The genome of the early spider-orchid *Ophrys sphegodes* provides insights into sexual deception and adaptation to pollinators

Alessia Russo
University of Zürich https://orcid.org/0000-0003-3856-5317
Mattia Alessandrini
University of Hohenheim
Moaine El Baidouri
French National Centre for Scientific Research https://orcid.org/0000-0002-5023-3457
Daniel Frei
Agroscope
Teresa Galise
University of Naples Federico II
Lara Gaidusch
University of Hohenheim
Hannah Oertel
Wageningen University https://orcid.org/0009-0005-4241-1232
Sara Garcia Morales
University of Hohenheim
Giacomo Potente
University of Zurich
Qin Tian
Naturalis Biodiversity Centre
Dmitry Smetanin
University of Zurich
Joris Bertrand
Université de Perpignan Via Domitia
Renske Onstein
Naturalis Biodiversity Centre
Olivier Panaud
Plant Genome and Development Laboratory https://orcid.org/0000-0002-9292-503X
Jürg Frey
Agroscope
Salvatore Cozzolino

University of Naples Federico II

Thomas Wicker

University of Zurich https://orcid.org/0000-0002-6777-7135

Shuqing Xu

University of Mainz

Ueli Grossniklaus

University of Zurich https://orcid.org/0000-0002-0522-8974

Philipp Schlüter

philipp.schlueter@uni-hohenheim.de

University of Hohenheim https://orcid.org/0000-0002-6057-0908

Article

Keywords:

Posted Date: November 15th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3463148/v1

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on July 26th, 2024. See the published version at https://doi.org/10.1038/s41467-024-50622-4.

Abstract

Pollinator-driven evolution of floral traits is thought to be a major driver of angiosperm speciation and diversification. The bee orchids (genus *Ophrys*) mimic their pollinators' female pheromone, shape and colour to lure male pollinators into pseudocopulation. This strategy, called sexual deception, is highly species-specific, thereby providing strong premating reproductive isolation. Identifying the genomic architecture underlying pollinator adaptation and speciation may shed light on the mechanisms of angiosperm diversification. We report the 5.2 Gb chromosome-scale genome sequence of *Ophrys sphegodes* and show that chromosomal rearrangements took place in the *Ophrys* lineage. We find evidence for a transposable element expansion event that preceded the radiation of the *O. sphegodes* group, and for gene duplication having contributed to the evolution of chemical mimicry. Moreover, one region on chromosome 2 is highly differentiated between *O. sphegodes* and other *Ophrys* species, indicating a candidate genomic region for pollinator-mediated evolution.

Main

Understanding the genetic mechanisms of adaptation to pollinators is a central question in plant evolutionary biology. The prominent role of pollinators in flower evolution is related to their dual function: enabling sexual reproduction and imposing selection on floral traits[1,2]. This is particularly evident in plant species with a specialised pollination mechanism associated with floral traits that evolved to attract one or a few functionally alike pollinator species[3,4]. Therefore, pollinator-mediated evolution of floral traits is considered a major force driving angiosperm diversity by contributing to their radiation[5,6]. It has been hypothesised that radiations are facilitated by entering a new ecological niche with little or no competition from similar species (ecological opportunity), and the genetic potential allowing the necessary adaptations to evolve (genetic variation), ultimately leading to rapid bursts of speciation[7–9].

Orchids of the Mediterranean genus *Ophrys* ensure their reproduction through sexual deception, that is, they mimic the olfactory, visual and tactile signals of females of their pollinating insects to entice conspecific males to pseudocopulate with the flower, leading to pollination (Fig. 1a-b). The extreme specialisation of *Ophrys* floral traits makes the plant-insect interaction highly species-specific, and the lack of shared pollinators between species leads to strong premating reproductive isolation[10,11]. The genus *Ophrys* is of relatively recent origin (c. 4.9 Ma crown age), with the earliest-diverged *Ophrys* lineage likely having been wasp-pollinated[12,13]. Two independent *Ophrys* lineages later experienced a pollinator shift to *Andrena* bee pollinators, preceding a burst of speciation in the *O. sphegodes* group within the last million years[12]. This resulted in a species radiation with one of the highest reported diversification rates among angiosperms[12]. While several facets of *Ophrys* speciation are relatively well understood[10,11,12], the genomic features allowing its adaptation to diverse pollinators are largely unknown. Our ability to elucidate how genome architecture has contributed to *Ophrys* diversification and adaptation to pollinators has hitherto been hampered by the lack of a reference genome. Here, we present the chromosome-level genome assembly of *Ophrys sphegodes*, a key representative of the

genus and the adaptive radiation, to address questions on the genetic mechanisms underlying such rapid pollinator-driven speciation.

Results and discussion

The chromosome-level genome assembly

The orchid *Ophrys sphegodes* Mill. has a karyotype of 2*n* = 2x = 36 [10,14,15]. The haploid genome size was estimated to be 4.83 Gb by flow cytometry, and heterozygosity was estimated at 1.28% via *k-mer* analysis (Supplementary Figs. 1 and 2, Supplementary Table 1). To assemble the genome, we generated a total of 409 Gb data on the Nanopore PromethION platform (Supplementary Table 2, Supplementary Fig. 3) and sequenced Hi-C chromatin conformation capture libraries (Supplementary Table 3, Supplementary Fig. 4). We used Miniasm assembly[16] to generate a total of 11 148 contigs in 6.4 Gb, with an N50 value of 754 kb (Supplementary Table 4). We removed under-collapsed heterozygous contigs (1.2 Gb) and used Hi-C data to anchor the scaffolds into 18 pseudomolecules corresponding to the 18 chromosomes expected for the haploid genome (Fig. 1c, Supplementary Fig. 5a; Supplementary Table 5). The final assembled genome size was 5.2 Gb, in line with the estimated genome size, with a scaffold N50 of 218 Mb, L50 of 10 and L75 of 17 (Supplementary Table 4). Overall, 97.8% of raw Illumina sequence reads could be mapped to the assembly, suggesting that our assembly contains the complete genetic information (Supplementary Table 6). Gene region completeness was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs)[17]: 1453 out of 1612 (91.1%) conserved core land plant genes were found in our assembly, of which 1361 (84.9%) were complete (Supplementary Fig. 6a). We confidently annotated a total of 42 549 protein-coding genes, of which 90.01% had functional annotation or RNA-seq support (see Supplementary Tables 7–10 for genes of interest). Cytosine methylation was inferred from Nanopore sequencing data (Fig. 1c, Supplementary Fig. 7).

A burst in transposon activity preceded adaptive radiation

We manually characterised transposable elements (TEs) in the genome following a protocol[18] for TE classification that characterises TEs through a combination of homology-based prediction and manual inspection to find structural motifs and define sequence boundaries. To this end, we created the first species-specific database of TEs of an orchid genome, containing a total of 436 novel sequences specific to *O. sphegodes.* Using this database, we identified a total of 4.05 Gb repetitive elements, occupying 78.06% of the *O. sphegodes* genome (Fig. 1c, Supplementary Table 11). Overall, *O. sphegodes* represents the largest orchid genome assembled to date and exhibits the highest abundance of long terminal repeat (LTR) elements (74.57% of the genome), more than *Vanilla planifolia* (10.00%)[19], *Apostasia shenzenica* (16.81%)[20], *Dendrobium catenatum* (39.88%)[20], *Phalaenopsis equestris* (44.19%)[20], *Gastrodia elata* (54.75%)[21], *Cymbidium sinense* (54.76%)[22], and similar to *Platanthera guangdongensis* and *P. zijinensis* (73.23% and 71.78%[23]; Supplementary Table 12). Transposon activity is known to influence genome size variation in eukaryotes[24], and particularly LTR/*Gypsy* and LTR/*Copia*

have previously been shown to correlate with genome size expansion in orchids[25,26]. In *O. sphegodes*, analysis of recent LTR insertions shows that LTR activity had an initial increase at around 3 Ma, to reach its maximum at around 1.3 to 0.8 Ma (Fig. 1d). During this period, the Mediterranean Basin experienced climatic oscillations with glacial/interglacial periods[27,28] that may have acted as environmental disturbances[29,30] and/or facilitated hybridisation[31] after distribution range shifts. Any such event might have triggered bursts of TE proliferation in *O. sphegodes*, thus inflating its genome size. Interestingly, the peak of LTR element insertions precedes or overlaps with the radiation of the most species-rich *Ophrys* lineage, including the *O. sphegodes* species group, less than 1.0 Ma[12]. The ability to cope with genome size changes has allowed angiosperms to successfully diversify[32] and TEs have played an important role in enhancing angiosperm evolution[33] through their effect upon gene expression[30], as well as gene duplication and genomic rearrangements[34,35]. By contributing to the generation of intraspecific genetic diversity[36], TE bursts may have provided the *O. sphegodes* lineage with the genetic capacity to adapt to a new pollinator niche (i.e., *Andrena* bees), thereby facilitating the adaptive radiation of the *O. sphegodes* group.

Genome evolution through chromosome fusions in the Ophrys lineage

To understand the evolutionary history of the *Ophrys* lineage, we constructed a phylogenomic tree and estimated divergence times across *O. sphegodes* and 20 other plant species with fully sequenced genomes, based on single-copy orthologues. *Ophrys* diverged from *Platanthera* (both from Orchidoideae subtribe Orchidinae), the most closely related orchid with a fully sequenced genome[23,37,38], approximately 22.42 Ma ago with a 95% confidence interval (CI) of 20.77–23.95 Ma, and Orchidoideae separated from other orchids around 54.49 Ma (CI 53.00–56.15 Ma; Fig. 2a, Supplementary Fig. 8). Our analysis further suggests that Orchidaceae separated from the common ancestor of Asparagales approximately 99.96 Ma ago (CI 98.66–105.66 Ma) and places monocot/eudicot divergence around 147.51 Ma (CI 139.87-154.45 Ma), in line with previous studies[39,40].

To track chromosome evolution of *O. sphegodes*, we compared it with the most closely related sequenced orchid genomes[23], focusing on the comparison with *Platanthera zijinensis* (Fig. 2b). The two genera differ in their chromosome numbers, the karyotype organisation of *Platanthera* (n = 21)[41] reflecting the ancestral and *Ophrys* (n = 18) the derived state[37,42]. Overall, most chromosomes maintained their structure between *Platanthera* and *Ophrys*, but some notable major rearrangements are apparent, particularly with regard to chromosome fusions. Chromosome (chr) 4 in *O. sphegodes* appears to be the product of a fusion between chr 7 and part of chr 4 in *Platanthera*, and *Ophrys* chr 10 derives from a fusion between *Platanthera* chr 15 and part of chr 14. Moreover, *Platanthera* chr 8 has no homologous chromosome in *Ophrys*, and significant parts of *Platanthera* chr 3, 4, 11, 14, 20 and 21 lack syntenic regions in *O. sphegodes* (Fig. 2b and Supplementary Fig. 9). Taken together, these findings are consistent with a reduction in chromosome number via fusions in the *Ophrys* lineage.

Expanded gene families may be involved in flower development

To gain insights into the genomic basis of the unique pollination system, we first identified orthologous gene families using OrthoFinder[43]. We identified a total of 495 819 genes in 26 709 orthogroups among the 21 species, of which 3 054 are shared among all species. A total of 1 351 families containing 4 537 genes were unique to *O. sphegodes*. We then identified expanded and contracted gene families using CAFE[44] (Fig. 2a). In *O. sphegodes*, 3 712 gene families underwent an expansion, whereas 756 underwent a contraction. This is the highest level of gene family expansion reported among orchids so far, followed by *V. planifolia* (+ 3 248) and *D. catenatum* (+ 1 817). Of those gene families, 291 and 59 exhibited significant (p-value ≤ 0.01) expansions and contractions, respectively. Among the significantly expanded gene families were genes encoding transcription factors (TFs) involved in plant reproduction and flower development, such as MADS-domain TFs (55 genes encoding type I and 6 genes of type II MADS; Supplementary Fig. 10a), MYB TFs (73 genes, of which 32 MYB-P; Supplementary Fig. 10b), LATERAL ORGAN BOUNDARIES (LOB) domain TFs (10 genes), C2C2-GATA TFs (12 genes), WRKY TFs (11 genes) involved in trichome development[45], SERINE CARBOXYPEPTIDASE-LIKE-1 (SCPL-1) proteins (15 genes) controlling anthocyanin acylation[46], TCP TFs (14 genes) regulating flavonoid biosynthesis and floral symmetry[47,48], and YABBY TFs (6 genes) involved in establishment of adaxial-abaxial polarity[49]. Secondly, disease resistance-related and stress response genes were found, such as NAC TFs (30 genes), glutathione S-transferases (28 genes) and plant PLEIOTROPIC DRUG RESISTANCE (PDR) proteins (10 genes). Finally, two putative candidate genes for floral odour production and anthocyanin biosynthesis[50] showed significant expansion, too, namely fatty acyl-CoA reductases (FARs) and chalcone synthases, also involved in defence response, (29 and 14 genes, respectively).

Plant adaptation to pollinators in Ophrys is associated with local duplication of candidate genes

The key component for pollinator attraction in *Ophrys* species is the chemical mimicry of the pollinator female's sex pheromone, the composition of which has previously been characterised for *O. sphegodes s.l.*[51,52]. Here, alkene hydrocarbons are especially important, as different proportions of (*Z*)-12-alkenes, (*Z*)-9-alkenes and (*Z*)-7-alkenes are the major odour differences among *O. sphegodes* and the closely related *O. exaltata*, responsible for attracting different pollinators[53,54]. Thus, genes involved in hydrocarbon biosynthesis are likely important for pollinator attraction[50]. We annotated previously identified candidate genes[50] in the genome of *O. sphegodes* (Supplementary Fig. 11, Supplementary Table 7). Among these, structural annotation of stearoyl-ACP desaturases (SADs) shows that the key genes, *SAD1* and *SAD2* (*SAD2*-type), are duplicated in tandem in a single cluster on chromosome 4 (283.17–283.30 Mb; containing four copies; Fig. 3d), whereas the house-keeping desaturase encoding *SAD3* [55] resides in one copy in chromosome 5 (280.45–280.47 Mb) and *SAD4* (*SAD5*-type) is also present as a single copy (scaffold 75: 1.04–1.11 Mb; Supplementary Fig. 11; Supplementary Table 7). For other *O. exaltata SAD5*-type genes, we identified four full-length copies in the *O. sphegodes* genome (at

least two of them on the same scaffold 210), although none of these appeared to be a functional copy (Supplementary Table 7). This is in line with previous findings that *O. sphegodes* only expresses functional *SAD2*-type alleles, while functional *SAD5*-type alleles are expressed in *O. exaltata*[54,55]. Phylogenetic analysis confirms the presence of separate *SAD* gene lineages corresponding to *SAD3, SAD2*-type (*SAD1/2/7/8*), and *SAD5*-type (*SAD4/5/6/9/10*, an incomplete *SAD11*) genes (Fig. 3a). Both *SAD2-* and *SAD5*-type lineages appear as single copies in *Platanthera*, suggesting they underwent recent gene duplication events. Of note, *SAD2* homologues were only present in orchid genomes from the subfamily Orchidoideae.

Like SADs, membrane-bound fatty acid desaturases (FADs)[56], which contribute to fatty acid desaturation and, potentially, alkene production[57], were also found duplicated in the *O. sphegodes* genome. We found a phylogenetic lineage containing four *FAD* gene copies, of which at least three reside on chromosome 1 (368.58–368.63 Mb; Fig. 3e), a second lineage of two copies clustered on scaffold 33, while one gene, *FAD2*, was not duplicated. Both lineages showing gene duplications are present as single copies in the *Platanthera* genome (Fig. 3b).

Fatty acyl-CoA reductase (FAR) homologues show gene duplication in *Ophrys.* FARs likely catalyse the conversion of fatty acyl-CoA to primary alcohols and different FARs produce fatty alcohols with different acyl chain lengths[58]. In the *O. sphegodes* genome, 16 *FAR* homologues were found, of which 15 gene models form a phylogenetic Orchidoideae lineage together with a single *Platanthera* sequence. Of these, 14 genes are distributed over only three scaffolds (Fig. 3c; chr 5: 1; scaffold 279: 4; scaffold 133: 4; scaffold 578: 6). Six of these clustered *FAR* genes reside in one region of 109 kb on scaffold 578 (Fig. 3f). The close vicinity of these copies in a unique clade shows a recent duplication event in the *O. sphegodes* lineage (Fig. 3c). Gene duplication plays a crucial role in shaping the evolutionary landscape of genomes, as they provide the main raw material for the evolution of new genes[59]. Single or tandem gene duplications are also involved in the origin of many plant genes[60]. Often, retention of duplicate genes occurs non-randomly, as changes in the concentration of gene products can have a selective advantage for the organism[61]. These novel genes may provide *Ophrys* with an opportunity to respond to selection by pollinators, e.g., through positive dosage effects or neofunctionalisation and pseudogenisation of the less effective variants[62].

Population genomic analyses reveal putative barrier loci under pollinator-driven selection

Closely related *Ophrys* species provide plausible examples of pollinator driven-speciation[11,63]. We therefore investigated the genetic differentiation between *O. sphegodes* and three other closely related, sympatric co-flowering species, *O. exaltata, O. garganica* and *O. incubacea* in Gargano, South Italy[10,11]. These four species are pollinated by sexual deception of different solitary bee males: *Andrena nigroaenea, Colletes cunicularius, A. pilipes, A. morio*, respectively[64]. These *Ophrys* species show variation in floral traits, ranging from labellum coloration (markedly blacker in *O. garganica* and *O.*

incubacea[11]), to different floral odours mimicking their pollinators' sex pheromones, which majorly contribute to reproductive isolation[10,11,52,54,65]. We used Genotype-By-Sequencing (GBS) data to investigate the genetic differentiation between species and scan the genome for signatures of selection. Chord between-population genetic distance[66] was calculated per 1 Mb window to identify the most similar/dissimilar species at a given chromosomal region in the *O. sphegodes* genome. This analysis revealed that segregating polymorphisms between species are distributed across the genome, and that overall genetic (dis)similarity between *O. sphegodes* and the three other species is roughly equal (Fig. 4a, Supplementary Fig. 12), as may be expected in a species radiation. Yet cumulatively, the four species were clearly separable in a principal coordinate analysis (PCoA; Fig. 4b). These findings are in line with previous population analyses suggesting that many polymorphisms in the genome are shared among all species[11], whilst few barrier loci may separate them. Furthermore, global F_{ST} outlier analysis (FDR < 0.01) of a GBS dataset from 126 individuals[11] revealed highly differentiated regions in chr 2 (327–346 Mb; Fig. 1c and 4c). Interestingly, in contrast to the genome-wide pattern, especially *O. sphegodes* and *O. exaltata* are clearly separated in the highly differentiated c. 20 Mb region on chr 2, as seen by PCoA of this region (Fig. 4d). Whether this pattern is due to divergent selection and suppression of effective recombination via e.g., divergence hitchhiking or an inversion, is currently unknown. Overall, using the metaphor of genomic islands of speciation[67], these results suggest that adaptation involving individual loci (or small genomic segments) rather than entire genomes, characterises these reproductively isolated sympatric species. The overall high level of allele sharing and the genomic mosaic pattern of species relationships (without long contiguous stretches of similar relationships) cannot easily be explained by gene flow after a secondary contact of species that are separated by strong floral isolation[10]. An alternative explanation could be that the high level of segregating polymorphisms is due to shared ancestry and large effective population sizes, suggesting that the species are in an early stage of genomic divergence[10,53,67,68]. Our population genomic analyses identified candidate regions potentially under pollinator-driven selection, thus calling for future research into the roles and functions of the genes in these regions.

Conclusions

The chromosome-level genome sequence of *O. sphegodes*, a plant with high pollinator specificity, provides important insights into the evolution of plant adaptation to pollinators and its role in species diversification. It seems that, ecologically, the recent adaptive radiation of this group was fuelled by the availability of pollinator niches and, genetically, a burst in TE activity and rampant gene duplication provided the genetic raw material for pollinator-mediated selection to act upon. These mechanisms may also provide a blueprint for other angiosperm radiations, as floral trait-mediated reproductive isolation may often be underlain by a simple genetic architecture[2]. More broadly, this case study supports the idea that the generation of genomic diversity, as genomic potential in the form of genome duplication, hybridisation or TE activity, often precedes adaptive radiation.

Methods

Sample preparation

The *O. sphegodes* individual for the genome assembly (accession SPH_8) was selected among several samples previously collected in Capoiale, Gargano area, South Italy[10], and grown in a pot at the Dept. of Biology, University of Naples Federico II, under natural light at ambient temperature. Fresh young leaves were harvested, snap-frozen, and genomic DNA was isolated as described[69].

Genome size estimation

Nuclear DNA content and thus genome size was estimated via flow cytometry. Since genome size estimation via flow cytometry from orchid leaves is challenging due to DNA endoreduplication[70], we used pollinia (carrying haploid pollen) for the analysis. We followed a previously published protocol[71] with slight modifications. Briefly, a pair of pollinia was crushed in a 1.5 ml tube with Otto I buffer using a clean pestle, transferred to a Petri dish, and co-chopped with 2×2 cm tissue of a reference leaf (*Solanum lycopersicum* cultivar 'Stupicke polni tyckove rane'; tomato 1C = 0.98 pg, as measured in ref. [72]). The suspension was filtered, mixed with Otto II buffer, and stained with propidium iodide in the dark at 4°C for 1 hour. At least 10 000 nuclei were analysed on a Cytoflex S (Beckman Coulter) flow cytometer. The average fluorescence value (Mean PerCP-A, see Supplementary Fig. 1) of reference 2C nuclei was 4 207 840; the average fluorescence for *O. sphegodes* 1C nuclei was 10 506 736 (Supplementary Table 1). Thus, following a published formula[73], and converting picograms of DNA to number of nucleotide pairs[74], we estimated a haploid genome size of 4.83 Gb (average of three measurements, see Supplementary Table 1).

Nanopore library construction, sequencing, and genome assembly

Isolated DNA was used to prepare eight ONT libraries (Supplementary Table 2), following the general guidelines provided by Oxford Nanopore Technologies® for the 1D Genomic DNA by Ligation (SQK-LSQ109) protocol, with modifications proposed by New England Biolabs® (NEB)[69]. Six libraries were sequenced on a PromethION PTC0031 platform (Oxford Nanopore Technologies), and two on MinION Mk1B. All sequencing was performed at Agroscope Wädenswil, Switzerland. We used Miniasm 0.3 and Minimap2 v2.17[16] to assemble the initial contigs. Pilon v1.23[75] and Racon v1.4.3[76] were used to correct indels and mis-assemblies with Illumina reads (Supplementary Table 3). To deal with the heterozygosity of the *O. sphegodes* genome, we used Redundans v0.11[77] to remove under-collapsed contigs and obtain a final assembled genome size of 5.2 Gb (see Supplementary Information). Raw Nanopore sequencing data were base-called with Guppy v5.0.11 and Nanopolish v0.13.3 was used to detect 5-methylcytosine (see Supplementary Information).

Hi-C library preparation and assembly scaffolding

A total of 0.5 g leaf tissue were fixed in nuclei isolation buffer and 36% formaldehyde, followed by cell lysis, chromatin digestion with the enzyme *Hin*dIII-HF (New England Biolabs, Ipswich, MA, USA, R3104L), re-ligation, DNA re-extraction and library preparation as described[78,79]. Four Hi-C libraries were sequenced on an Illumina NovaSeq 6000 SP FlowCell. To assemble the chromosomes, raw Hi-C data was mapped against the genome assembly using the ArimaGenomics mapping script (https://github.com/ArimaGenomics/mapping_pipeline/blob/master/01_mapping_arima.sh). Subsequently, Salsa v2.3[80,81] was used to build scaffolds (total scaffolds 2 511, N50 = 4.8 Mb); Juicebox Assembly Tools (JBAT)[82] and the 3D-DNA pipeline v-180114[83] were used to order scaffolds to chromosomes.

Gene and non-coding RNA annotation

Gene annotation was carried out via an integrative approach that included *de novo* gene prediction, homology-based prediction, and transcriptome-based prediction. AUGUSTUS was used as part of the BRAKER2 pipeline[84] and trained with RNA-seq data[13] from floral tissues to predict coding regions in the repeat-masked genome. ProtHint (https://github.com/gatech-genemark/ProtHint) was used to generate protein hints using the protein database liliopsida_odb10-v.2020-09-10 from OrthoDB10, to score intron intervals, start and stop codons from ultra-conserved proteins of the monocot lineage. This extrinsic evidence was given to GeneMark[85] to self-train, and improve prediction accuracy of AUGUSTUS. Proteomes from *P. equestris*[20], *C. goeringii*[86], *O. sativa* ssp. *japonica* cultivar Nipponbare IRGSP-1.0[87], and *A. officinalis* Aspof.V1[88] were downloaded and used as input for GeMoMa[89] for homology-based prediction. The transcriptome of *O. sphegodes*[13] was mapped against the genome using PASA (Program to Assemble Spliced Alignment) https://github.com/PASApipeline/PASApipeline/blob/master/docs/index.asciidoc). Valid transcript alignments were clustered based on genome mapping location and assembled into gene structures. PASA assemblies were incorporated into gene predictions to correct exon boundaries, add UTRs, and update gene structures. AHRD (https://github.com/groupschoof/AHRD) was used to filter genes based on SwissProt, trEMBL, and TAIR10 databases, and candidate genes for pollinator attraction from previous studies were manually included. Functional annotation was carried out using TRAPID[90] for Gene Onthology (GO) terms, InterPro-Scan, KEGG, and iTAK Classifier[91] for transcription factors and protein kinases. Genes for tRNA and rRNA were identified using tRNAscan-SE[92] and Barrmap v0.9 (https://github.com/tseemann/barrnap; see Supplementary Information for details).

Repetitive element annotation

We used a published approach[18] to classify and manually annotate repetitive elements in the genome. Putative TEs identified with RepeatModeler v2.0.1 were blasted against the TRansposable Elements Platform (TREP) database (v.2019, https://trep-db.uzh.ch/). Sequences with strong hits to a known family were selected (plus their flanking regions) and blasted against the *O. sphegodes* genome to identify other copies belonging to the same TE family. Multiple sequence alignments between those clusters of sequences were performed to create consensus TE sequences, and to search for structural

motifs that could help classify them. Sequences with no strong hits to a known family were blasted against PTREP (Protein TRansposable Elements Platform) or Dfam (https://dfam.org) or examined for the presence of structural motifs to identify novel TEs in the genome. All novel TE sequences were classified according to a three-letter code[18].

Analysis of LTR divergence and insertion time

LTR_harvest[93] was used to identify LTRs with complete structure (-xdrop 37 -motif tgca -motifmis 1 -minlenltr 100 -maxlenltr 3000 -mintsd 2), expected to be detectable for young LTR copies. The insertion time was estimated based on the sequence divergence between two LTRs of the same type. 5' and 3' LTR of each full-length paralogue were aligned using Mafft[94] and divergence between them was estimated using the Kimura two-parameter model using the distmat program implemented in the EMBOSS package (https://emboss.sourceforge.net/). LTRs insertion age was estimated using a substitution rate of 1.3×10^{-8} substitutions per site per year[95].

Phylogenomic analysis and estimation of divergence time

The protein-coding regions of 21 plant species with a sequenced genome (Supplementary Table 17) were used for phylogenomic analysis. Orthologue groups were identified using OrthoFinder v2.5.4[43], using Mafft for multisequence alignment and fastTree[96] for building the trees. Among 526 955 genes from the 19 species, 495 819 (92.1%) were assigned into 26 709 orthologue groups. The concatenated supermatrix of 34 single-copy nuclear genes was used to infer phylogenetic relationships using the GTR-GAMMA model with 1000 bootstrap replicates in RAxML v8.2.11[97]. The same supermatrix and corresponding concatenated ML tree were used for dating analysis. We estimated divergence times using the penalised likelihood method implemented in TreePL[98]. We considered 83.78 Ma and 114.92 Ma as the minimum- and maximum-age calibrations of the stem age of orchids[99]. We used an optimal smoothing parameter determined by the "random subsample and replicate" cross-validation method to accommodate rate heterogeneity. A total of 1000 bootstrap replicates with the topology fixed to the concatenated ML tree and branch lengths allowed to vary were also generated using RAxML for calculating the confidence intervals of age estimates. Results from the dating of the bootstrapped trees were then summarised and visualised on the concatenated ML tree using TreeAnnotator (part of the BEAST2 package[100]).

We used CAFE v5.0.0[44] to identify gene family contractions and expansions, using the reconstructed species tree. Only gene families were kept in which among-species difference in size was less than 100. First, nucleotide sequences of the previously identified single-copy orthologues were translated into amino acids with TranslatorX v1.1 within the MitoPhAST-master v3.0 package[101], then given as input for MUSCLE v5.[102] for multiple sequence alignment (MSA). The resulting MSAs were trimmed with trimAL v1.2[103] (parameter -automated1 to select optimal cut-off based on alignment similarity) to remove poorly aligned regions or spurious sequences, and the best-fit model for each tree was computed with ModelTest-NG v0.1.7[104] (parameters -f ef -h f -s 3 -tr phyml). The final trees were built with phyml

v3.3.20220408 (parameters -d nt -c 4 -m AIC -f modelfreq -o modelparam) (https://github.com/stephaneguindon/phyml) using the parameters from ModelTest-NG and 1001 bootstrap replicates.

Comparative genomics analysis

Synteny analyses between *O. sphegodes* and *P. guangdongensis* and *P. zijnensis* were performed using the MCScan[105] functions of the JCVI utility libraries v1.2.9 (https://github.com/tanghaibao/jcvi). First, *jcvi.compara.catalog ortholog* (--min_size 5 -dist 35) used LAST v2.34[106] to identify syntenic blocks between the *O. sphegodes* genome and another genome. Second, *jcvi.compara.synteny screen* (--minspan 30 -simple) was used to create simplified versions of the anchor files containing the syntenic blocks. Finally, linear synteny plots were made with *jcvi.graphics.karyotype* and dot-plots with *jcvi.graphics.dotplot*.

Population genetic analyses

Previously published GBS data from 126 individuals[11] and RNA-seq data from floral tissues of 32 accessions[13] were used to analyse the genetic variation between *Ophrys* species. Both datasets were mapped against the genome using BWA-MEM2 v2.[107] with default parameters, and variants (SNPs and small InDels) were identified using freebayes v1.0.2 (https://github.com/freebayes/freebayes) with minimum depth 2 (-C 2). Samtools v1.10[108] was used to remove duplicated reads prior to variant calling. Raw variants (1 680 249, of which 1 621 433 biallelic GBS markers) were filtered for depth and frequency, and pairwise genetic distances calculated between individuals using SPA v0.1 (https://peb.uni-hohenheim.de/spa) as described in Supplementary Information. Biallelic SNPs were used with BayeScan v2.1[109] to identify candidate loci under selection in a “global” manner, i.e., treating each species as a subpopulation, and in a “pairwise” fashion, comparing two species at time. Chord distance was calculated according as described.[66,110] PCoA was carried out genome-wide and per chromosome. A detailed description of population genetic methods can be found in Supplementary Information.

Declarations

Acknowledgements

We are grateful to Martin Mascher and Chang Liu for fruitful discussions about the project, and to Noa Schwabe for kindly providing pictures of *Ophrys* and *Andrena.*

Funding

This work was supported by the University of Zurich, a Ph.D. project funded by the University Research Priority Program “Evolution in Action: from Genomes to Ecosystems” (to UG and PMS), the Swiss National Science Foundation (SNSF; grant 31003A_155943 to PMS and grant IZLRZ3_163885 to UG), the German Research Foundation (DFG; project 446145319 to PMS) and the 2022 PRIN Program (to SC). This work made use of infrastructure services provided by the Science IT team of the University of Zurich

(www.s3it.uzh.ch) and was supported via bwHPC by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, state of Baden-Württemberg, and the DFG (grant INST 37/935-1 FUGG).

Author contributions

A.R., P.M.S. and U.G. conceived and designed this study. A.R., P.M.S. and S.C. collected *Ophrys* samples. A.R. performed genome size estimation, gDNA extraction and the Hi-C experiment. A.R. and D.S. conducted the initial genome assembly. A.R. improved the assembly and performed annotation analysis. D.F. performed Nanopore library preparation and sequencing. A.R. and M.A. performed the JBAT visualisation to final chromosome scaffold. A.R., T.R.G., L.G. and T.W. conducted the transposable element annotation. M.E.B. conducted the LTR divergence time analysis. H.F.O. performed the methylation analysis. S.X., Q.T., R.E.O. and S.E.G.M. carried out the phylogenetic and gene family analyses. G.P. conducted the synteny analysis. A.R. and P.M.S. performed the population genomic analyses. A.R. and M.A. performed figure curation. A.R. wrote the original draft of the manuscript with inputs from P.M.S., S.C., S.X. and U.G. All authors discussed and approved the final manuscript. J.E.F., J.B., S.C., O.P. and U.G. provided critical resources for this study. P.M.S. and U.G. supervised the project. P.M.S., U.G., S.C. and J.E.F. acquired funding.

Data availability

The genome assembly and raw sequencing data generated for this study, including ONT data, Illumina data and HiC data, were submitted to NCBI under BioProject number PRJNA994461 [@Reviewers, we are in the process of completing the NCBI submission under this accession number]. The RNA-seq data used for genome annotation can be found on NCBI under accession PRJNA574279 [13]. The GBS data used for population genetic analysis can be found in the NCBI accession PRJNA257331 [11].

References

1. Waser, N. M. Pollination, angiosperm speciation, and the nature of species boundaries. *Oikos* **82**, 198–201 (1998).
2. Schlüter, P. M. The magic of flowers or: speciation genes and where to find them. *Am. J. Bot.* **105**, 1957–1961 (2018).
3. Armbruster, W. S. Floral specialization and angiosperm diversity: phenotypic divergence, fitness trade-offs and realized pollination accuracy. *Ann. Bot. Plants.* **6**, plu003 (2014).
4. Gómez, J. M. & Zamora, R. Generalization vs. specialization in the pollination system of *Hormathophylla spinosa* (Cruciferae). *Ecology* **80**, 796–805 (1999).
5. Van der Niet, T. & Johnson, S. D. Phylogenetic evidence for pollinator-driven diversification of angiosperms. *Trends Ecol. Evol.* **27**, 353–361 (2012).
6. Grant, V. Pollination systems as isolating mechanisms in Angiosperms. *Evolution.* **3**, 82–97 (1949).

7. Bouchenak-Khelladi, Y., et al. On the complexity of triggering evolutionary radiations. *New Phytol.* **207**, 313–326 (2015).
8. Seehausen, O. Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytol.* **207**, 304–312 (2015).
9. Nürk, N. M., et al. Diversification in evolutionary arenas—Assessment and synthesis. *Ecol. Evol.* **10**, 6163–6182 (2020).
10. Xu, S., et al. Floral isolation is the main reproductive barrier among closely related sexually deceptive orchids. *Evolution.* **65**, 2606–2620 (2011).
11. Sedeek, K. E. M., et al. Genic rather than genome-wide differences between sexually deceptive *Ophrys* orchids with different pollinators. *Mol. Ecol.* **23**, 6192–6205 (2014).
12. Breitkopf, H., et al. Multiple shifts to different pollinators fuelled rapid diversification in sexually deceptive *Ophrys* orchids. *New Phytol.* **207**, 377–389 (2015).
13. Piñeiro Fernández, L., et al. A phylogenomic analysis of the floral transcriptomes of sexually deceptive and rewarding European orchids, *Ophrys* and *Gymnadenia*. *Front. Plant Sci.* **10**, 1553 (2019).
14. Greilhuber, J. & Ehrendorfer, F. Chromosome numbers and evolution in *Ophrys* (Orchidaceae). *Plant Syst. Evol.* **124**, 125–138 (1975).
15. Bianco, P., et al. Karyological studies of some taxa of the genus *Ophrys* (Orchidaceae) from Apulia (Italy). *Caryologia.* **42**, 57–63 (1989).
16. Li, H. Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics.* **32**, 2103–2110 (2016).
17. Simão, F. A., et al. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* **31**, 3210–3212 (2015).
18. Wicker, T., et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
19. Hasing, T., et al. A phased *Vanilla planifolia* genome enables genetic improvement of flavour and production. *Nat. Food.* **1**, 811–819 (2020).
20. Zhang, G. Q., et al. The *Apostasia* genome and the evolution of orchids. *Nature.* **549**, 379–383 (2017).
21. Yuan, Y., et al. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat. Commun.* **9**, 1625 (2018).
22. Yang, F. X., et al. The genome of *Cymbidium sinense* revealed the evolution of orchid traits. *Plant Biotechnol. J.* **19**, 2501–2516 (2021).
23. Li, M. H., et al. Genomes of leafy and leafless *Platanthera* orchids illuminate the evolution of mycoheterotrophy. *Nat. Plants.* **8**, 373–388 (2022).
24. Kidwell, M. G. Transposable elements and the evolution of genome size in eukaryotes. *Genetica.* **115**, 49–63 (2002).

25. Chumová, Z., et al. Repeat proliferation and partial endoreplication jointly shape the patterns of genome size evolution in orchids. *Plant J.* **107**, 511–524 (2021).
26. Eriksson, M. C., et al. Repeat dynamics across timescales: a perspective from sibling allotetraploid marsh orchids (*Dactylorhiza majalis* s.l.). *Molec. Biol. Evol.* **39,** msac167 (2022).
27. Quézel, P. La flore du bassin méditerranéen: origine, mise en place, endémisme. *Ecol. Mediterr.* **21**, 19–39 (1995).
28. Thompson, J. D. *Plant Evolution in the Mediterranean. Insights for conservation. Second Edition* Ch. 1 (Oxford University Press, 2020).
29. Chuong, E. B., et al. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017).
30. Dubin, M. J., et al. Transposons: a blessing curse. *Curr. Opin. Plant Biol.* **42**, 23-29 (2018).
31. Meier, J. I., et al. Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. *Science* **381**, eade2833 (2023).
32. Puttick, M. N., et al. Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms. *Proc. R. Soc. B Biol. Sci.* **282**, 20152289 (2015).
33. Oliver, K. R., et al. Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biol. Evol.* **5**, 1886–1901 (2013).
34. Zhao, D., et al. What makes up plant genomes: The vanishing line between transposable elements and genes. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859**, 366–380 (2016).
35. Balachandran, P., et al. Transposable element-mediated rearrangements are prevalent in human genomes. *Nat. Commun.* **13**, (2022).
36. Lisch, D. How important are transposons for plant evolution? *Nat. Rev. Genet.* **14**, 49–61 (2013).
37. Bateman, R. M., et al. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). *Botanical Journal of the Linnean Society.* **142**, 1–40 (2003).
38. Wong, D. C. J. & Peakall, R. Orchid phylotranscriptomics: the prospects of repurposing multi-tissue transcriptomes for phylogenetic analysis and beyond. *Front. Plant Sci.* **13**, 910362 (2022).
39. Bell, C. D., et al. The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* **97**, 1296–1303 (2010).
40. Hertweck, K. L., et al. Phylogenetics, divergence times and diversification from three genomic partitions in monocots. *Bot. J. Linn. Soc.* **178**, 375–393 (2015).
41. Turco, A., et al. Contribution to the study of wild Orchidaceae, genus *Platanthera* L.C.M. Richard. Karyotype and C-banding analysis of two species from Italy. *Caryologia* **4**, 11–16 (2021).
42. Inda, L. A., et al. 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 (2012).
43. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

44. Bio, F., et al. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518 (2020).
45. Johnson, C. S., et al. *TRANSPARENT TESTA GLABRA2*, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell.* **14**, 1359–1375 (2002).
46. Curaba, J., et al. Identification of an SCPL gene controlling anthocyanin acylation in carrot (*Daucus carota* L.) root. *Front. Plant Sci.* **10**, 1770 (2020).
47. Corley, S. B., et al. Floral asymmetry involves an interplay between TCP and MYB transcription factors in *Antirrhinum*. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5068–5073 (2005).
48. Li, S. & Zachgo, S. TCP3 interacts with R2R3-MYB proteins, promotes flavonoid biosynthesis and negatively regulates the auxin response in *Arabidopsis thaliana*. *Plant J.* **76**, 901–913 (2013).
49. Chen, Y. Y., et al. Genome-wide identification of *YABBY* genes in Orchidaceae and their expression patterns in *Phalaenopsis* orchid. *Genes.* **11**, 1–17 (2020).
50. Sedeek, K. E. M., et al. Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. *PLoS One.* **8**, e64621 (2013)
51. Schiestl, F. P., et al. Orchid pollination by sexual swindle. *Nature.* **399**, 421–421 (1999).
52. Schiestl, F. P., et al. Sex pheromone mimicry in the early spider orchid (*Ophrys sphegodes*): patterns of hydrocarbons as the key mechanism for pollination by sexual deception. *J. Comp. Physiol. A.* **186**, 567–574 (2000).
53. Schlüter, P. M., et al. Stearoyl-acyl carrier protein desaturases are associated with floral isolation in sexually deceptive orchids. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5696–5701 (2011).
54. Xu, S., et al. The genetic basis of pollinator adaptation in a sexually deceptive orchid. *PLoS Genet.* **8**, e1002889 (2012).
55. Sedeek, K. E. M., et al. Amino acid change in an orchid desaturase enables mimicry of the pollinator's sex pheromone. *Curr. Biol.* **26**, 1505–1511 (2016).
56. Shanklin, J. & Cahoon, E. B. Desaturation and related modifications of fatty acids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 611–641 (1998).
57. Perera, M. A. D. N., et al. Biological origins of normal-chain hydrocarbons: a pathway model based on cuticular wax analyses of maize silks. *Plant J.* **64**, 618–632 (2010).
58. Wang, M., et al. Five fatty acyl-coenzyme a reductases are involved in the biosynthesis of primary alcohols in *Aegilops tauschii* leaves. *Front. Plant Sci.* **8**, 1012 (2017).
59. Ohno, S. Evolution by gene duplication. Springer (1970).
60. Magadum, S., et al. Gene duplication as a major force in evolution. *J. Genet.* **92**, 155–161 (2013).
61. Innan, H. & Kondrashov, F. The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* **11**, 97–108 (2010).
62. Francino, M. P. An adaptive radiation model for the origin of new gene functions. *Nat. Genet.* **37**, 573–578 (2005).

63. Xu, S., et al. Pollinator-driven speciation in sexually deceptive orchids. *Int. J. Ecol.* **2012**, 285081 (2012).
64. Paulus, H. & Gack, C. Pollinators as prepollinating isolation factors: evolution and speciation in *Ophrys* (Orchidaceae). *Isr. J. Bot.* **39**, 43–79 (1990).
65. Schiestl, F. P. & Schlüter, P. M. Floral isolation, specialized pollination, and pollinator behavior in orchids. *Annu. Rev. Entomol.* **54**, 425–446 (2009).
66. Cavalli-Sforza, L. L. & Edwards, A. W. F. Phylogenetic analysis: models and estimation procedures. *Evolution* **21**, 550-570 (1967).
67. Nosil, P. & Feder, J. L. Genomic divergence during speciation: causes and consequences. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 332–342 (2012).
68. Mant, J., et al. Cuticular hydrocarbons as sex pheromone of the bee *Colletes cunicularis* and the key to its mimicry by the sexually deceptive orchid, *Ophrys exaltata*. *J. Chem. Ecol.* **31**, 1765– 1787 (2005).
69. Russo, A., et al. Low-input high-molecular-weight DNA extraction for long-read sequencing from plants of diverse families. *Front. Plant Sci.* **13**, 883897 (2022).
70. Trávníček, P., et al. Challenges of flow-cytometric estimation of nuclear genome size in orchids, a plant group with both whole-genome and progressively partial endoreplication. *Cytom. Part A.* **87**, 958–966 (2015).
71. Temsch, E. M., et al. Heavy metal pollution, selection, and genome size: the species of the Žerjav study revisited with flow cytometry. *J. Bot.* **2010**, 596542 (2010).
72. Doležel, J., et al. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol. Plant.* **85**, 625–631 (1992).
73. Doležel, J., et al. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2**, 2233–2244 (2007).
74. Doležel, J., et al. Nuclear DNA content and genome size of trout and human. *Cytometry.* **51A**, 127–128 (2003).
75. Walker, B. J., et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
76. Nagarajan, N., et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
77. Pryszcz, L. P. & Gabaldón, T. Redundans: An assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**, e113 (2016).
78. Grob, S., Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the *flamenco* locus of *Drosophila*. *Mol. Cell* **55**, 678–693 (2014).
79. Liu, C. In situ Hi-C library preparation for plants to study their three-dimensional chromatin interactions on a genome-wide scale. *Methods Mol. Biol.* **1629**, 155–166 (2017).

80. Ghurye, J., et al. Scaffolding of long read assemblies using long range contact information. *BMC Genomics.* **18**, 527 (2017).
81. Ghurye, J., et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLOS Comput. Biol.* **15**, e1007273 (2019).
82. Durand, N. C., et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
83. Dudchenko, O., et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science.* **356**, 92–95 (2017).
84. Hoff, K. J., et al. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95 (2019).
85. Brůna, T., et al. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics Bioinformatics.* **2**, lqaa026 (2020).
86. Chung, O., et al. A chromosome-scale genome assembly and annotation of the spring orchid (*Cymbidium goeringii*). *Mol. Ecol. Resour.* **22,** 1168–1177 (2021).
87. Kawahara, Y., et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice.* **6**, 3–10 (2013).
88. Harkess, A., et al. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.* **8**, 1279 (2017).
89. Keilwagen, J., et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89–e89 (2016).
90. Bucchini, F., et al. TRAPID 2.0: a web application for taxonomic and functional analysis of de novo transcriptomes. *Nucleic Acids Res.* **49**, e101–e101 (2021).
91. Zheng, Y., et al. iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* **9**, 1667–1670 (2016).
92. Chan, P. P., et al. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).
93. Ellinghaus, D., et al. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
94. Katoh, K., et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
95. Ma, J. & Bennetzen, J. L. Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12404–12410 (2004).
96. Price, M. N., et al. FastTree 2-Approximately maximum-likelihood trees for large alignments. *PLoS One.* **5**, e9490 (2010).
97. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* **30**, 1312–1313 (2014).
98. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics.* **28**, 2689–2690 (2012).

99. Kim, Y.K., et al. Plastome evolution and phylogeny of Orchidaceae, with 24 new sequences. *Front. Plant Sci.* **11**, 22 (2020).
100. Bouckaert, R., et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).
101. Tan, M. H., et al. MitoPhAST, a new automated mitogenomic phylogeny tool in the post-genomic era with a case study of 89 decapod mitogenomes including eight new freshwater crayfish mitogenomes. *Mol. Phylogenet. Evol.* **85**, 180–188 (2015).
102. Edgar, R. C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* **5**, 113 (2004).
103. Capella-Gutiérrez, S., et al. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* **25**, 1972–1973 (2009).
104. Darriba, Di., et al. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2020).
105. Tang, H., et al. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **18**, 1944–1954 (2008).
106. Kiełbasa, S. M., et al. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487 (2011).
107. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arXiv:1303.3997 (2013).
108. Danecek, P., et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
109. Fischer, M. C., et al. Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Mol. Ecol.* **20**, 1450–1462 (2011).
110. Takezaki, N. & Nei, M. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics.* **144,** 389–399 (1996).

Figures

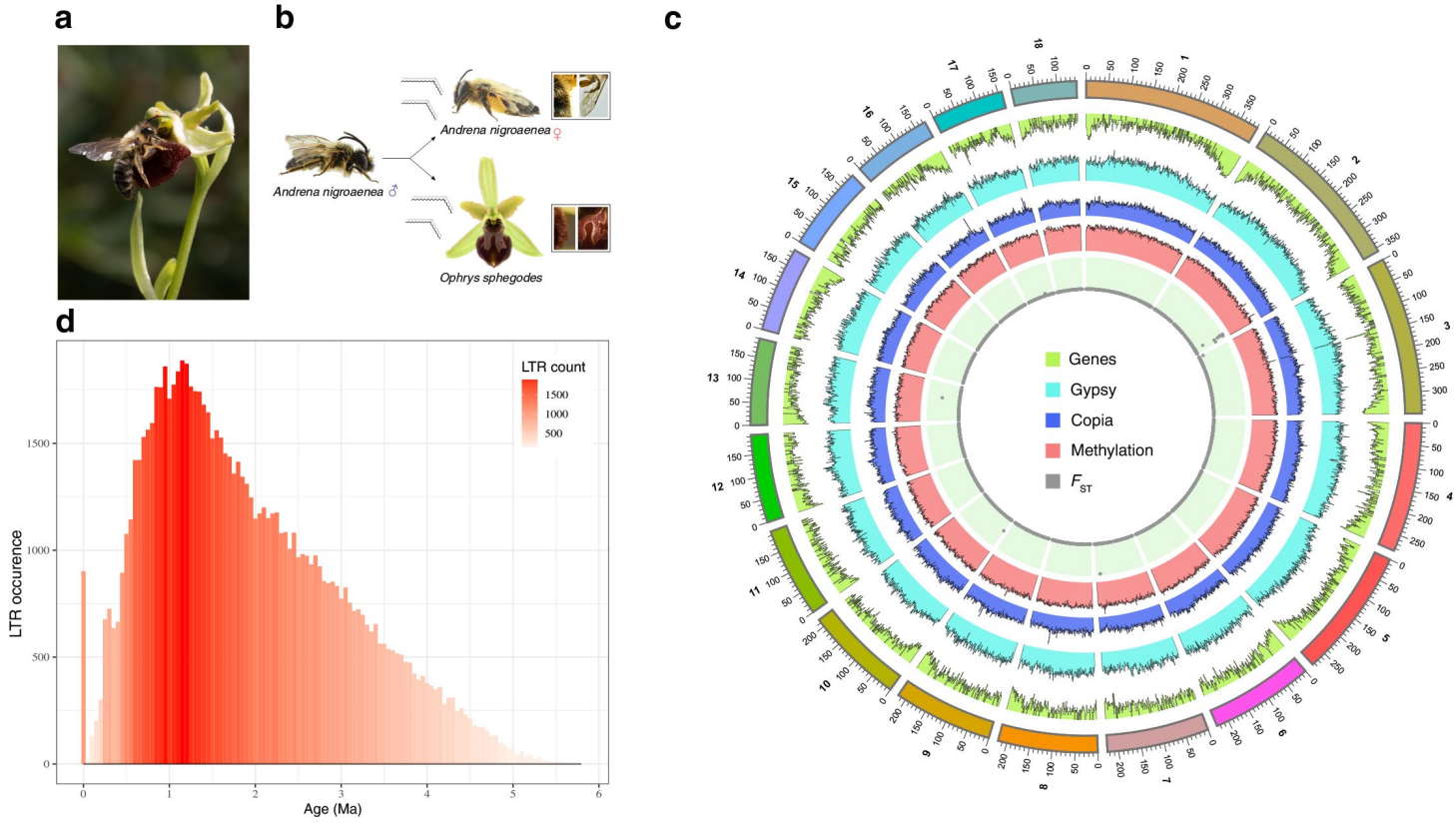

Fig. 1 |

Figure 1

Pollination and genome of *Ophrys sphegodes.*

a. Pseudocopulation of an *Andrena nigroaenea* male with an *Ophrys sphegodes* flower. Photo courtesy of Noa Schwabe. **b.** Schematic representation of sexual deception. An *Ophrys* orchid mimics the pollinator's female insect's hairs, wings and pheromones to lure the male into pseudocopulation. Photos courtesy of Noa Schwabe. **c.** The 18 assembled pseudochromosomes of *O. sphegodes* are labelled from 1 to 18. From outside to inside: chromosomes; gene density in light green; *Gypsy* retroelements in turquoise; *Copia* retroelements in blue; cytosine methylome in light red; between-species differentiation (global F_{ST}) along chromosomes in grey. **d.** Recent LTR insertions vs. their age, showing that LTR expansion reached its maximum at around 1.3 to 0.8 Ma.

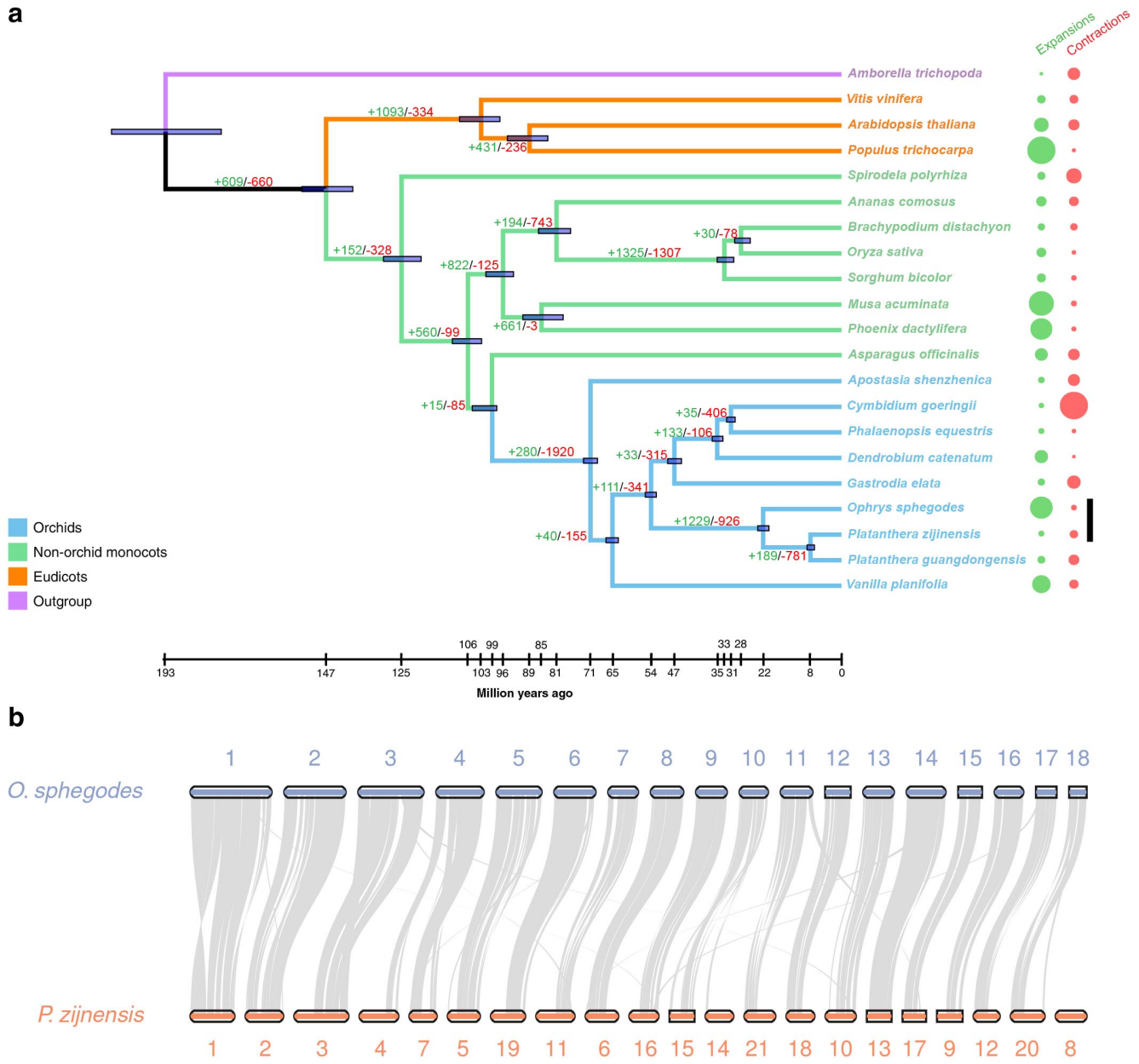

Fig. 2 |

Figure 2

Evolutionary relationships among orchid genomes.

a. Phylogenomic tree showing estimated divergence time and evolution of gene families for 21 plant species, colour indicating different plant groups (see inset). Divergence times are indicated by light blue bars at the internodes; their ranges indicate the 95% confidence intervals (CI) of the divergence time. Numbers at the branches indicate the number of expanded (green) and contracted (red) gene families in the lineages. Bubbles at the tips indicate expanded (green) and contracted (red) gene families per species (see Supplementary Fig. 8 and Supplementary Table 17 for CI intervals and number of expanded/contracted gene families). **b.**Genome comparison between *O. sphegodes* and *P. zijinensis.*

Chromosome comparison shows a high degree of collinearity and some chromosome rearrangements (see also Supplementary Fig. 9).

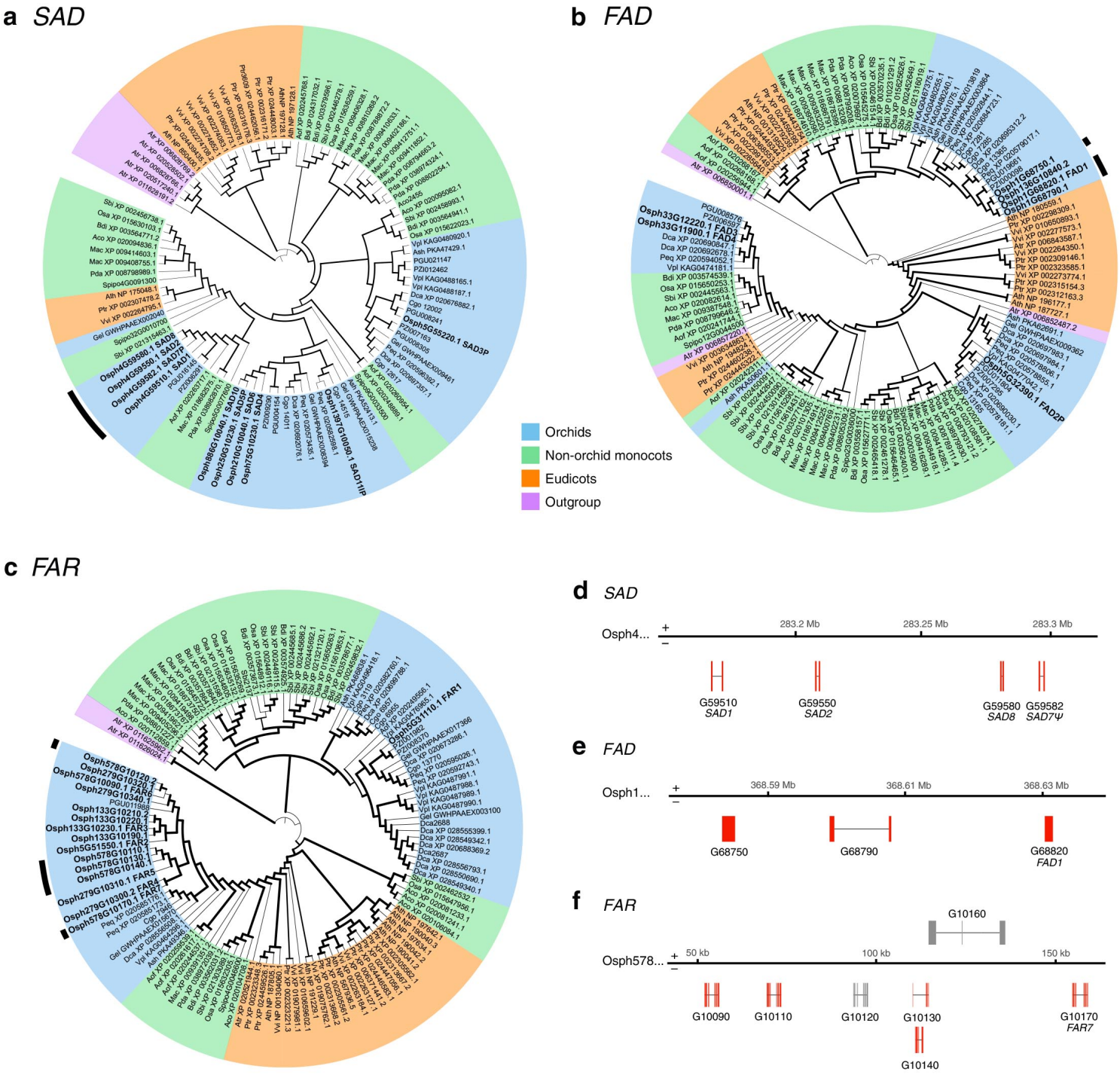

Fig. 3 |

Figure 3

Phylogenetic analyses of three gene families showing gene duplications.

a. *SAD* gene family tree. **b.** *FAD* gene family tree. **C.** *FAR* gene family tree. Colour shading for taxa as in Fig 2a. *O. sphegodes* genes are shown in bold, branch thickness indicating bootstrap support. The first

letters of each gene sequence indicate species (as in Supplementary Table 17), where Osph indicates *O. sphegodes*; Pgu and Pzi indicate *Platanthera guangdongensis* and *P. zijinensis*, respectively. Black bars highlight clustered genes shown in subsequent Figure panels. **d.** *SAD2* gene cluster on chr 4, containing *SAD1/2/8/7*$^{\psi}$ (gene details in Supplementary Table 7). **e.** Details of a cluster on chr 1 containing 3 *FAD* homologues. **f.** Details of scaffold 578 with 5 *FAR* gene homologues. Gene models of interest are drawn in red.

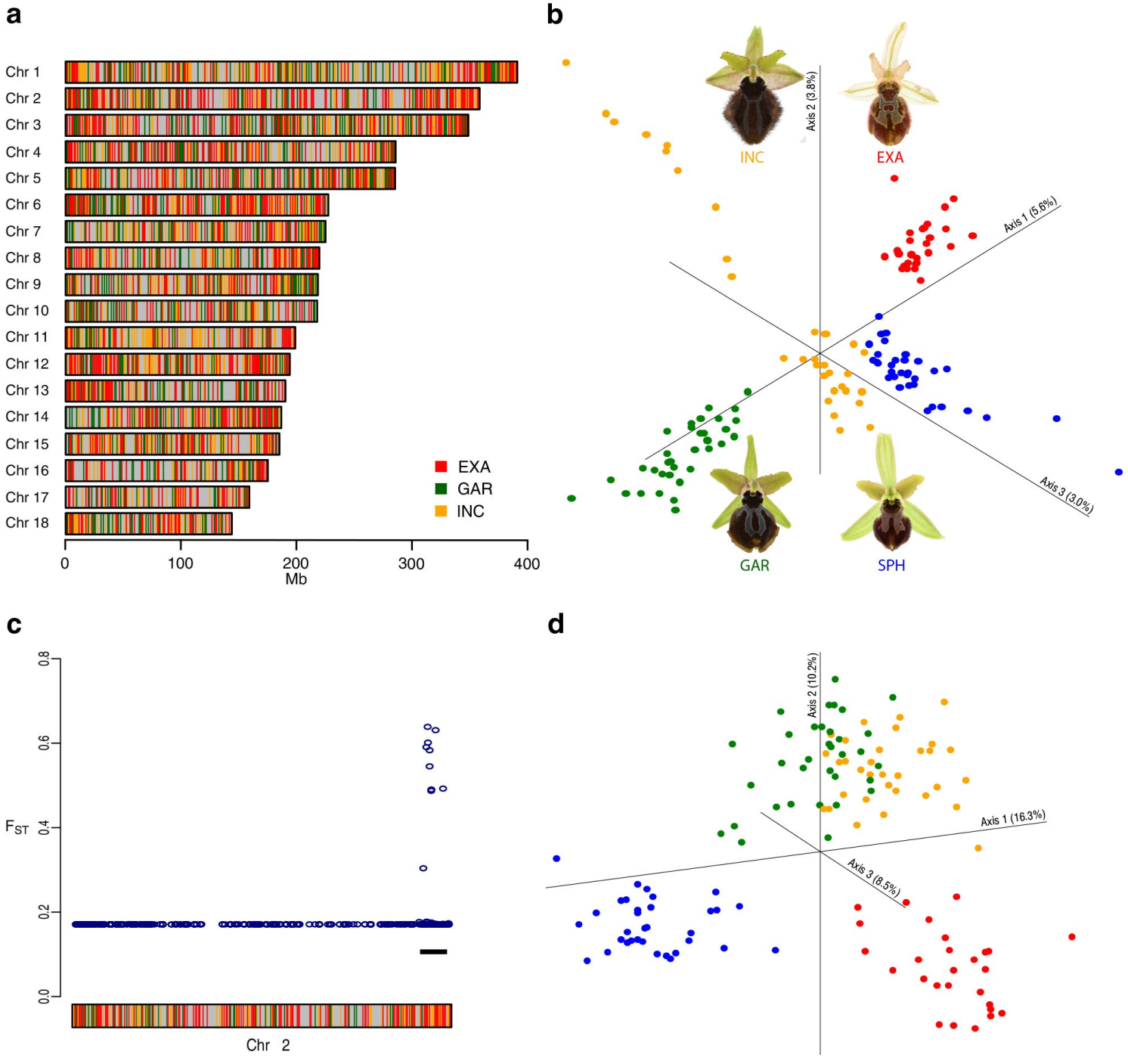

Fig. 4 |

Figure 4

GBS-based population genetic analysis between four sympatric *Ophrys* species.

a. Genetic distance among the different species of the *O. sphegodes s.l.* group plotted along the genome, colours showing the most genetically dissimilar species for a given 1 Mb window (see Supplementary Fig. 12 for genetic similarity). **b.** Genome-wide PCoA plot of pairwise distances between plant individuals revealed genetic separation between the four species. **c.** Global F_{ST} among the 4 species was elevated in one region on chr 2 (327-346 Mb, shown as black bar). **d.** PCoA of this region of chr 2 shows increased separation between *O. sphegodes* and *O. exaltata.* Colours show the most dissimilar species in panels a and c (grey: no data), and species identity in panels b and d, where *O. sphegodes* (SPH) is shown in blue; *O. exaltata* (EXA), *O. garganica* (GAR) and *O. incubacea* (INC) are shown in red, green and yellow, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportingInformationOphrysGenome.pdf