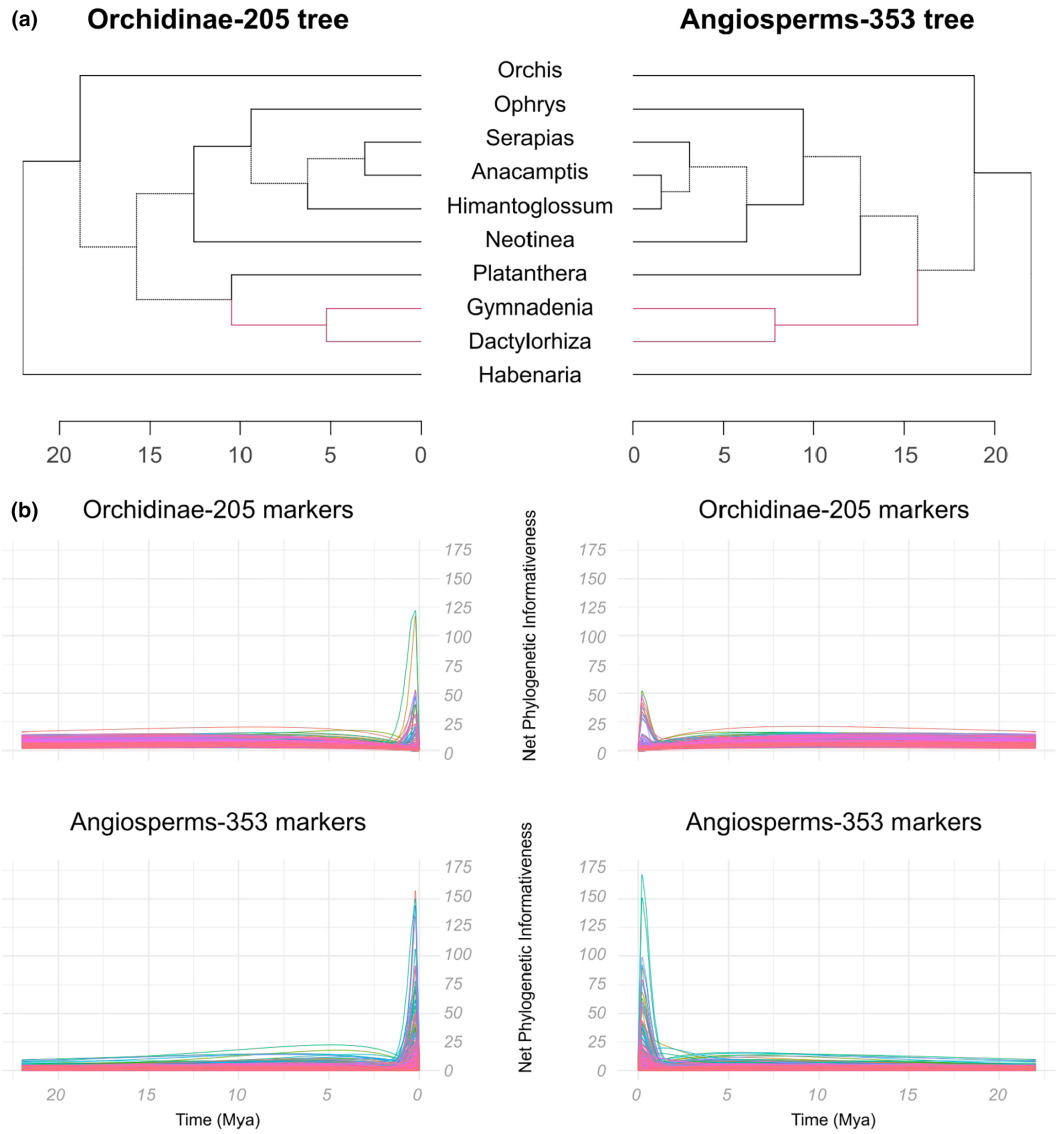

FIGURE 4 Phylogenetic informativeness profiles of the Angiosperms-353 loci and the Orchidinae-205 loci for nine selected species as measured against two chronograms with different topologies based on maximum likelihood inference. (a) Tanglegram (co-phylo plot) of two Orchidinae genus trees generated with different marker sets, 205 custom Orchidinae-205 markers (left), and 310 universal Angiosperms-353 markers (right). Lines connect identical species. Identical clades are highlighted pink. (b) Phylogenetic informativeness profiles of both marker sets as inferred from the ML tree based on a supermatrix of the 205 Orchidinae-205 alignments (left), or a supermatrix of the 310 Angiosperms-353 alignments (right).

Average node support in the ML species trees was high, but slightly higher when based on alignments trimmed with option 1 (96.4) than option 2 (95.2). The ultra-fast bootstrap (UF-BS) conveys strong support when it is around 95 or higher. This means that most clades in the ML trees are credible, with more well-supported nodes for the tree obtained with trimming option 1 (87%) than trimming option 2 (80%). The opposite is observed for the unconstrained coalescent species trees, where the final normalized quartet scores were nearly identical for both options (0.89) and the average posterior probability was slightly higher for trimming option 2 (0.91) than for option 1 (0.89). The posterior probability (PP) generally gives less support to the same clades than regular bootstrapping based on concatenation and is therefore more conservative. Using a threshold of 0.8, more nodes are well-supported for the tree obtained with trimming option 2 (80%) than trimming option 1 (76%), but this difference decreases when the threshold is relaxed to 0.7 (87% versus 85%). Given the similarity of the tree topologies and branch support values (Figure S4), and the higher number of parsimony informative sites for option 1, we here show results based on trimming option 1.

The sample-based ML and coalescent trees are broadly comparable, with a few key differences (Figure 5). All genera are monophyletic and form well-supported groups that (with the exception of *Neotinea* and *Orchis*) fall into two main clades: one grouping *Dactylorhiza* and *Gymnadenia* with *Platanthera*, *Pseudorchis* and *Traunsteinera* (clade A), and one grouping *Anacamptis* and *Serapias* with *Himantoglossum* and *Ophrys* (clade B). *Orchis*, whose placement has been disputed in

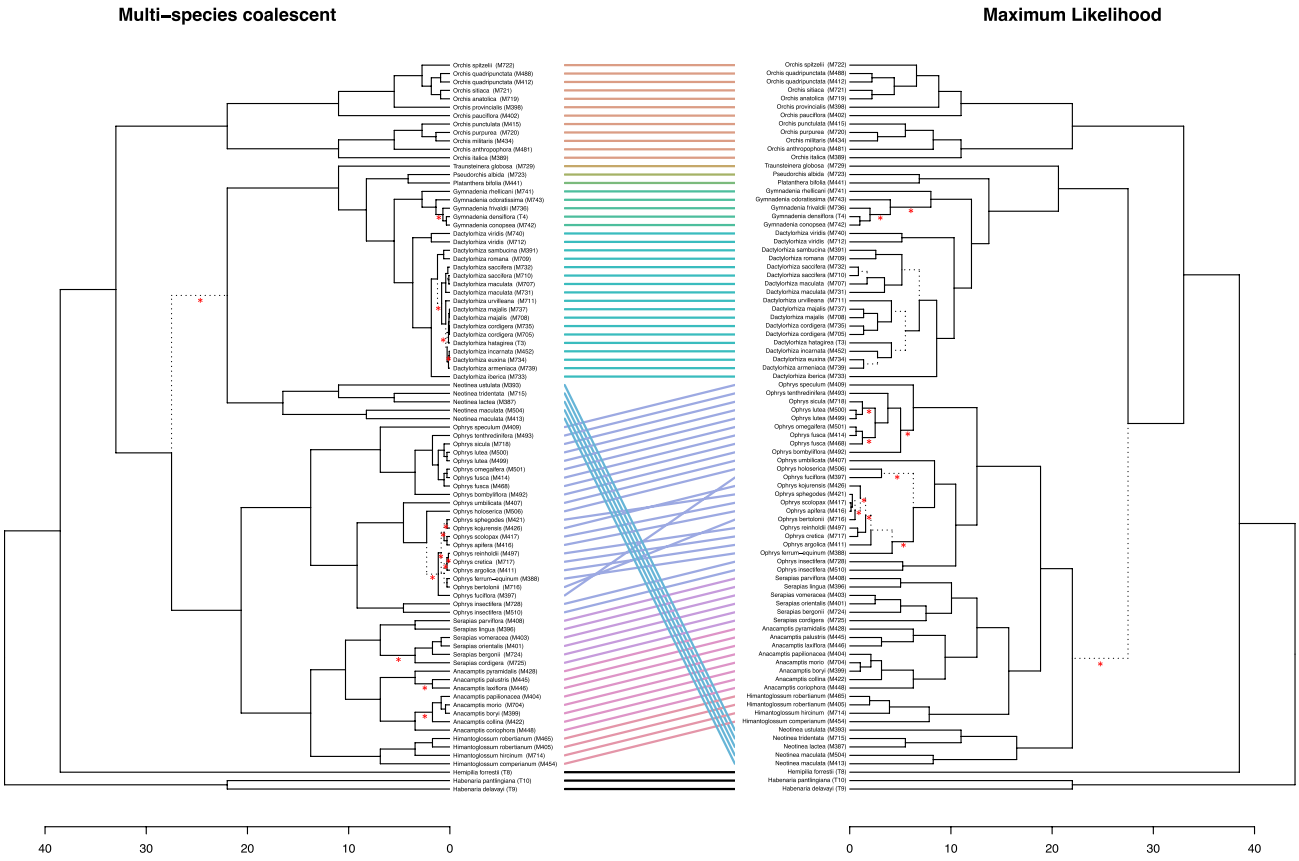

FIGURE 5 Tanglegram (co-phylo plot) of two Orchidinae species trees built with Orchidinae-205 loci analysed with two different methods: the multispecies coalescent as implemented in ASTRAL-III (left), and maximum likelihood as implemented in IQ-TREE (right). Coloured lines connect identical samples. Red asterisks indicate internal branches where support values are lower than 0.7 local posterior probability (left) and lower than 80 SH-aLRT or 95 ultrafast bootstrap (right). Support values for all branches are given in Figure S4.

recent studies, is consistently placed at the base of the ingroup with maximum support (100 UF-BS, 1.0 PP), as one of the first genera to branch off among our target taxa. The main difference between the concatenation and coalescent-based inference of genus relationships concerns *Neotinea*, whose placement has been similarly disputed and is here alternatively grouped as a sister to clade B (in the concatenated ML tree) or a sister to clade A (in the coalescent ASTRAL tree). Both options receive relatively low support (73 UF-BS and 0.67 PP). The branches surrounding this node are short, and a comparison of their length (in coalescent units) suggests that this area of the multispecies tree may be in the anomaly zone (Text S4, Figures S5 and S6). Despite the short internal branches at the basis of *Orchis* and *Neotinea*, a polytomy test rejected the null hypothesis that any of the branches leading to a split between genera are better represented by a polytomy (Figure S7). Most branches for which the null hypothesis was not rejected are more shallow and indicate recent splits between closely related species, suggesting that these might be soft polytomies that could be statistically rejected with more data.

Similarly, gene tree discordance is more widespread in shallow nodes than deeper nodes, with the exception of three consecutive nodes at the basis of the split between *Orchis* and clades A and B, the split between clade A and B and the early split of *Traunsteinera* from the rest of clade A (Figure S9). This indicates widespread conflict in the signal between genes, despite medium to high support for the main topology. Within the genera, most species relationships are well-supported, with some notable exceptions in *Dactylorhiza* and *Ophrys*. These nodes broadly correspond to those with low gene and site concordance factors in both trees (Figure S10). However, most species for which multiple individuals were sequenced were monophyletic, and the constrained species trees generated with ASTRAL did not have a noticeably worse performance than the unconstrained trees, with quartet scores of 0.89 (Figure S8).

3.5 | Diversifying selection events

A branch site test shows that most glucomannan target genes underwent episodic diversifying selection at least once in the evolutionary history of the Orchidinae tribe (Table S13). The only exception is OG0000643, a cellulose synthase-like D family protein, which might indicate a conserved nature and the presence of purifying rather than diversifying selection in our target species. The most frequent diversifying selection events were observed for OG0009824 (starch branching enzyme) and OG0003395 (ADP-glucose pyrophosphorylase), with less events detected in invertase and (phospho-)fructokinase genes. Certain branches experienced simultaneous positive selection events

in multiple loci (Figure S11). While this is mostly apparent in the tips (most notably *Dactylorhiza urvilleana* (Steud.) H. Baumann & Künkele, and several *Anacamptis*, *Orchis* and *Ophrys* species), our taxon sampling precludes conclusions about whether this selection is species-wide. Rather, some internal branches stand out for their position in the speciation history of the tribe, leading to larger clades or genera which subsequently radiated. Specifically, node 18 (at the basis of *Serapias*) has experienced selection on OG0005853 (mannose-6-phosphate isomerase) and OG0009824 (starch branching enzyme), and node 49 (at the basis of *Dactylorhiza*) appears to have experienced selection on OG0001154a (a sucrose synthase), OG0004882 (a starch synthase) and OG0001983 (a fructokinase). One of the main clades within *Orchis* (node 64) demonstrates diversifying selection on OG0002522 (another starch synthase) and OG0009112 (another ADP-glucose pyrophosphorylase family protein). Clades that stand out for their relative lack of selection events include *Himantoglossum* and *Neotinea*.

4 | DISCUSSION

Target capture has gained popularity in recent years as a method of choice for the phylogenomics of non-model organisms without reference genomes (Gasc et al., 2016). For flowering plants specifically, the release of the Angiosperms-353 kit has led to a boom in the number of studies employing target enrichment in a variety of plant families (Baker et al., 2021). Comparisons between the universal Angiosperms-353 loci and custom target loci for specific clades are becoming more common (Larridon et al., 2019; Ogutcen et al., 2021; Yardeni et al., 2022). This study adds to the growing body of literature that compares the merits of universal versus custom loci specifically, and of target capture more generally, and fits within a wider trend to design custom baits that target the entire orchid family (Eserman et al., 2021), specific (sub-)tribes (Peakall et al., 2021) and even genera (Bogarín et al., 2018; Wettewa & Wallace, 2021); with kits designed for broader taxonomic groups often showing merit at shallower evolutionary scales too (Granados Mendoza et al., 2019; Lagou et al., 2024; Wong et al., 2022).

4.1 | Factors of bait design influencing enrichment success

The effectiveness of custom and universal bait kits depends on a number of factors in the bait design, including first and foremost the evolutionary distance between the taxa used for developing the probes and the taxa that are enriched (Andermann et al., 2019). In this regard, the relatively poor target recovery of the Orchidaceae-963 for Orchidoideae is not surprising, given that the baits were designed with coding sequences from *Phalaenopsis equestris* (Schauer) Rchb.f., a member of the Epidendroideae, which are separated by at least 60 million years from other orchid subfamilies (Givnish et al., 2015; Gustafsson et al., 2010; Ramírez et al., 2007). This explains why only about half of the target genes and a quarter of the total target space could be retrieved by Eserman et al. (2021), and why their average alignment length, variability and taxon occupancy were not better than those generated by the Angiosperms-353 probes. The Angiosperms-353 kit is more universal in the true sense of the word, in that the target sequences were sourced from a wide range of taxa, yielding baits that are therefore in theory never more than 30% divergent from any species of flowering plant (Johnson et al., 2019). This matches results from our reciprocal blast, which did not show any hits below <70% identity. But while this kit is near-universal in its design, in practice the locus recovery for taxonomic groups that were not included in the bait design remains much lower than the total target space, and for Orchidinae even below the average recovery found by Johnson et al. (2019). In contrast, even though our kit was not designed with representatives from each of our target genera, the enrichment of genera not included in the design was not noticeably lower. Given the inclusion of *Habenaria* and *Hemipilia* sequences and the chosen thresholds for sequence similarity, we therefore expect our kit to be broadly applicable to all genera in the Orchidinae tribe.

In addition to the genetic proximity of the taxa used for bait design, a second decisive factor for the effectiveness of a bait kit is determined by the length of the target loci that is effectively enriched. Recent studies have suggested that universal markers do not always have less phylogenetic power than custom loci, and that the number of segregating sites is mostly impacted by locus recovery and length rather than variations in the number of single nucleotide polymorphisms (Larridon et al., 2019; Ogutcen et al., 2021). Locus recovery is partly determined by the length and therefore the choice of the gene itself, but is also modulated by probe capture efficiency, which is a function of sequence similarity and hence the taxon sampling during bait design (Andermann et al., 2019). Locus choice in our design was obviously limited by the transcriptome assemblies used, which had varying levels of completeness. While sequencing depth of the transcriptomes used for probe design can explain part of this variation, some of it may also be the result of tissue sampling, as most transcriptomes were generated only from one tissue (flower or leaf) in which not all genes may be expressed. Other genes with better length, informativeness and recovery might therefore exist that did not pass our quality filters. Yet, the chosen genes are optimized for effective enrichment across the tribe, and the expected differences in capture efficiency of the three kits based on their design (with Orchidinae-205 expected to be the most efficient and Orchidaceae-963 the least) are confirmed by their observed relative target recoveries for the species we analysed. Nonetheless, studies aiming to examine phylogenetic relationships on a broader taxonomic scale extending beyond the Orchidoideae subfamily may wish to look into the Orchidaceae-963 as an alternative resource, as it offers more sequence information than the Angiosperms-353 probes particularly for Epidendroideae, while the Angiosperms-353 will remain instrumental for investigations spanning multiple families.

4.2 | Phylogenetic information and congruence

Sufficient phylogenetic information is crucial for researchers wishing to employ a coalescent approach to tree construction, which is sensitive to the quality of the gene trees that serve as input for the species

tree (Xu & Yang, 2016). Short sequences tend to contain less genetic variation and produce poorly supported gene trees, interfering with species tree reconstruction in two-step multispecies coalescent approaches that rely on estimated gene trees. As the absolute target recovery of the Angiosperms-353 loci within Orchidoideae is the lowest of all three kits, closely followed by the useable (because overlapping between species) target recovery of the Orchidaceae-963 loci, both are expected to lead to less informative loci, more missing data and poorer support for the species tree. The longer alignments and higher taxon occupancy of the Orchidinae-205 loci therefore make it the more suitable kit for gene tree reconstruction and species tree reconstruction under the multispecies coalescent within the subtribe. The larger recovered gene space and higher amounts of phylogenetic information also increase the robustness of ML tree reconstruction relative to the Angiosperms-353 loci, which explains the higher support values and its closer similarity to the coalescent tree, which is less impacted by gene tree uncertainty. In conclusion, phylogenetic inference with the Orchidinae-205 markers appears to be more reliable, and given the similar performance of Orchidaceae-963 and Angiosperms-353 among closely related Orchidoideae, we expect this conclusion to hold in comparison to both alternatives. Given that the Orchidinae-205 markers are tailored to the subtribe, these differences are expected to be amplified at shallower phylogenetic scales, where longer sequences with more phylogenetic information are even more important.

In cases where exon recovery is insufficient, it may be necessary to look into the flanking sequences of the exons, which are often enriched as by-catch of target sequence enrichment. In addition to introns, these non-exonic sequences could also include promoter regions and untranslated regions flanking the first and last exon. Non-coding sequences are not subjected to the same functional constraints and may therefore evolve faster than coding sequences, yielding a higher percentage of segregating sites. Since we observe over three times more intron recovery than exon recovery for our target species, outstripping also the non-exon lengths generated with the Orchidaceae-963 and Angiosperms-353 probes, we expect that the non-coding sequences recovered by Orchidinae-205 offer a vast untapped potential for studying the evolution of Orchidinae in more recent evolutionary timeframes. This will require the use of different analysis pipelines than for exons, but examples are emerging that effectively use target capture sequencing data for *kmer* block analysis of target and off-target reads (Peakall et al., 2021; Wong et al., 2022), and for traditional read mapping and variant calling (Bi et al., 2013; Slimp et al., 2021), as is customary in population genomic analyses. These can potentially be used to differentiate between closely related populations and aid in more detailed conservation genomic and forensic analyses.

4.3 | Unresolved species relationships and new insights

The high support for monophyly of species for which multiple individuals were sequenced indicates that the exons recovered are able to correctly cluster con-specific samples in at least six genera. However, due to sampling limitations, we were unable to test monophyly for all species, and exons might not be able to resolve all relationships. Recovering introns could be especially relevant for recently diverged lineages in Orchidinae that currently suffer from poor phylogenetic resolution. In general, nodes with low support in the coalescent species tree are characterized by low gene and site concordance factors, but the inverse is not always true, which means that some well-supported clades in the species trees are not necessarily supported by a majority of gene trees. The discrepancy between the gene concordance factors (gCF) and site concordance factors (sCF) that exists for some nodes suggests that incomplete lineage sorting is not solely responsible for this conflicting signal, but that certain genes may simply lack sufficient variation to be informative for certain splits. The low numbers of informative genes for shallow splits in particularly *Serapias*, *Ophrys* and *Dactylorhiza* indicate that there might be a lack of segregating sites available in the exons for resolving these nodes. A possible remedy could therefore be to generate clade-specific alignments of non-coding sequences generated in this study to obtain more resolution in these species complexes.

While the analysis of non-coding sequences may resolve some of the discordance we see here due to recent and rapid speciation (*Ophrys* and *Serapias*, see Breitkopf et al., 2015; Inda et al., 2012), a fraction is expected to remain in readily hybridizing lineages (*Dactylorhiza* and *Gymnadenia*, see Brandrud et al., 2020; Hedrén et al., 2018; Pillon et al., 2007) where the discordance is partially caused by gene flow. In these cases, separating haplotypes by putative parental species may further help to clarify the evolutionary histories of different fractions of the genome. In other cases, such as rapid radiations and consecutive short branches deeper in the phylogeny, gene discordance might never be fully resolved. This is especially the case for divergence events that fall within the anomaly zone, where the majority of gene trees will contradict the true species tree due to short coalescent times. However, where gene flow is the cause for discordance on deep nodes, more detailed evolutionary genomic analyses could elucidate the extent to which hybridisation has led to basal reticulation patterns (Cai et al., 2021; Morales-Briones et al., 2021). The conflicting estimates of the position of *Neotinea* within Orchidinae and widespread discordance around the splits of *Orchis* and *Traunsteinera* could be further explored by taking into account these different scenarios.

A detailed systematic (re-)evaluation of the tribe is beyond the scope of this study, but there are two novel insights regarding species relationships that warrant mention here, because they question recent taxonomic consensus. The first is that *Ophrys insectifera* L. is not the most basal lineage in the genus as long thought based on traditional markers (Breitkopf et al., 2015; Devey et al., 2008), but has a well-supported inner placement as a sister group to several more derived lineages, including the *O. sphegodes*, *O. fusca* Link and *O. scolopax* Cav. species complexes. This corroborates recent findings from whole plastid genomes (Bertrand et al., 2021), which is remarkable given the frequently observed mismatch between plastid and nuclear genetic histories (Pérez-Escobar et al., 2021). The second observation is that *Serapias bergonii* E.G. Camus, which is sometimes considered a subspecies of *S. vomeracea* (Burm.f.) Briq. (Bellusci et al., 2008; Kühn et al., 2019), is sister to both *S. orientalis* (Greuter) H.Baumann & Künkele and *S. vomeracea* (Burm.f.) Briq. This

means that, unless we believe *S. orientalis* to be a subspecies of the same species complex, we should consider the possibility that *S. bergonii* is a separate species based on the phylogenetic species concept. Since our species selection focuses mainly on the Eastern Mediterranean, new insights will undoubtedly emerge as more species and subspecies are added to this reference database.

4.4 | Future applications

Notwithstanding remaining uncertainties in the circumscription of some orchid species, Orchidinae-205 is an important step forward in the progressive refinement of the phylogeny of European terrestrial orchids, and the identification of orchid-derived products at the species and population level. While it is increasingly feasible to utilize fresh and silica-dried material for large-scale phylotranscriptomic studies (He et al., 2022; Zhang et al., 2023), as well as reuse existing datasets for this purpose (Wong & Peakall, 2022), target capture's unique ability to obtain high-quality sequence information from even severely degraded plant material renders it the sequencing method of choice for identification and systematics of low DNA input samples – even on shallow evolutionary scales where RAD-seq would normally be adopted (Harvey et al., 2016). However, the costs of library preparation and enrichment still prohibit widespread implementation of this method in practice. Luckily, strategies exist that can reduce the per-sample processing costs considerably (Hale et al., 2020). Another promising outlook comes from previous studies which suggest that rather than the pure number of genes, the choice and length of genes matter more for phylogenetic resolution (Ai & Kang, 2015; Wortley et al., 2005). By maximizing locus length and coverage, the Orchidinae-205 increases the chance that a smaller number of genes will be sufficient to obtain well-supported trees. This opens the door to a reduced set of target loci or a multi-locus barcode that is tailored to Orchidinae, and that can be deployed cost-effectively at a larger scale.

The bait set presented here also offers the possibility to assess functional differences between orchid species that are preferred for salep and those that do not satisfy consumer preferences for a specific polysaccharide composition. While the role of the glucomannan target genes (if any) in speciation and/or adaptation is unclear, evidence of diversifying selection may point to clades with an altered metabolic pathway for the production of polysaccharides, and hence a different glucomannan content. Phenotypic measurements of the traits affected by these genes, such as cell wall composition and glucomannan concentration and in different tissues and at different developmental stages, will allow us to detect which traits (if any) display a phylogenetic signal. The phylogenetic framework generated here could thus form the basis for future studies on trait evolution. To elucidate the genetic basis of such trait variation, site-specific tests of selection (Murrell et al., 2015) and variant effect prediction (Cingolani et al., 2012) will be useful follow-up analyses. Experimental validation of the functional effects of sequence variation in glucomannan target genes, coupled with analysis of gene duplication and expression data, could give further insight into which enzymes involved in the polysaccharide biosynthesis pathway exhibit a strong link with glucomannan production. Such analyses will be facilitated by the emergence of novel genomes and transcriptome data, and could benefit orchid breeding efforts (Zhang et al., 2022). Lastly, the utilization of these baits in herbarium and museum contexts offers an opportunity to study the changes in species diversity and provenance of salep over time, as well as any effects on genetic diversity and adaptation as a result of overexploitation.

AUTHOR CONTRIBUTIONS

Margaretha A. Veltman, Barbara Gravendeel and Hugo J. de Boer designed and conceived the research. Bastien Anthoons and Audun Schrøder-Nielsen performed laboratory work. Margaretha A. Veltman analysed transcriptomic and targeted sequence capture data. Margaretha A. Veltman wrote the manuscript with input from all co-authors. All authors contributed to the manuscript and approved the final version.

ACKNOWLEDGEMENTS

This research would not have been possible without the generous contribution of many people who shared samples and offered their expertise on European and Mediterranean orchids. We are indebted to Baset Ghorbani (Uppsala University) for providing reference material from Iran, Andreas Drouzas (Aristotle University of Thessaloniki) and Spyros Tsiftsis (International Hellenic University) for providing reference material from Greece, Jaco Kruizinga (Hortus botanicus Leiden) for providing reference material from living collections in the Netherlands, as well as Mikael Hedrén (Lund University), Gábor Sramkó (University of Debrecen), Jana Jersáková (University of South Bohemia) and Philipp Schlüter (University of Hohenheim) for sharing the collections of their research groups. We acknowledge the Royal Botanic Gardens Kew for providing DNA aliquots from their DNA and Tissue bank, which allowed us to fill in gaps in our reference collection. We would also like to thank Huub van Proemeren (Staatsbosbeheer) for providing access to protected areas and to Marcel Eurlings (Naturalis Biodiversity Center) for aiding with DNA extractions. Finally, we are grateful to Alessia Russo (University of Zürich) and Philipp Schlüter (University of Hohenheim) for providing advance access to a draft version of the *Ophrys sphegodes* genome. 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.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Targeted sequencing reads analysed as part of this study are deposited in the NCBI SRA under project number PRJNA1053613. Associated accession numbers can be found in the Table S10. Assembled transcriptome and target capture sequences are available on Dryad (https://doi.org/10.5061/dryad.sj3tx96bn).

BENEFIT-SHARING STATEMENT

This research uses biological samples from a number of countries. These samples have been provided for research purposes as institutional exchange. The research results are intended to empower national law enforcement to monitor trade in terrestrial orchids.

ORCID

Margaretha A. Veltman https://orcid.org/0000-0002-4106-4516

REFERENCES

Ai, B., & Kang, M. (2015). How many genes are needed to resolve phylogenetic incongruence? *Evolutionary Bioinformatics Online*, *11*, 185–188.

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. *Frontiers in Genetics*, *10*, 1407.

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., … Forest, F. (2022). A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology*, *71*(2), 301–319.

Baker, W. J., Dodsworth, S., Forest, F., Graham, S. W., Johnson, M. G., McDonnell, A., Pokorny, L., Tate, J. A., Wicke, S., & Wickett, N. J. (2021). Exploring Angiosperms353: An open, community toolkit for collaborative phylogenomic research on flowering plants. *American Journal of Botany*, *108*(7), 1059–1065.

Bellusci, F., Pellegrino, G., Palermo, A. M., & Musacchio, A. (2008). Phylogenetic relationships in the orchid genus *Serapias* L. based on noncoding regions of the chloroplast genome. *Molecular Phylogenetics and Evolution*, *47*(3), 986–991.

Bertrand, J. A. M., Gibert, A., Llauro, C., & Panaud, O. (2021). Whole plastid genome-based phylogenomics supports an inner placement of the *O. insectifera* group rather than a basal position in the rapidly diversifying *Ophrys* genus (Orchidaceae). *Botany Letters*, *168*(3), 452–457.

Bi, K., Linderoth, T., Vanderpool, D., Good, J. M., Nielsen, R., & Moritz, C. (2013). Unlocking the vault: Next-generation museum population genomics. *Molecular Ecology*, *22*(24), 6018–6032.

Bogarín, D., Pérez-Escobar, O. A., Groenenberg, D., Holland, S. D., Karremans, A. P., Lemmon, E. M., Lemmon, A. R., Pupulin, F., Smets, E., & Gravendeel, B. (2018). Anchored hybrid enrichment generated nuclear, plastid and mitochondrial markers resolve the *Lepanthes horrida* (Orchidaceae: Pleurothallidinae) species complex. *Molecular Phylogenetics and Evolution*, *129*, 27–47.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

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

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). *Systematic Biology*, *69*(1), 91–109.

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. *The New Phytologist*, *207*(2), 377–389.

Bulpitt, C. J. (2005). The uses and misuses of orchids in medicine. *QJM: An International Journal of Medicine*, *98*(9), 625–631.

Cai, L., Xi, Z., Lemmon, E. M., Lemmon, A. R., Mast, A., Buddenhagen, C. E., Liu, L., & Davis, C. C. (2021). The perfect storm: Gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. *Systematic Biology*, *70*(3), 491–507.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421.

Chase, M. W., Cameron, K. M., Freudenstein, J. V., Pridgeon, A. M., Salazar, G., van den Berg, C., & Schuiteman, A. (2015). An updated classification of Orchidaceae. *Botanical Journal of the Linnean Society*, *177*(2), 151–174.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92.

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*. *Annals of Botany*, *101*(3), 385–402.

Di Franco, A., Poujol, R., Baurain, D., & Philippe, H. (2019). Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evolutionary Biology*, *19*(1), 21.

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 in Plant Science*, *24*(10), 887–891.

Ece Tamer, C., Karaman, B., & Utku Copur, O. (2006). A traditional Turkish beverage: Salep. *Food Reviews International*, *22*(1), 43–50.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238.

Eserman, L. A., Thomas, S. K., Coffey, E. E. D., & Leebens-Mack, J. H. (2021). Target sequence capture in orchids: Developing a kit to sequence hundreds of single-copy loci. *Applications in Plant Sciences*, *9*(7), e11416.

Galili, T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22), 3718–3720.

Gasc, C., Peyretaillade, E., & Peyret, P. (2016). Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids Research*, *44*(10), 4504–4518.

Ghorbani, A., Gravendeel, B., Naghibi, F., & de Boer, H. (2014). Wild orchid tuber collection in Iran: A wake-up call for conservation. *Biodiversity and Conservation*, *23*(11), 2749–2760.

Ghorbani, A., Gravendeel, B., Selliah, S., Zarré, S., & de Boer, H. (2017). DNA barcoding of tuberous Orchidoideae: A resource for identification of orchids used in Salep. *Molecular Ecology Resources*, *17*(2), 342–352.

Givnish, T. J., Spalink, D., Ames, M., Lyon, S. P., Hunter, S. J., Zuluaga, A., Iles, W. J. D., Clements, M. A., Arroyo, M. T. K., Leebens-Mack, J., Endara, L., Kriebel, R., Neubig, K. M., Whitten, W. M., Williams, N. H., & Cameron, K. M. (2015). Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1814), 20151553. https://doi.org/10.1098/rspb.2015.1553

Govaerts, R. (2019). *World checklist of selected plant families in the catalogue of life*.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652.

Granados Mendoza, C., Jost, M., Hágsater, E., Magallón, S., van den Berg, C., Lemmon, E. M., Lemmon, A. R., Salazar, G. A., & Wanke,

S. (2019). Target nuclear and off-target plastid hybrid enrichment data inform a range of evolutionary depths in the orchid genus *Epidendrum*. *Frontiers in Plant Science*, *10*, 1761.

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 Evolutionary Biology*, *10*, 177.

Hale, H., Gardner, E. M., Viruel, J., Pokorny, L., & Johnson, M. G. (2020). Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in Plant Sciences*, *8*(4), e11337.

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, *65*(5), 910–924.

He, J., Lyu, R., Luo, Y., Xiao, J., Xie, L., Wen, J., Li, W., Pei, L., & Cheng, J. (2022). A phylotranscriptome study using silica gel-dried leaf tissues produces an updated robust phylogeny of Ranunculaceae. *Molecular Phylogenetics and Evolution*, *174*, 107545.

Hedrén, M., Lorenz, R., & Ståhlberg, D. (2018). Evidence for bidirectional hybridization between *Gymnadenia* and *Nigritella*. *Journal Europäischer Orchideen*, *50*(1), 43–60.

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. (2017). A review of the trade in orchids and its implications for conservation. *Botanical Journal of the Linnean Society*, *186*(4), 435–455.

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS One*, *6*(5), e19254.

Hollingsworth, P. M., Li, D.-Z., van der Bank, M., & Twyford, A. D. (2016). Telling plant species apart with DNA: From barcodes to genomes. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences*, *371*(1702), 20150338. 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. *Annals of Botany*, *110*(1), 71–90.

Jahanbanifard, M., Veltman, M. A., Veldman, S., Hartvig, I., Cowell, C., Lens, F., Janssens, S., & Smets, E. (2022). Wildlife trade. In H. de Boer, M. O. Rydmark, B. Verstraete, & B. Gravendeel (Eds.), *Molecular identification of plants: From sequence to species*. Advanced Books.

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. *Applications in Plant Sciences*, *4*(7), 1600016. 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. *Systematic Biology*, *68*(4), 594–606.

Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, *25*(1), 185–202.

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, *26*(13), 1669–1670.

Kadlec, M., Bellstedt, D. U., Le Maitre, N. C., & Pirie, M. D. (2017). Targeted NGS for species level phylogenomics: "Made to measure" or "one size fits all"? *PeerJ*, *5*, e3569.

Kasparek, M., & Grimm, U. (1999). European trade in Turkish salep with special reference to Germany. *Economic Botany*, *53*(4), 396–406.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J., Zhou, S., Childs, K. L., Davidson, R. M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S. S., Kim, J., Numa, H., … Matsumoto, T. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, *6*(1), 4.

Kosakovsky Pond, S. L., Poon, A. F. Y., Velazquez, R., Weaver, S., Hepler, N. L., Murrell, B., Shank, S. D., Magalis, B. R., Bouvier, D., Nekrutenko, A., Wisotsky, S., Spielman, S. J., Frost, S. D. W., & Muse, S. V. (2020). HyPhy 2.5-A customizable platform for evolutionary hypothesis testing using phylogenies. *Molecular Biology and Evolution*, *37*(1), 295–299.

Kremer, L. (2017). *MSA_trimmer: Straightforward & minimalistic removal of poorly aligned regions in Multiple Sequence Alignments*. https://github.com/LKremer/MSA_trimmer

Kreziou, A., de Boer, H., & Gravendeel, B. (2016). Harvesting of salep orchids in north-western Greece continues to threaten natural populations. *Oryx*, *50*(3), 393–396.

Kühn, R., Pedersen, H., & Cribb, P. (2019). *Field guide to the orchids of Europe and the Mediterranean*. Kew Publishing.

Kurt, A. (2021). Salep glucomannan: Properties and applications. In Inamuddin, M. I. Ahamed, R. Boddula, & T. Altalhi (Eds.), *Polysaccharides* (pp. 177–203). Wiley. https://doi.org/10.1002/9781119711414.ch9

Lagou, L. J., Kadereit, G., & Morales-Briones, D. F. (2024). Target enrichment data uncovers rapid radiation, whole genome duplication, and extensive hybridization in slipper orchid genus *Cypripedium* L. *bioRxiv*. https://doi.org/10.1101/2024.01.24.577114

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. *Frontiers in Plant Science*, *10*, 1655.

Larsson, J. (2022). *eulerr: Area-proportional Euler and Venn diagrams with ellipses* (Version 7.0.0). https://cran.r-project.org/package=eulerr

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, *32*(10), 2798–2800.

Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296.

Li, M.-H., Liu, K.-W., Li, Z., Lu, H.-C., Ye, Q.-L., Zhang, D., Wang, J.-Y., Li, Y.-F., Zhong, Z.-M., Liu, X., Yu, X., Liu, D.-K., Tu, X.-D., Liu, B., Hao, Y., Liao, X.-Y., Jiang, Y.-T., Sun, W.-H., Chen, J., … Liu, Z.-J. (2022). Genomes of leafy and leafless *Platanthera* orchids illuminate the evolution of mycoheterotrophy. *Nature Plants*, *8*(4), 373–388.

López-Giráldez, F., & Townsend, J. P. (2011). PhyDesign: An online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology*, *11*, 152.

Mai, U., & Mirarab, S. (2018). TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, *19*(Suppl 5), 272.

Margulies, J. D., Bullough, L.-A., Hinsley, A., Ingram, D. J., Cowell, C., Goettsch, B., Klitgård, B. B., Lavorgna, A., Sinovas, P., & Phelps, J. (2019). Illegal wildlife trade and the persistence of "plant blindness". *Plants, People, Planet*, *1*(3), 173–182.

Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature*, *536*(7615), 143–145.

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. *Molecular Biology and Evolution*, *37*(5), 1530–1534.

Morales-Briones, D. F., Kadereit, G., Tefarikis, D. T., Moore, M. J., Smith, S. A., Brockington, S. F., Timoneda, A., Yim, W. C., Cushman, J. C., & Yang, Y. (2021). Disentangling sources of gene tree discordance in phylogenomic data sets: Testing ancient hybridizations in Amaranthaceae s.L. *Systematic Biology*, *70*(2), 219–235.

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D. P., Smith, D. M., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Gene-wide identification of episodic selection. *Molecular Biology and Evolution*, *32*(5), 1365–1371.

Ogutcen, E., Christe, C., Nishii, K., Salamin, N., Möller, M., & Perret, M. (2021). Phylogenomics of Gesneriaceae using targeted capture of nuclear genes. *Molecular Phylogenetics and Evolution*, *157*, 107068.

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528.

Peakall, R., Wong, D. C. J., Phillips, R. D., Ruibal, M., Eyles, R., Rodriguez-Delgado, C., & Linde, C. C. (2021). A multitiered sequence capture strategy spanning broad evolutionary scales: Application for phylogenetic and phylogeographic studies of orchids. *Molecular Ecology Resources*, *21*(4), 1118–1140.

Pérez-Escobar, O. A., Dodsworth, S., Bogarín, D., Bellot, S., Balbuena, J. A., Schley, R. J., Kikuchi, I. A., Morris, S. K., Epitawalage, N., Cowan, R., Maurin, O., Zuntini, A., Arias, T., Serna-Sánchez, A., Gravendeel, B., Torres Jimenez, M. F., Nargar, K., Chomicki, G., Chase, M. W., … Baker, W. J. (2021). Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *American Journal of Botany*, *108*(7), 1166–1180.

Pillon, Y., Fay, M. F., Hedrén, M., Bateman, R. M., Devey, D. S., Shipunov, A. B., van der Bank, M., & Chase, M. W. (2007). Evolution and temporal diversification of western European polyploid species complexes in *Dactylorhiza* (Orchidaceae). *Taxon*, *56*(4), 1185–1208.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS One*, *5*(3), e9490.

Pridgeon, A. M., Cribb, P. J., Chase, M. W., & Rasmussen, F. N. (2001). *Genera Orchidacearum: Volume 2. Orchidoideae (Part 1)*. Oxford University Press.

Pridgeon, A. M., Cribb, P. J., Chase, M. W., & Rasmussen, F. N. (2003). *Genera Orchidacearum: Volume 3. Orchidoideae (Part 2), Vanilloideae*. Oxford University Press.

Ramírez, S. R., Gravendeel, B., Singer, R. B., Marshall, C. R., & Pierce, N. E. (2007). Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature*, *448*(7157), 1042–1045.

Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N., & Delsuc, F. (2018). MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Molecular Biology and Evolution*, *35*(10), 2582–2584.

Russo, A., Alessandrini, M., El Baidouri, M., Frei, D., Galise, T., Gaidusch, L., Oertel, H., Morales, S. G., Potente, G., Tian, Q., Smetanin, D., Bertrand, J., Onstein, R., Panaud, O., Frey, J., Cozzolino, S., Wicker, T., Xu, S., Grossniklaus, U., & Schlüter, P. (2023). The genome of the early spider-orchid Ophrys sphegodes provides insights into sexual deception and adaptation to pollinators. *Research Square*. https://doi.org/10.21203/rs.3.rs-3463148/v1

Sayyari, E., & Mirarab, S. (2018). Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes*, *9*(3), 132. https://doi.org/10.3390/genes9030132

Slimp, M., Williams, L. D., Hale, H., & Johnson, M. G. (2021). On the potential of Angiosperms353 for population genomic studies. *Applications in Plant Sciences*, *9*(7), e11419. https://doi.org/10.1002/aps3.11419

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, *32*(5), 1342–1353.

Tekinşen, K. K., & Güner, A. (2010). Chemical composition and physicochemical properties of tubera salep produced from some Orchidaceae species. *Food Chemistry*, *121*(2), 468–471.

Wettewa, E., & Wallace, L. E. (2021). Molecular phylogeny and ancestral biogeographic reconstruction of *Platanthera* subgenus *Limnorchis* (Orchidaceae) using target capture methods. *Molecular Phylogenetics and Evolution*, *157*, 107070.

Wolfe, T. M., Balao, F., Trucchi, E., Bachmann, G., Gu, W., Baar, J., Hedrén, M., Weckwerth, W., Leitch, A. R., & Paun, O. (2023). Recurrent allopolyploidizations diversify ecophysiological traits in marsh orchids (*Dactylorhiza majalis* s.l.). *Molecular Ecology*, *32*(17), 4777–4790.

Wong, D. C. J., & Peakall, R. (2022). Orchid Phylotranscriptomics: The prospects of repurposing multi-tissue transcriptomes for phylogenetic analysis and beyond. *Frontiers in Plant Science*, *13*, 910362.

Wong, D. C. J., Perkins, J., & Peakall, R. (2022). Conserved pigment pathways underpin the dark insectiform floral structures of sexually deceptive Chiloglottis (Orchidaceae). *Frontiers in Plant Science*, *13*, 976283.

Wortley, A. H., Rudall, P. J., Harris, D. J., & Scotland, R. W. (2005). How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. *Systematic Biology*, *54*(5), 697–709.

Wu, T. D., Reeder, J., Lawrence, M., Becker, G., & Brauer, M. J. (2016). GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. *Methods in Molecular Biology*, *1418*, 283–334.

Xu, B., & Yang, Z. (2016). Challenges in species tree estimation under the multispecies coalescent model. *Genetics*, *204*(4), 1353–1368.

Yang, Y., & Smith, S. A. (2014). Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution*, *31*(11), 3081–3092.

Yardeni, G., Viruel, J., Paris, M., Hess, J., Groot Crego, C., de La Harpe, M., Rivera, N., Barfuss, M. H. J., Till, W., Guzmán-Jacob, V., Krömer, T., Lexer, C., Paun, O., & Leroy, T. (2022). Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Molecular Ecology Resources*, *22*(3), 927–945.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(Suppl 6), 153.

Zhang, D., Zhao, X.-W., Li, Y.-Y., Ke, S.-J., Yin, W.-L., Lan, S., & Liu, Z.-J. (2022). Advances and prospects of orchid research and industrialization. *Horticulture Research*, *9*, uhac220.

Zhang, G., Hu, Y., Huang, M.-Z., Huang, W.-C., Liu, D.-K., Zhang, D., Hu, H., Downing, J. L., Liu, Z.-J., & Ma, H. (2023). Comprehensive phylogenetic analyses of Orchidaceae using nuclear genes and evolutionary insights into epiphytism. *Journal of Integrative Plant Biology*, *65*(5), 1204–1225.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: 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. *Molecular Ecology Resources*, *00*, e13986. https://doi.org/10.1111/1755-0998.13986