

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

Russo *et al.*

Supplementary Information

Table of Contents

<i>The genome of the early spider-orchid <i>Ophrys sphegodes</i> provides insights into sexual deception and adaptation to pollinators</i>	1
Table of Contents	2
Extended Methods	4
1 Genome assembly	4
1.1 Genome assembly and long read polishing	4
1.2 Illumina sequencing and short read polishing	4
1.3 Heterozygosity in the genome and gap closing	5
1.4 Hi-C library preparation and genome assembly scaffolding	6
1.5 Quality assessment and validation	7
2 Genome Annotation	7
2.1 TE Annotation	7
2.2 Annotation of protein coding genes	8
2.3 Functional annotation of protein coding genes	10
2.4 Annotation of transcription factors and protein kinase genes	10
2.5 Annotation of tRNAs and rRNAs	11
3 Chromosome-wide DNA Nanopore methylation map	11
4 Population genetics analyses	12
4.1 GBS and RNA-Seq variant calling and data filtering strategy	12
4.2 Distance-based analysis	13
4.3 F_{ST} and population distance analysis	13
5 Figures	14
Supplementary Figure 1. Flow cytometry measurements of genome size	14
Supplementary Figure 2. Heterozygosity estimation with k-mer analysis	15
Supplementary Figure 3. Summary of sequencing statistics for the Nanopore data	16
Supplementary Figure 4. Genome assembly strategy	17
Supplementary Figure 5. Assembled chromosomes	18
Supplementary Figure 6. Comparison of BUSCO single-copy orthologous genes across orchids	19
Supplementary Figure 7. Methylation patterns across all chromosomes	20
Supplementary Figure 8. Phylogenomic tree of 21 plant species showing 95% confidence intervals	22
Supplementary Figure 9. Synteny analysis	23
Supplementary Figure 10. Phylogenetic analyses of two transcription factor gene families showing significant gene expansion	24

36	Supplementary Figure 11. Positions of candidate genes	25
37	Supplementary Figure 12. Most similar species to <i>O. sphegodes</i> per 1 Mb window.....	26
38	Supplementary Figure 13. GBS PCoA plots for each chromosome.	27
39	Supplementary Tables	28
40	Supplementary Table 1. Flow cytometry measurements.	28
41	Supplementary Table 2. Statistics of sequencing raw data from Oxford Nanopore platforms. ..	29
42	Supplementary Table 3. Statistics of sequencing raw data from Illumina platforms.	30
43	Supplementary Table 4. Summary of the <i>Ophrys sphegodes</i> genome assembly	31
44	Supplementary Table 5. Chromosome lengths in the <i>Ophrys sphegodes</i> genome assembly.....	32
45	Supplementary Table 6. Mapping rate of Illumina reads against the genome.....	33
46	Supplementary Table 7. Candidate genes for odour production included in the gene annotation	
47		34
48	Supplementary Table 8. Additional candidate genes identified from annotation data.	37
49	Supplementary Table 9. List of transcription factors identified	39
50	Supplementary Table 10. List of protein kinases identified	50
51	Supplementary Table 11. Repetitive content in the <i>Ophrys sphegodes</i> genome	55
52	Supplementary Table 12. Repetitive elements in orchids	56
53	Supplementary Table 13. Statistics of sequence variants.....	57
54	Supplementary Table 14. Distance matrix for comparing genotype calls.....	58
55	Supplementary Table 15. Point mutation transition matrices	59
56	Supplementary Table 16. Statistics of syntenic analysis of <i>O. sphegodes</i> (21 462 genes)	
57	compared with other orchid genomes	60
58	Supplementary Table 17. Gene family expansion and contraction across 21 plant species with	
59	sequenced genome.....	61
60	Supplementary Table 18. Accession number of orchid genome assemblies used in this study ..	62
61	6 References	63
62		
63		

Extended Methods

1 Genome assembly

1.1 Genome assembly and long read polishing

Nanopore long reads were base-called with Guppy v3.0.4¹, and their quality assessed with NanoPlot v1.0.0² (Supplementary Fig. 3). All reads were filtered using Filtlong v0.2.0 (<https://github.com/rrwick/Filtlong>) with the options `--min_length 10000 --target_bases 125000000000` to discard all reads shorter than 10 000 bp and remove low quality reads until 125 Gb remained. To generate the genome assembly, filtered reads were used as input to run Minimap2 v2.17 and Miniasm v0.3³ (with options `-R -c 2`). Initial contigs were polished with long-read data, first by mapping the long reads against the assembly with Minimap2, then using the mapped reads and the draft assembly as input for Racon v1.4.3⁴ (parameters `--match 8 -mismatch -6 --gap -8 --window-length 500`). This polishing process was iterated four times. Then, a consensus sequence was created with the medaka_consensus module (medaka v0.11.4; <https://github.com/nanoporetech/medaka>) (parameters `--model r941_prom_high_g303`).

1.2 Illumina sequencing and short read polishing

Genomic DNA extracted for Nanopore sequencing (individual SPH_8 from the Gargano region, South Italy) was sheared using a Covaris sonicator, then used to prepare two size-selected DNA libraries with the TruSeq DNA Nano kit (Illumina). Libraries were sequenced on an Illumina NovaSeq 6000 platform (2× 150 nt and 2× 250 nt; Supplementary Table 3). Four additional libraries were prepared from the individual SPH_511B (also from the Gargano region), with the TruSeq DNA Nano Kit (Illumina), and then sequenced on Illumina HiSeq 2500/4000 (2× 150 nt; Supplementary Table 3). The NovaSeq sequence reads were trimmed using Cutadapt v2.8⁵, with the following parameters for adaptor removal: `--minimum-length 1 --trim-n -a GATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT`. The HiSeq sequence reads were trimmed to remove poor-quality reads, using: `--quality-base 20 --minimum-length 1` (only one dataset needed trimming at both ends, specified with `--quality-base 28,15`). Illumina short reads were mapped against the assembly using BWA-MEM v2.1⁶, and the resulting bam files and the assembly (after the medaka step) were used as input for Pilon v1.23⁷ to further polish the assembly with short reads. This process was repeated

three times in total: the first and second round were performed with the NovaSeq dataset (individual SPH_8) with parameters `--fix all`. A third round was performed including the HiSeq dataset (individual SPH_511B) and parameter `--fix indels`, to fix small indels with short reads from another individual. A schematic representation of the assembly strategy is provided in Supplementary Fig. 4.

1.3 Heterozygosity in the genome and gap closing

Prior to genome assembly, the overall heterozygosity rate of the *O. sphegodes* genome was estimated based on the k -mer count distribution with GenomeScope (<http://qb.cshl.edu/genomescope/>) and Illumina reads. The model calculates the heterozygosity rate based on duplicated k -mers from maternal and paternal origin and the relative heights of the homozygous and heterozygous peaks, and revealed a heterozygosity rate of 1.28% (meaning that the heterozygous peak matches the height of the homozygous peak at around 1.28 % for $k = 21$) (Supplementary Fig. 2). The assembly of heterozygous genomes is a challenging task because it is difficult to accurately differentiate between heterozygous haplotypes^{8–11}. When two heterozygous sequences of the same region exceed a certain threshold of nucleotide diversity, the genome assembler will consider them as separate contigs, i.e., sequences belonging to different regions of the genome. Instead of being collapsed as single-haplotype contig, or discarded, these sequences are included as extra contigs in the genome assembly, thus inflating assembly size^{12,13}. Furthermore, these extra contigs make the assembly more fragmented, as gaps between sequences cannot be solved when two possible alternative solutions are present¹⁴. In the case of the *Ophrys* genome, after assembling and polishing the contigs, the assembly size accounted for 6.4 Gb, more than 1 Gb larger than the estimated genome size. To remove under-collapsed heterozygous contigs, therefore, the Redundans v0.11 pipeline¹⁵ was used (Supplementary Fig. 4). The pipeline considers the similarity between two sequences and read coverage depth to detect and remove redundant contigs (used with parameters `-b -m 400 --identity 0.51 --overlap 0.80`). Then, short and long reads are used by Redundans for gap closing (2 iterations). Final assembly size (Ospv1.0.fa) at this stage was 5.2 Mb, with a total of 8 074 contigs (Supplementary Table 4).

1.4 Hi-C library preparation and genome assembly scaffolding

To scaffold the genome assembly to chromosomes, an *in situ* Hi-C library was prepared as described previously¹⁶ with modifications¹⁷. Individual SPH_2 from Capotiale, Italy, was chosen for the experiment. Briefly, 0.5 g of leaf tissue was treated with nuclei isolation buffer and 36% formaldehyde in a vacuum pump, to extract cell nuclei and cross-link chromatin. Plant material was then crushed with mortar and pestle in liquid nitrogen, centrifuged to extract the supernatant (nuclei), digested with the enzyme *HindIII*-HF (New England Biolabs, Ipswich, MA, USA, R3104L), and re-ligated for cross-link reversal. DNA was then extracted by phenol:chloroform:isoamyl alcohol (25:24:1) and cleaned up. Four Hi-C libraries were prepared with the KAPA HyperPrep Kit (Roche, Switzerland, KK0502), and sequenced on an Illumina NovaSeq 6000 SP FlowCell (2× 150 nt) to produce a total of 1 billion reads (Supplementary Table 3). To assemble the chromosomes, the *de novo* assembly and Hi-C reads were used as input for a combined approach including different pipelines. First, raw Hi-C data was mapped against the genome assembly with the ArimaGenomics mapping script, specifically developed for Hi-C short reads (https://github.com/ArimaGenomics/mapping_pipeline/blob/master/01_mapping_arima.sh). Then, the bam file was used as input for building scaffolds with Salsa v2.3^{14,18} (parameters: --enzyme AAGCTT --clean yes --cutoff 1000). After Salsa, 2 520 scaffolds remained (Osph-v1.2.fa). Finally, scaffolds were linked together to chromosomes using Juicebox Assembly Tools (JBAT)¹⁹ and the 3D-DNA pipeline v-180114²⁰. First, raw Hi-C short reads were mapped against the assembly version Osph-v1.2.fa with Juicer v-1.5.6 (parameters: -s *HindIII* -C 40000000 -D early). Then, final chromosomes were assembled by the 3D-DNA pipeline (parameters: --input 20000 --editor-coarse-resolution 500000 --editor-coarse-region 2000000 --editor-fine-resolution 25000 --editor-coarse-stringency 40 --editor-repeat-coverage 5), and manual curation of the chromosomes was performed with the visualisation tool embedded in JBAT. As a last step, HiC-Hiker²¹ v1.0.0 was used to further reduce scaffold misorientation at small scale, with option -K 90000, to produce the final *O. sphegodes* genome assembly (Osph-v2.3.fa, Supplementary Table 4).

1.5 Quality assessment and validation

The completeness and quality of the final *O. sphegodes* assembly was assessed with two methods. First, 267 billion Illumina paired-end reads (Supplementary Table 3) were mapped to the genome using BWA-MEM2 v2.1⁶ with default parameters. In total, 97.8% of the reads could be mapped to the genome overall and 85.9% were properly paired (Supplementary Table 6). Furthermore, BUSCO v5.1.3^{22,23} was run in genome mode using the embryophyta_odb10 database (release 2020-09-10) based on 1 614 universal single-copy orthologues. In total, 1 369 (84.9%) ultra-conserved genes were found in full length (complete) and 100 (6.2%) were found to be partial (fragmented) in the final assembly. The overall completeness of the final *O. sphegodes* assembly estimated based on BUSCO is similar to the *V. planifolia* haplotype A genome, and only slightly less than other orchid genomes (except for *G. elata*) (Supplementary Fig. 6a). A comparison of the genome's BUSCO results with the ones of the *O. sphegodes* transcriptome²⁴ shows consistent 'complete' and 'missing' values (BUSCO on *O. sphegodes* transcriptome run with embryophyta_odb10 results: Complete, 88.3%; Fragmented, 4.4%; Missing, 7.3%). Taken together, these analyses suggest that the overall quality of the assembly is high.

2 Genome Annotation

2.1 TE Annotation

Transposable elements (TE) were manually characterised according to the method described in ref.²⁵. First, putative TEs were identified with RepeatModeler v2.0.1 and RepeatMasker v4.1.0²⁶, and then blasted against chromosome 1 of the *O. sphegodes* genome. Resulting sequence hits were retrieved from the genome fasta file with their flanking regions (6 000 nt to 10 000 nt upstream and downstream, depending on the sequence class), and used as input for a multiple sequence alignment (MSA) with CLUSTAL W²⁷. Resulting alignments were visually inspected to identify start and end motifs and to characterise the consensus sequence. All consensus sequences were renamed according to the three-letter code described in ref.²⁵. This procedure was employed for all putative repeats originally pre-assigned by RepeatMasker as part of a superfamily. Unfortunately, ~30% repetitive elements were classified as unknown by RepeatMasker. For those, the putative consensus sequences obtained from the MSA were blasted against the TRansposable Elements Platform (TREP) database (v.

20Aug2019, available at <https://trep-db.uzh.ch/>), and PTREP (Protein TRansposable Elements Platform, release 19) and Dfam (<https://dfam.org>), and then revised accordingly. If the unknown sequence did not have any match in any of the databases, we manually examined the MSA for presence of structural motifs that could help to characterise and classify the TE. By following this approach, we were able to characterise a total of 436 novel TEs in *O. sphegodes*, and created the first curated TE database of the Orchidaceae family. This data set has now been included in the TREP database (available at <https://trep-db.uzh.ch/>). Finally, we used this species-specific library to annotate the TEs in the genome with RepeatMasker v4.1.0 (`-engine rmbblast -nolow -norna -no_is`).

2.2 Annotation of protein coding genes

The soft-masked *O. sphegodes* genome was annotated using a combined approach that included BRAKER2 v2.1.6²⁸ with AUGUSTUS v3.4.0²⁹, GeneMark-EX v4.64_lic³⁰ with ProtHint v2.6.0, and DIAMOND v0.9.24³¹, for *de novo* gene prediction; GeMoMa v1.8³² for homology-based prediction; PASA v2.5.1³³ to model gene structures based on transcriptome evidence, annotate UTRs and deal with alternative splicing variants. First, protein hints were generated by ProtHint 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 monocots lineage. This extrinsic evidence was given to GeneMark for self-training. AUGUSTUS was specifically trained for *O. sphegodes* using RNA-seq data from floral tissues of *O. sphegodes* and the three closely related species: *O. exaltata*, *O. incubacea*, *O. garganica* (previously published²⁴, available under SRA accession PRJNA574279). First, all RNA-seq fastq were mapped against the unmasked genome assembly with STAR v2.7.5c³⁴ (parameters: `--readQualityScoreBase 28 -outFilterScoreMinOverLread 0 -outFilterMismatchNmax 2`). Then, the bam files were used as input for the BRAKER2 pipeline to train AUGUSTUS. A second BRAKER2 run was performed with `-etpmode` and hints from the first run, to improve the prediction accuracy of AUGUSTUS and include predictions from GeneMark. The resulting gff file was used as an input for GeMoMa, for gene prediction based on homology with the genome annotations of *Phalaenopsis equestris*³⁵, *Cymbidium goeringii*³⁶, *Oryza sativa* ssp. *japonica* cultivar Nipponbare IRGSP-1.0³⁷ and *Asparagus officinalis* Aspof.V1³⁸. PASA was run on the final prediction to model gene structures based on the transcriptome assembly of *O.*

*sphegodes s.l.*²⁴. First, the transcriptome was mapped against the genome assembly with Minimap2, and alignments were clustered and assembled. Then, the PASA assemblies were incorporated into our final gene predictions to correct exon boundaries, add UTRs, and update gene structures. At this stage, the total number of gene models was 155 025.

The predicted gene models were filtered based on their repeat content (after TE annotation). All genes with a repeat content $\geq 40\%$ of their length were removed. Gene models were also removed if the exact same gene structure was predicted multiple times independently by different algorithms. Moreover, when different isoforms of the same gene were predicted as different genes, they were clustered together into the same gene structure. At this stage, the total number of coding sequences (CDS) was 136 555. An extra filtering step was carried out based on RNA-seq support, i.e., genes with little RNA-seq evidence or no functional annotation were removed from downstream analysis (keeping genes with RNA-seq TPM ≥ 5). All gene models were integrated in AHRD (Automated assignment of Human Readable Descriptions), that classifies predicted sequences based on bit score of the blast results against SwissProt, TAIR10, trEMBL, overlap score, and token score of assigned Human Readable Description, where tokens are sequences of characters with a human readable meaning. We used AHRD with default parameters and considered valid only those genes that met the following three criteria: bit score > 50 with e-value $< e-10$; blast overlap $> 60\%$; token score > 0.5 . Genes with premature stop codons were considered to be pseudogenes and were removed from the final list. The final number of CDS after this filtering step was 42 542.

Candidate genes for hydrocarbon biosynthesis³⁹⁻⁴² and other traits of interest were also included in our annotation. This was done semi-manually for each candidate gene and then manually curated. Coding sequences of a priori candidate genes were used in BLASTN searches against the genome assembly (both reference and alternative heterozygous haplotigs, i.e., contigs discarded from the final assembly via Redundans) using an e-value of $1E-15$. Hits were filtered by percentage of sequence identity (typically between 90 and 95%) using varying threshold values reflecting phylogenetic distance between *Ophrys* and the organism of sequence origin. Valid hits were visualised in R to discern consecutive areas of homology along an *Ophrys* contig, putatively corresponding to different exons. When a homologue was found,

the corresponding contig sequence was extracted from the lowest to the highest BLASTN hit coordinates plus flanking regions (typically of 5kb), and the contig aligned to the query CDS using est2genome from the EMBOSS 6.6.0.0 package (emboss.sourceforge.net)⁴³. Resulting alignments were manually inspected in BioEdit 7.2.6⁴⁴, especially putative exon/intron boundaries, to obtain putative gene models maximally consistent with those previously described. All manually curated genes were included in the final gene prediction. In case a manually annotated gene overlapped with an automatically predicted one, we chose to keep the manual version. Supplementary Table 7 shows the complete list of manually annotated candidate genes.

2.3 Functional annotation of protein coding genes

The function, gene ontology (GO), and protein domain of filtered predicted coding genes were annotated with TRAPID⁴⁵. Protein coding sequences were first compared to the PLAZA 4.5 monocots database⁴⁶ for similarity search, using Orchidaceae as phylogenetic clade and DIAMOND to blast query sequences against the database with e-value cut-off of 1E-5. After initial processing, the InterPro database v.87.0 was used for Protein Domain search. In total, 38 978 transcripts were assigned to at least one GO term, 38 845 had at least one Protein Domain. The analysis also grouped predicted transcripts in 7 154 gene families. The top three gene families were HOM04x5M000001, with 774 transcripts, involved in protein binding; HOM04x5M000002, with 473 transcripts, with putative protein kinase activity; HOM04x5M000003, with 251 transcripts, with putative oxidoreductase activity. KEGG annotation was performed with KAAS (KEGG Automatic Annotation Server; <https://www.genome.jp/tools/kaas/>), and enzyme commission (EC) numbers were assigned to the predicted genes. Overall, 7 978 genes were assigned to KEGG orthologues (KO). Genes annotated in this way were again screened for genes of *a priori* interest, specifically for additional genes putatively involved in hydrocarbon or very-long-chain fatty acid (VLCFA) biosynthesis, anthocyanin biosynthesis and carotenoid biosynthesis (Supplementary Table 8).

2.4 Annotation of transcription factors and protein kinase genes

Transcription Factors (TF) and Protein Kinase (PK) genes were identified using the iTAK Classifier (<http://itak.feilab.net/cgi-bin/itak/index.cgi>)⁴⁷. A total of 2 078 coding genes were

identified as transcription factors or transcriptional regulators, and 917 of genes were identified as protein kinase (Supplementary Tables 9 and 10).

2.5 Annotation of tRNAs and rRNAs

The structure and number of transfer RNAs were predicted using tRNAscan-SE v2.0.9⁴⁸ and Infernal 1.1.4⁴⁹. A total of 2 239 candidate tRNA genes were found, of which 1 608 were confirmed by Infernal. Of these, 1 238 tRNA decode codons of the standard 20 amino acids, 9 are possible suppressor tRNAs, and 15 have undetermined isotypes; 346 were likely pseudogenes. Barrmap v0.9 (<https://github.com/tseemann/barrnap>) was used to annotate ribosomal RNA with default parameters and `--kingdom euk.` A total of 1 255 rRNA genes were found.

3 Chromosome-wide DNA Nanopore methylation map

Nanopore sequencing can detect methylated cytosine (mC) in plants, as they provide a characteristic raw electric signal compared to unmethylated cytosine⁵⁰. To measure 5-mC, sequencing data in FAST5 were basecalled with Guppy v5.0.11¹ with option `-c dna_r9.4.1_450bps_modbases_5mc_hac_prom.cfg -device cuda:0`. Guppy uses a deep learning recurrent neural network (RNN) model on the nanopore current signal to distinguish 5-mC from unmethylated cytosine (https://community.nanoporetech.com/technical_documents/). The methylation caller generates a log-likelihood value for the probability of methylated to unmethylated cytosines. The newly called fastq files were merged into a single one and indexed, to relate the Nanopore electric signal data to read IDs (`nanopolish index -s sequencing_summary.txt`). The fastq reads were mapped to the reference genome with Minimap2, with option `-ax map-ont --split-prefix string`. Then, Nanopolish v0.13.3 (<https://github.com/jts/nanopolish>) was used to detect methylated bases with option `call-methylation -r file.fastq -b mapping.bam -g genome.fa`. The result was a .tsv table with position of the CG dinucleotide (start/end), read ID, sequence, log-likelihood ratio of methylated/unmethylated base, where a positive value indicated support for methylation. Finally, methylation frequency was calculated with the `calculate_methylation_frequency.py` script provided in the nanopolish toolkit, using a log-likelihood ratio greater than 1.5

(methylated) as threshold for calling methylation⁵¹. Supplementary Fig. 7 shows methylation count per chromosome.

4 Population genetics analyses

4.1 GBS and RNA-Seq variant calling and data filtering strategy

Different data types were used for population genomic analysis, including (i) a genotyping-by-sequencing (GBS) dataset⁵², and (ii) an RNA-Seq data set²⁴. The samples were previously collected across five collection spots in the Gargano area, south Italy, and included the following closely related species: *O. sphegodes* Miller, *O. exaltata* subsp. *archipelagi* (Gözl & H.R. Reinhard) Del Prete, *O. garganica* Nelson ex O. & E. Danesch, *O. incubacea* Bianca⁵³. They are pollinated by *Andrena nigroaenea* (Kirby 1802), *Colletes cunicularius* (Linnaeus 1761), *Andrena pilipes* Fabricius 1781 (syn. *A. carbonaria*), *Andrena morio* Brullé 1832, respectively^{40,54,55}. The different data sets had varying numbers of biological individuals vs sequencing depth (and thus genotyped variants) per individual (RNA-Seq variant calling and data filtering strategy; see Supplementary Table 13). Data filtering of raw variant data in VCF format was carried out using SPA v0.1 (<https://peb.uni-hohenheim.de/SPA>), and putatively mis-assigned individuals were removed; the accession INC_527I was excluded from the GBS dataset, because its species assignment was unclear. Likewise, any samples with unclear population assignment and bud samples (pooled from different individuals)²⁴ were removed from the RNA-Seq dataset. The choice of variant filtering was designed to balance needs for genotyping quality and number of diploid, biallelic markers available for analysis, making sure each individual genotype call had a >99% chance of detecting both alleles given the number of reads (n , given in the sample depth of coverage field, DP, in a VCF file) and the average read quality (q), assuming (an unrealistically low) mean read quality of *phred* score $Q=10$ (yielding the probability q that a base read is correct of 0.9). The worst-case scenario for assessing if both alleles are detected within one individual occurs in homozygous state, where the probability that both alleles are correct is given by $P=1-p^{nq}-(1-p)^{nq}$, with $p=0.5$, the frequency of each allele at a heterozygous site. Additionally, both more and less stringent settings were assessed for consistency of results in exploratory analyses. For distance-based analyses, polymorphic, biallelic variant calls passing a minimal sample DP threshold of 3 when heterozygous, or of 10 when homozygous, were used for analysis. Where seemingly

homozygous variants exceeded a DP threshold of 3 (but <10), only the first allele at this variant was considered for distance analysis. For BayeScan analysis of GBS data, in addition, at least 6 allele calls had to be observed in *each* species (two for pairwise, four for global F_{ST} analysis). For BayeScan analysis of RNA-Seq with high read depth, only biallelic sites present in *all* individuals at per-sample DP ≥ 5 were analysed. SNP state transition matrices are given in Supplementary Table 15.

4.2 Distance-based analysis

Between-individual genetic distances were calculated from data sets filtered as outlined above. Here, following ref.⁵², we used the unphased diploid SNP distance described in ref.⁵⁶, allowing the inclusion of partial genotypes (in which only the first allele is evaluated, and partial combinations are evaluated as 0.5; Supplementary Table 14). Distance matrices were analysed via principal coordinate analysis (PCoA)⁵⁷; this analysis was implemented in SPA v0.1, which includes source code from the biOP library (<https://sourceforge.net/projects/biop/>; see ref.⁵²) for distance calculation and from FAMd v1.33⁵⁸ for PCoA. Distance analyses were carried out on biallelic variants passing the filtering steps as outlined above.

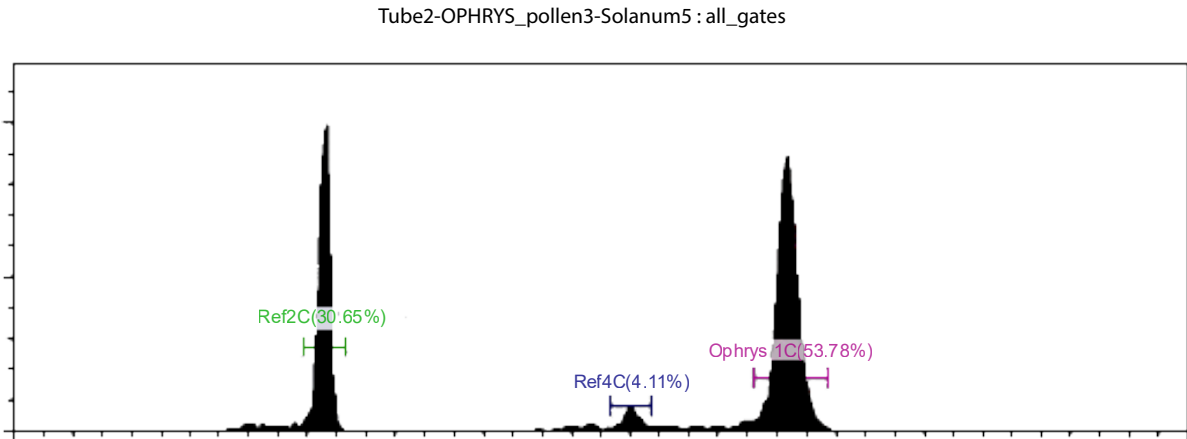
4.3 F_{ST} and population distance analysis

In order to calculate F_{ST} values for population differentiation and perform F_{ST} outlier scans, datasets were filtered and converted from VCF into BayeScan format using SPA v0.1. BayeScan v2.1 was run as described in ref.⁵², including their calculation of command-line PO settings from the number of available loci. BayeScan analysis was carried out both as a ‘global F_{ST} ’ analysis, treating each of the four species as a subpopulation, and as pairwise analyses between each possible combination of pairs of species. These analyses were carried out independently for GBS and RNA-Seq datasets. Additionally, to investigate genomic similarity among *O. sphegodes* and the other closely related species, for each chromosomal block, a genetic inter-population distance was calculated and averaged across multiple loci within a 1 Mb window to identify the species with the highest and lowest genetic distance. This was implemented using the single-locus chord distance for each variant⁵⁹, averaged into a multi-locus distance as described by ref.⁶⁰.

5 Figures

Supplementary Figure 1. Flow cytometry measurements of genome size

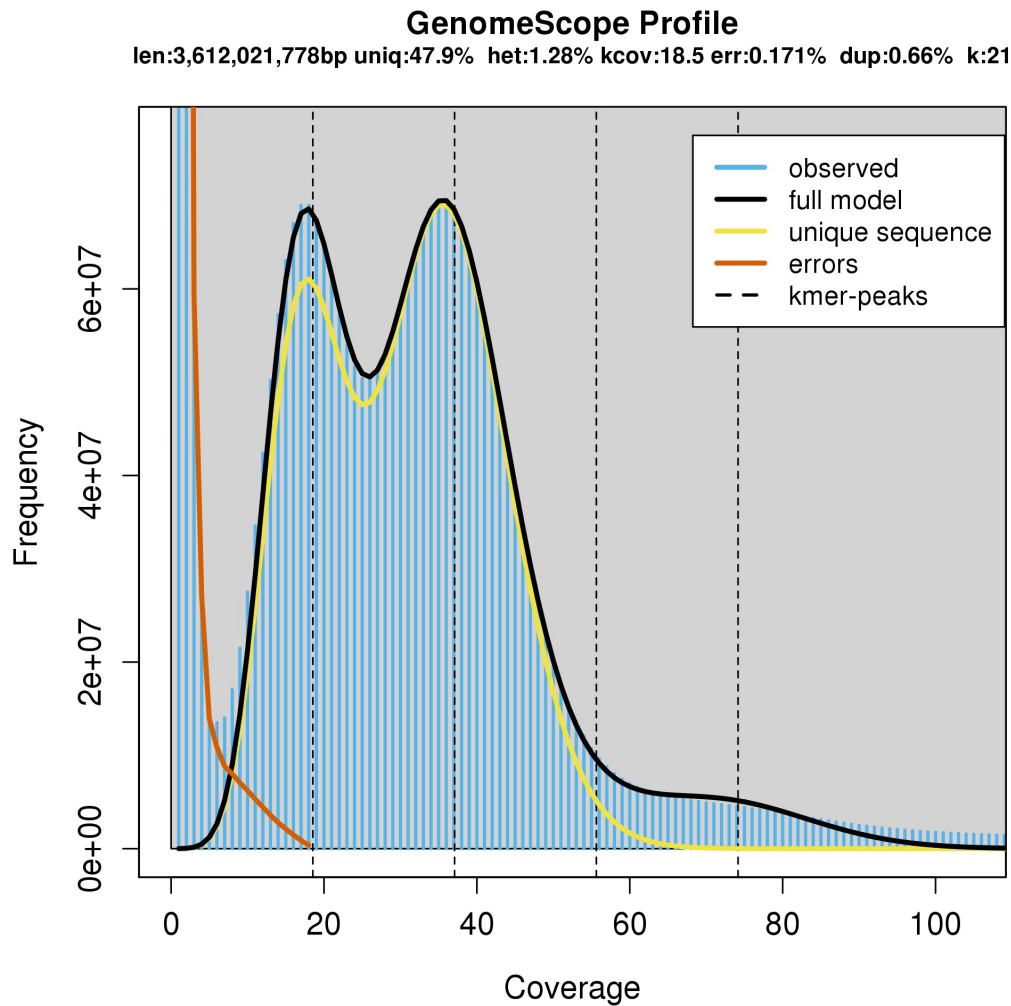
Fluorescence (PerCP-A) intensity of *O. sphegodes* nuclei from pollinia relative to the reference (*S. lycopersicum* leaf nuclei). The y axis shows the number of events (measured nuclei); the x axis shows relative fluorescence or light scatter intensity.



Population	Events	Total (%)	Parent (%)	Mean PerCP-A	Median PerCP-A	CV PerCP-A (%)	CV PerCP-H (%)	CV ECD-A (%)
All Events	18311	100	100	7395071.5	4422000	94.7	38.61	130.75
Ref2C	1059	5.78	30.65	4207840.0	4216375	1.72	2.64	2.38
Ref4V	142	0.78	4.11	8397176.0	8394184	1.45	2.17	1.91
Ophrys 1C	1858	10.15	53.78	10506736.0	10500936	1.43	0.16	1.38
all_gates	3455	18.87	18.87	8092343.0	10299056	36.48	18.6	37.92

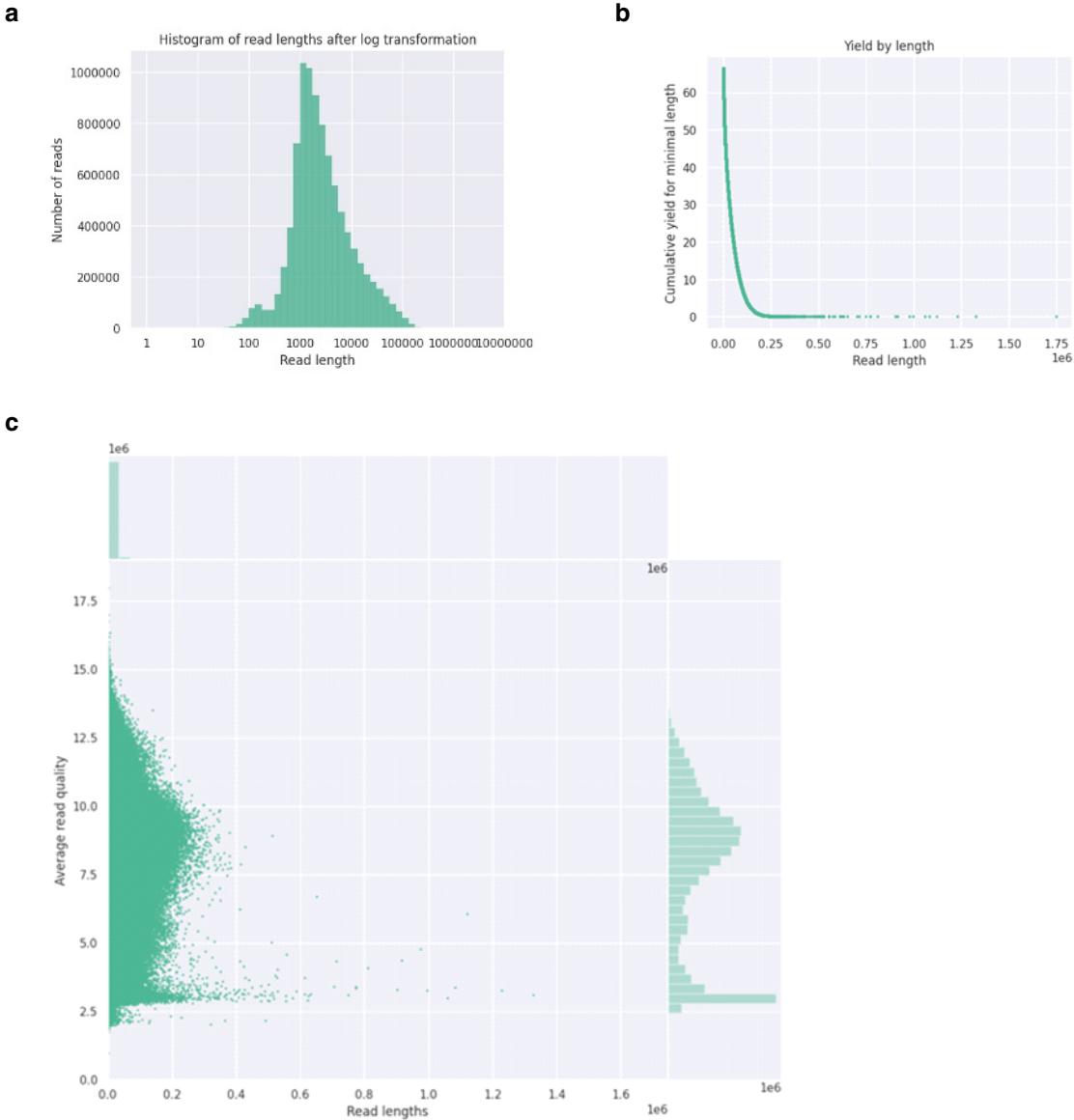
Supplementary Figure 2. Heterozygosity estimation with k-mer analysis.

k-mer profile of *O. sphegodes* based on Illumina reads. The heterozygous portion of the genome is represented by the first peak. The homozygous portion is represented by the second peak. GenomeScope estimates the genome heterozygosity, taking into account observed *k*-mer distribution, sequencing errors, unique sequences, and fits the best *k*-mer model.

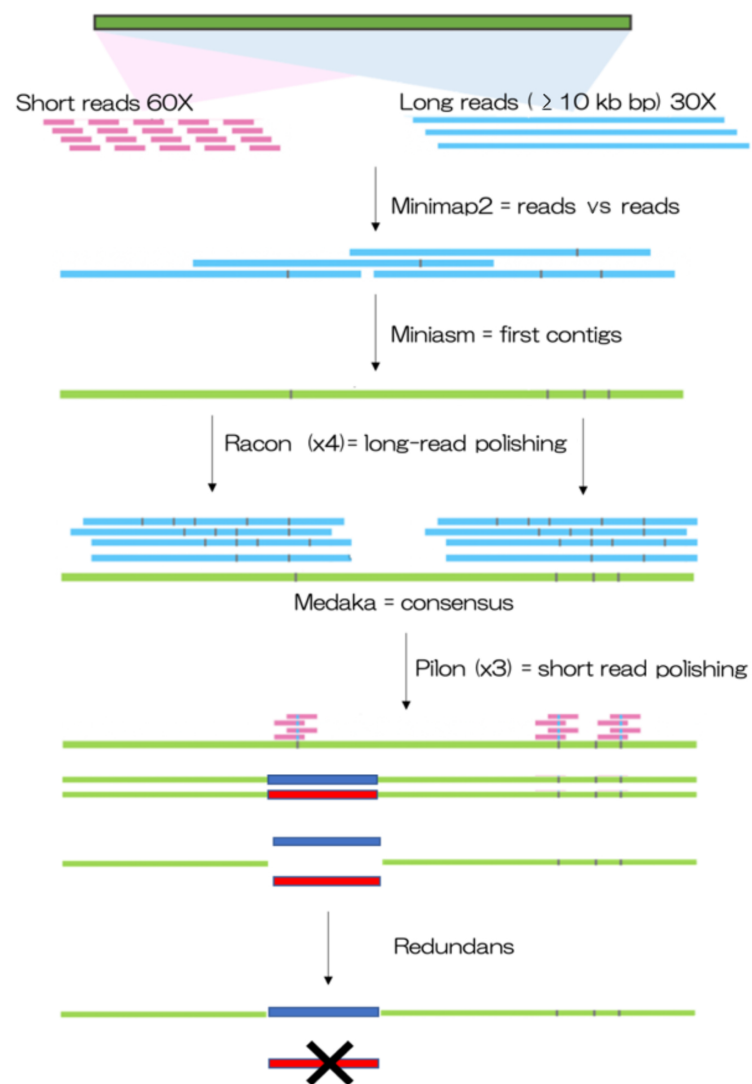


Supplementary Figure 3. Summary of sequencing statistics for the Nanopore data.

Sequencing summary for PromethION dataset 4. **a.** Histogram of log-transformed read length distribution. **b.** Cumulative yield in Gb per read length. **c.** Bivariate plot of read length compared to the average read quality

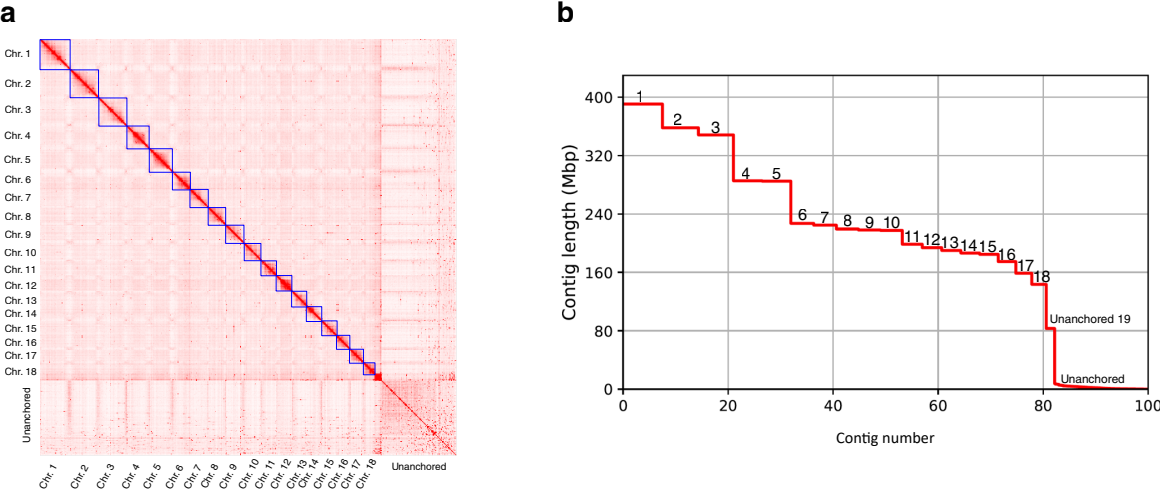


Supplementary Figure 4. Genome assembly strategy



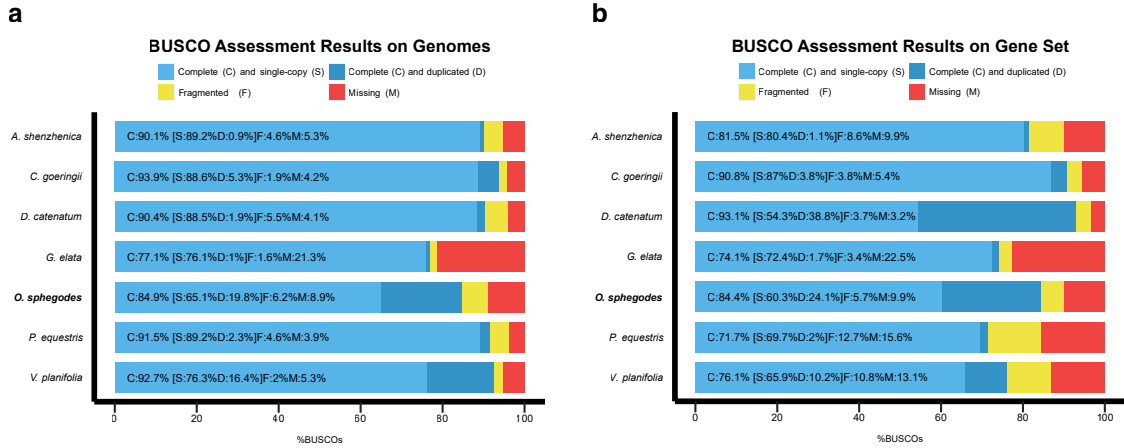
Supplementary Figure 5. Assembled chromosomes.

a. Contact maps of chromosomes from Hi-C. **b.** Chromosome length in Mbp per percentage of genome length. 82.2% of the genome is comprise in the first 19 scaffolds



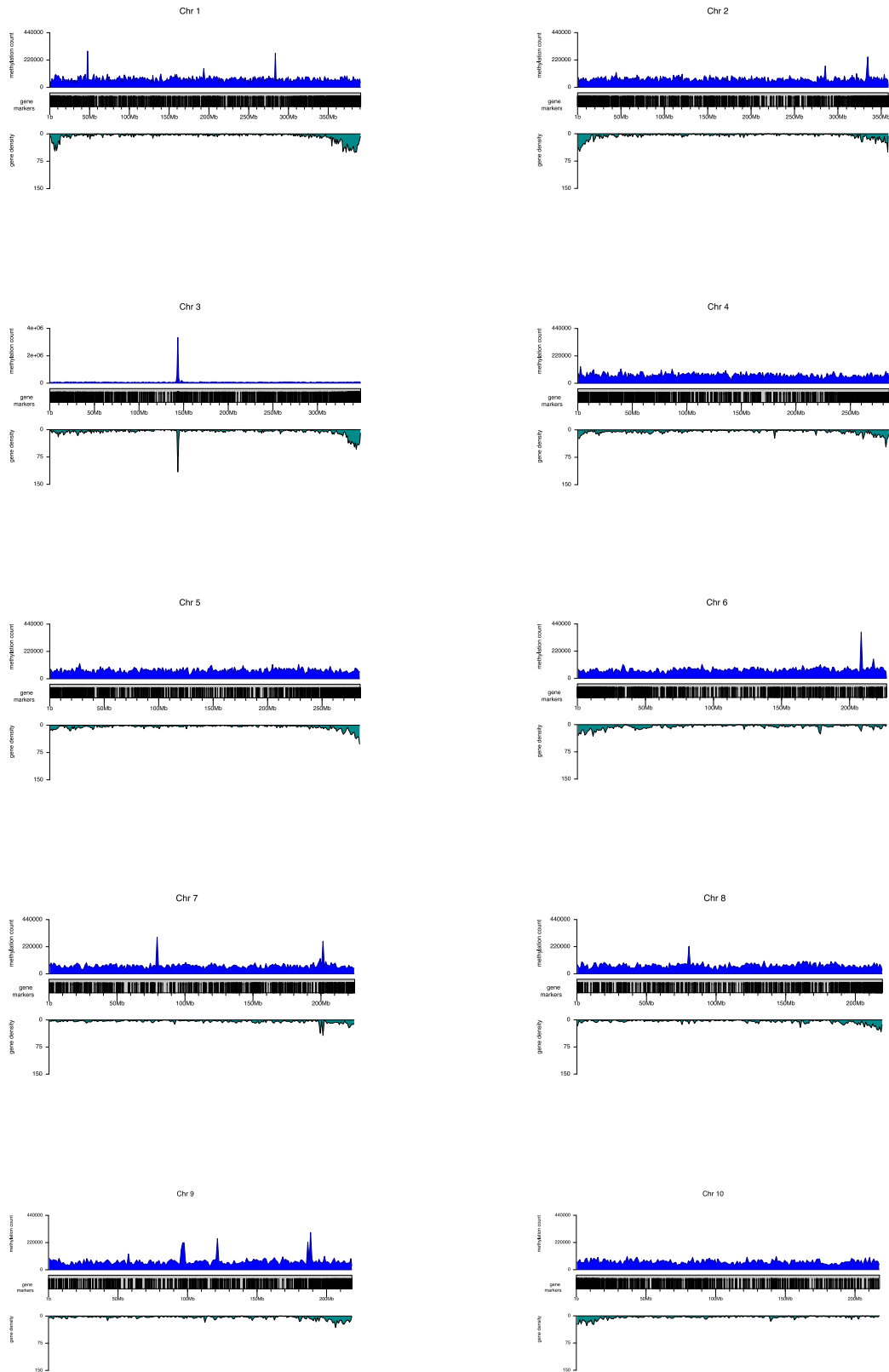
Supplementary Figure 6. Comparison of BUSCO single-copy orthologous genes across orchids.

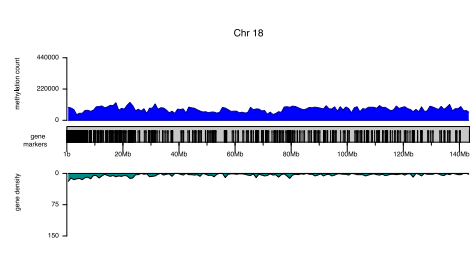
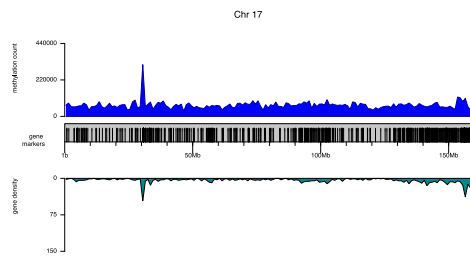
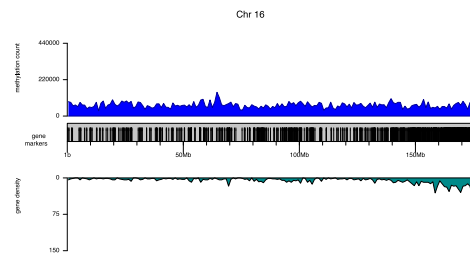
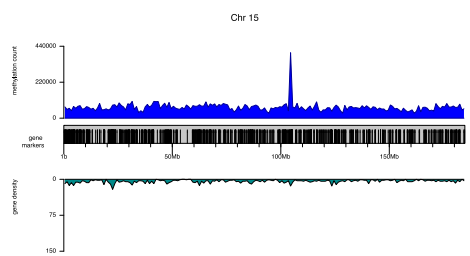
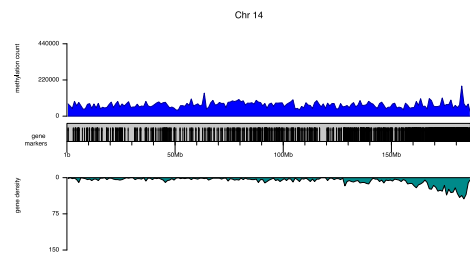
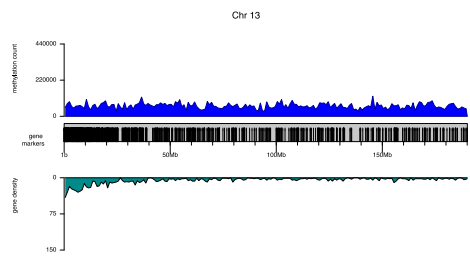
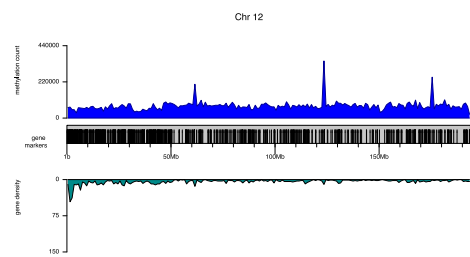
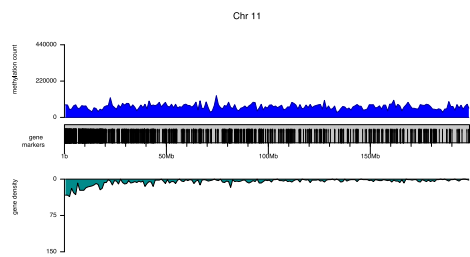
a. BUSCO results across orchid genomes. **b.** BUSCO results across orchid proteomes. BUSCO assessment was performed with the embryophyta OrthoDB_10 database.



Supplementary Figure 7. Methylation patterns across all chromosomes.

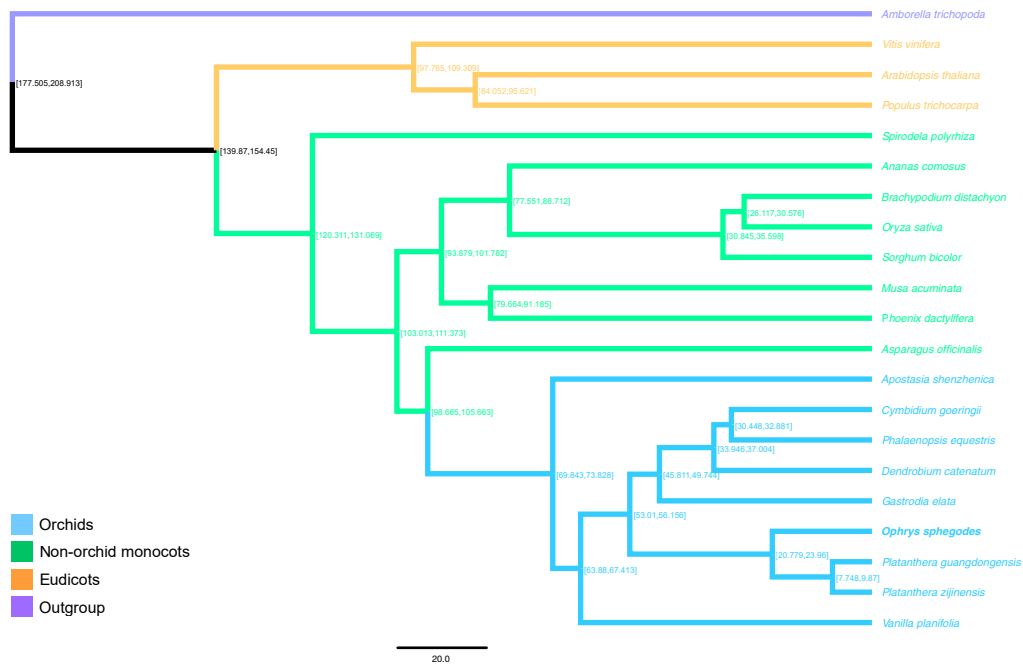
Cytosine methylation was called from Nanopore data. Methylation count and gene density were calculated per 1Mb window and plotted along chromosomes.





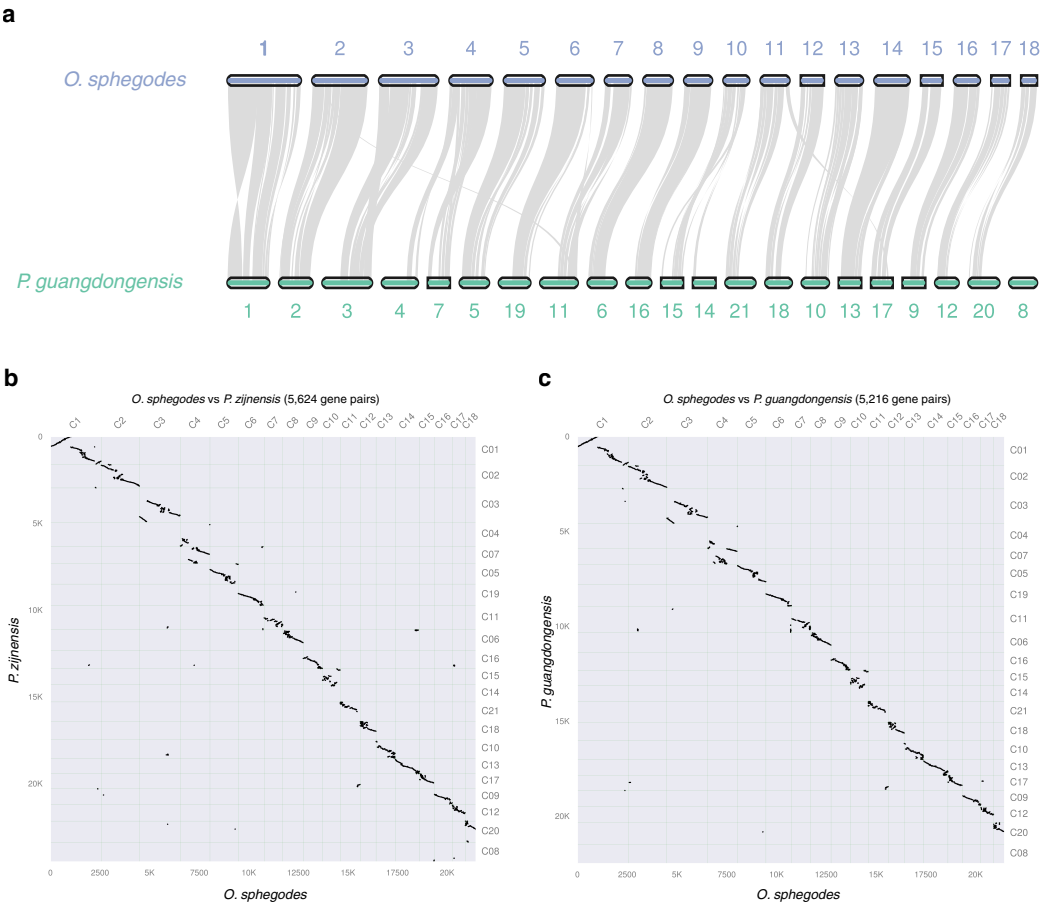
Supplementary Figure 8. Phylogenomic tree of 21 plant species showing 95% confidence intervals.

Related to Fig. 2a in the main text.



Supplementary Figure 9. Synteny analysis

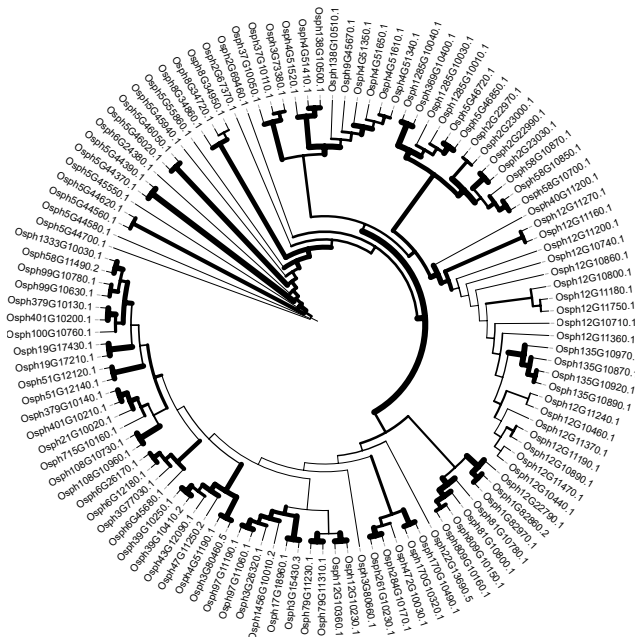
a. Synteny analysis between *O. sphegodes* and *P. guangdongensis*. **b** and **c** show collinearity analysis between *O. sphegodes* and *P. zijinensis* and *P. guangdongensis*, respectively



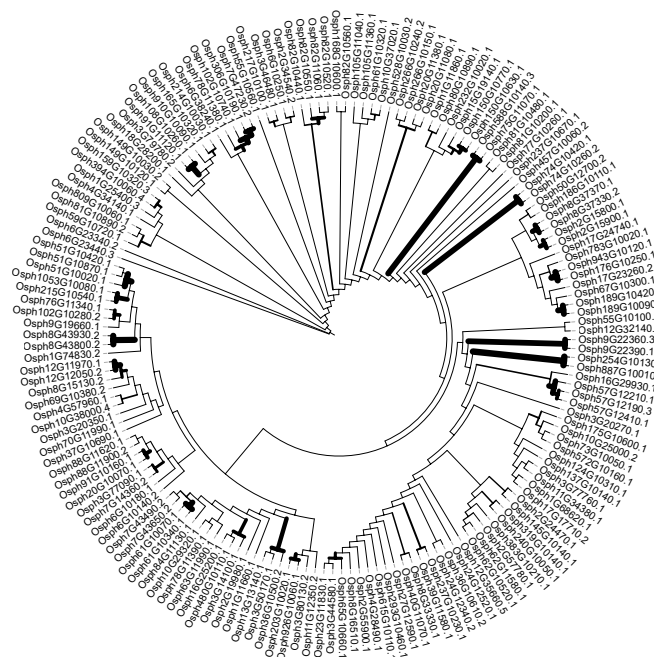
Supplementary Figure 10. Phylogenetic analyses of two transcription factor gene families showing significant gene expansion.

Line thickness indicates bootstrap support. **a** MADS-box family genes. **b** MYB family genes.

a

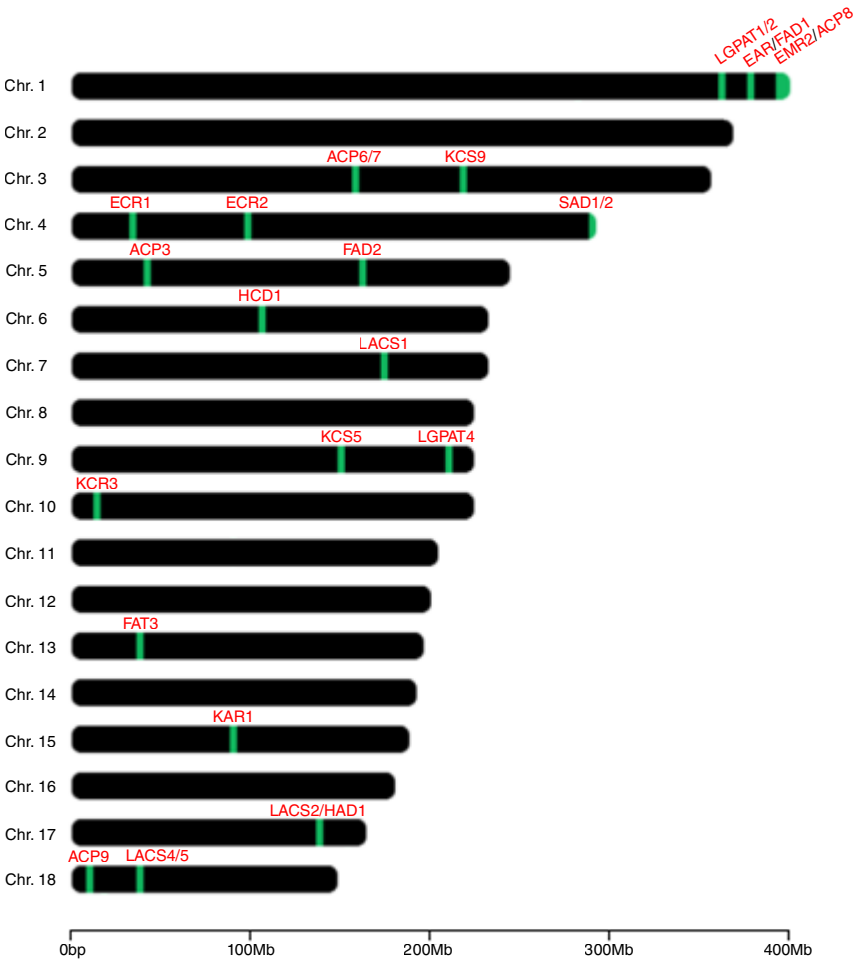


b



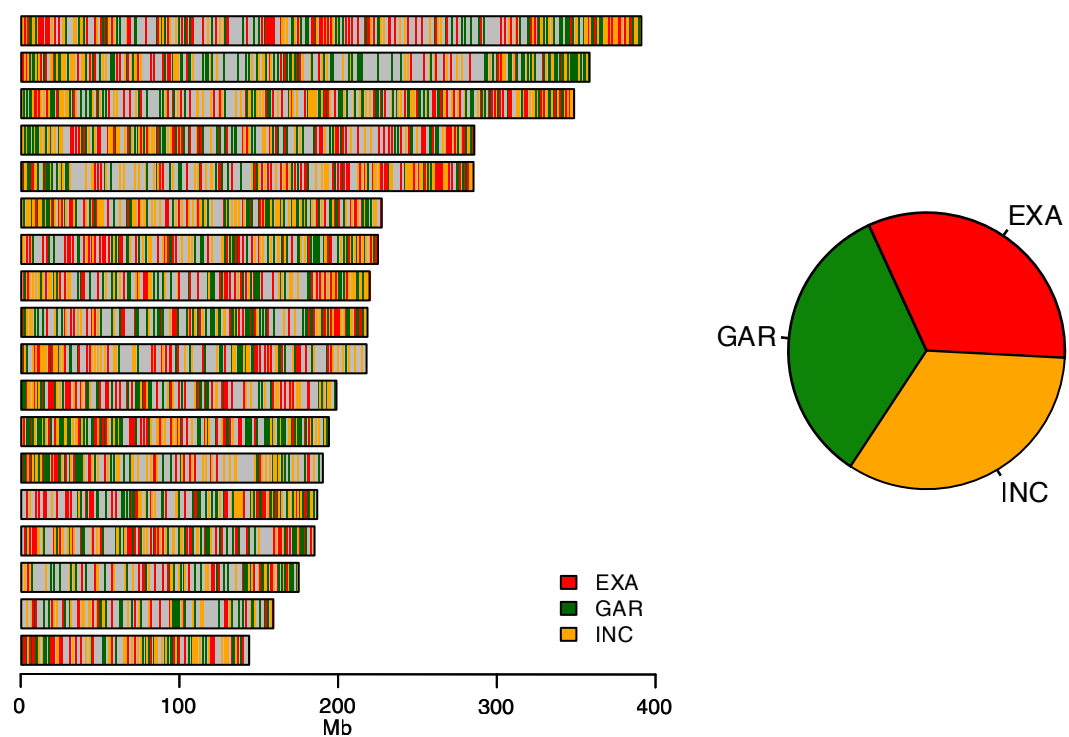
Supplementary Figure 11. Positions of candidate genes

Positions of candidate genes for hydrocarbon biosynthesis and pollinator attraction are not clustered in the genome overall but show local gene duplications.



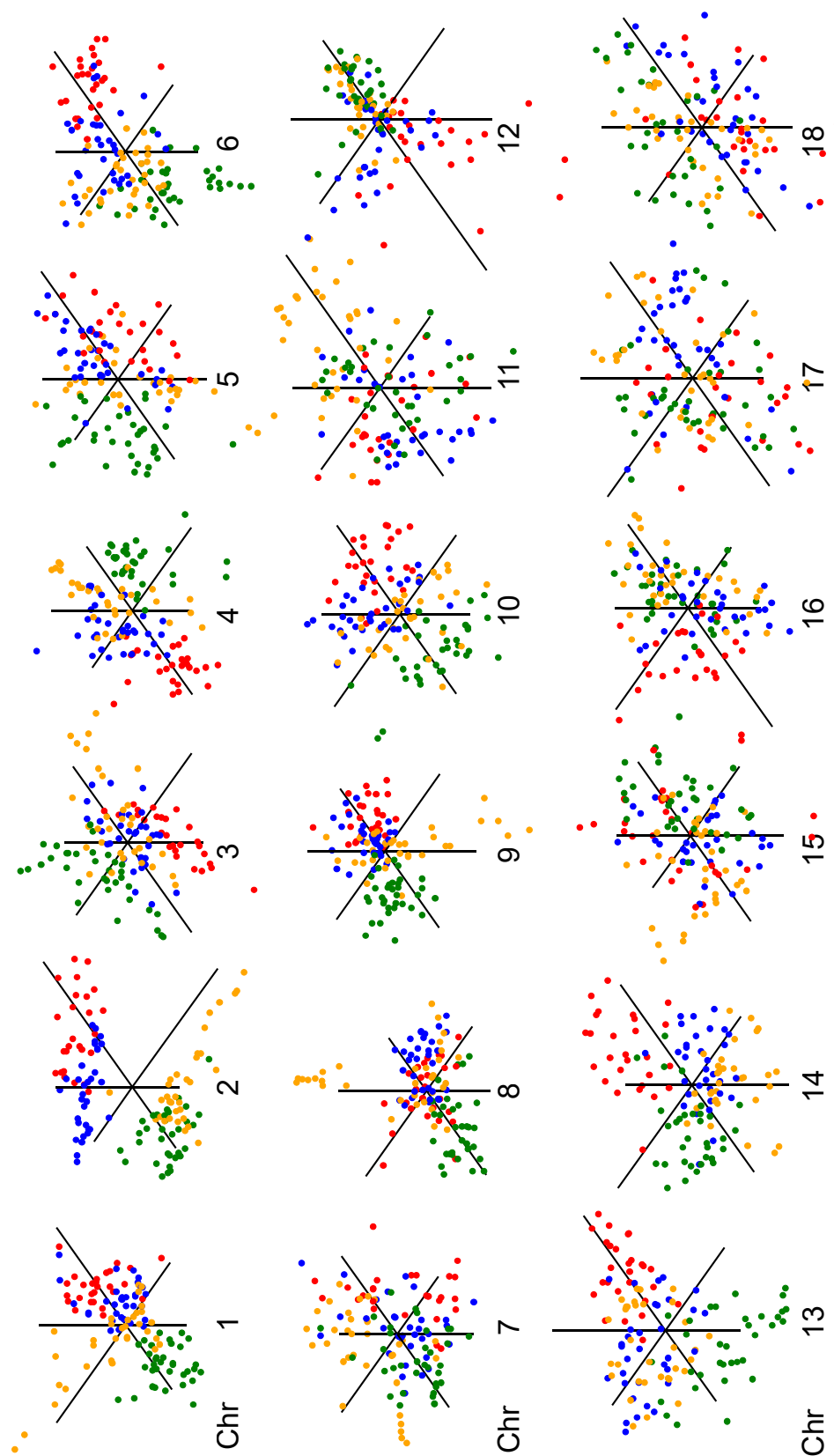
Supplementary Figure 12. Most similar species to *O. sphegodes* per 1 Mb window.

a. Most similar species as per chord distance analysis along *O. sphegodes* chromosomes indicated in colour. **b.** Pie chart showing the proportions of genome windows per most similar species. *O. exaltata* (EXA), *O. garganica* (GAR), *O. incubacea* (INC) are shown in red, green, yellow, respectively.



Supplementary Figure 13. GBS PCoA plots for each chromosome.

Species are colour coded as in Fig. 4.



Supplementary Tables

Supplementary Table 1. Flow cytometry measurements.

Flow cytometry data for *O. sphegodes* pollinia ('Sample', haploid) measured on a Cytoflex S machine, Beckman Coulter, using *Solanum lycopersicum* cultivar 'Stupicke polni tyckove rane' as a reference ('Ref') species. Each row represents a different biological *O. sphegodes* replicate. Run1, Run2, Run3 represent three technical replicates. Sample 1C is the average of Run1, Run2, Run3 per biological sample.

Ref 1C	Sample 1C	Interun CV%	Intrarun CV%	Haploid size (Mb)	Run1 1C	Run2 1C	Run3 1C
0.98	4.980	0.82	1.84	4870	5.025	4.970	4.944
0.98	4.914	0.40	1.79	4806	4.933	4.915	4.894
0.98	4.924	0.98	1.88	4816	4.978	4.909	4.885

CV = Coefficient of Variation

1C = DNA amount in picogram of the haploid nucleus

Supplementary Table 2. Statistics of sequencing raw data from Oxford Nanopore platforms.

Sequencing platform	Flow cell	Total read number	Total length (Gbp)
PromethION	1	5 720 322	90 Gb
PromethION	2	1 049 368	24 Gb
PromethION	3	4 765 204	88 Gb
PromethION	4	9 131 684	126 Gb
PromethION	5	1 467 129	23 Gb
PromethION	6	3 065 912	53 Gb
MinION Mk1B	1	234 692	2.7 Gb
MinION Mk1B	2	234 637	1.8 Gb
Total		25 199 619	409 Gb

472

Supplementary Table 3. Statistics of sequencing raw data from Illumina platforms.

Illumina Sequencing Platform	Individual	Library type	Read length (nt)	Files	Total base pairs after trimming
NovaSeq 6000	SPH_8	WGS PE	2 x 250	20190807.B-Alessia_3_R1/R2	328 059 212
NovaSeq 6000	SPH_8	WGS PE	2 x 150	20200128.A-Alessia_3_R1/R2	83 183 820 328
HiSeq 2500	SPH_511B	WGS PE	2 x 125	20170124.A-Sph511B_R1/R2	6 666 301 185
HiSeq 4000	SPH_511B	WGS PE	2 x 150	20170207.A-Sph511B_R1/R2	56 150 212 140
HiSeq 4000	SPH_511B	WGS PE	2 x 150	20170731.B-Sph511B_R1/R2	72 122 655 460
HiSeq 4000	SPH_511B	WGS PE	2 x 150	20171116.A-Sph511B_R1/R2	49 523 800 508
NovaSeq 6000	SPH_2	Hi-C PE	2 x 150	20191216.B-Ophrys_R1/R2	30 885 826
NovaSeq 6000	SPH_2	Hi-C PE	2 x 150	20200520.B-Ophrys_R1/R2	90 929 210
NovaSeq 6000	SPH_2	Hi-C PE	2 x 150	20200527.B-Ophrys_R1/R2	1 021 481 422

473

WGS, whole genome sequencing; PE, paired-end.

Supplementary Table 4. Summary of the *Ophrys sphegodes* genome assembly

Assembly version	<i>Osph-v0.7</i>	<i>Osph-v1.0</i>	<i>Osph-v1.2</i>	<i>Osph-v2.3</i>
Total contigs	11 148	8 074	2 511	1 500
Total length (bp)	6 429 540 610	5 215 626 102	5 218 768 602	5 219 753 694
Longest contig (Mb)	7.3	7.3	51.3	390.7
N50	754 kb	908 kb	4.8 Mb	218 Mb
N75	402 kb	537 kb	2.6 Mb	159 Mb
L50	2 466	1 759	308	10
L75	5 385	3 618	672	17

The versions of the *O. sphegodes* genome assembly refer to: v0.7, after polishing but before removal of under-collapsed contigs; v1.0, after removal of under-collapsed heterozygous contigs; v1.2, after Hi-C scaffolding with SALSA; v2.3, final assembly after 3D-DNA and HiC-Hiker step.

Supplementary Table 5. Chromosome lengths in the *Ophrys sphegodes* genome assembly

Chromosome/Contig	Total length (bp)
Chromosome 1	390 740 925
Chromosome 2	358 121 781
Chromosome 3	348 374 665
Chromosome 4	285 498 563
Chromosome 5	237 963 090
Chromosome 6	227 181 086
Chromosome 7	224 791 954
Chromosome 8	219 458 497
Chromosome 9	218 188 670
Chromosome 10	217 594 550
Chromosome 11	198 636 841
Chromosome 12	193 889 286
Chromosome 13	190 071 460
Chromosome 14	186 519 188
Chromosome 15	183 697 217
Chromosome 16	174 855 801
Chromosome 17	158 874 247
Chromosome 18	143 640 109
Scaffold 19	83 227 664
Unanchored contigs (N= 1481)	92 406 681

Supplementary Table 6. Mapping rate of Illumina reads against the genome

A paired-end read is considered properly mapped when both mates map to the same region (Properly paired %).

Data set	All mapped (%)	Properly paired (%)
20170124.A-Sph511B	98.05	83.13
20170207.A-Sph511B	96.34	82.68
20170731.B-Sph511B	96.63	83.02
20171116.A-Sph511B	96.72	84.90
20190807.B-Alessia_3	99.73	91.87
20200128.A-Alessia_3	99.28	89.65

Supplementary Table 7. Candidate genes for odour production included in the gene annotation

A priori candidate genes and pseudogenes identified by homology to genes of interest in very long-chain fatty acid or hydrocarbon biosynthesis as well as selected transcription factors. Unique *Ophrys sphegodes* gene identifiers referring to the haploid reference genome are denoted “Osph” followed by chromosome/scaffold number, “G” and gene number (sorted by gene position), “.” and a gene/transcript model number. Where a gene was annotated on an alternative haplotig (as identified by Redundans), the identifier refers to “A” and the haplotig number followed by “:” and the gene name. In several cases, two alleles of the same gene were identified. Letters following a gene name in superscript denote allele designation. Genes/alleles ending in Ψ appear pseudogenised (i.e. with at least one stop codon or frame-shift) when aligned to the full-length sequence of a *bona fide* homologue of interest. Where available, the Comments column adds pipeline-based EC and KO annotation information for biosynthetic genes or TF database annotation for putative regulators.

Some of the *SAD* loci identified in the *Ophrys* genome correspond to published sequences previously assumed to be alleles (see ref.⁴¹) rather than paralogues. We here note the correspondence of genomic loci to these previously identified sequences (arrow symbols). The present naming aims to preserve the gene names for functionally characterised *SAD* sequences encoding enzymatically active proteins^{40,61}.

Gene identifier	Gene/allele name	Description	Comments
VLCFA and hydrocarbon biosynthesis			
Osph222G10140.1	ACP1	acyl carrier protein	
Osph222G10400.1	ACP2		
Osph5G20630.1 / OsphA969:ACP3P	ACP3 / ACP3 Ψ		
Osph217G10600.1	ACP4		
Osph517G10100.1	ACP5		
Osph3G43820.1	ACP6		
Osph3G43810.1	ACP7 Ψ		
Osph1G82220.1	ACP8 Ψ		
Osph18G13840.1	ACP9 Ψ		
- / OsphA2986:ACP10P	- / ACP10 Ψ		
Osph169G10610.1 / OsphA3:AD1P	AD1 / AD1 Ψ	aldehyde decarbonylase (<i>A. thaliana</i> CER1-homologue)	
Osph169G10400.1 / OsphA3:AD2-Alt	AD2 / AD2 ^{Alt}		4.1.99.5 / K15404
Osph169G10470.1 / OsphA3:AD3P-Alt	AD3 Ψ / AD3 Ψ ^{Alt}		
Osph152G10450.1	CER3H1	<i>A. thaliana</i> CER3-homologue (putative aldehyde-forming acyl-CoA reductase)	
Osph152G10590.1	CER3H2 Ψ		
Osph21G10650.1	CYTB5H1	cytochrome b ₅ homologue	
Osph21G10840.2	CYTB5H2		
Osph1G68690.1	EAR1	enoyl-ACP reductase	
Osph4G18700.1 / OsphA1371:ECR1P	ECR1 / ECR1 Ψ	enoyl-CoA reductase	
Osph4G30620.1	ECR2		1.3.1.93
Osph62G12220.1	EMR1	enoyl-CoA or mitochondrial enoyl-ACP reductase	
Osph1G80960.1 / OsphA384:EMR2P-Alt	EMR2 Ψ / EMR2 Ψ ^{Alt}		1.3.1.38 / K07512
Osph1G68820.1	FAD1	(membrane-bound) fatty acid desaturase	1.14.19.6
Osph5G32390.1 / OsphA419:FAD2P-Alt	FAD2 Ψ / FAD2 Ψ ^{Alt}		
Osph33G12220.1	FAD3		
Osph33G11900.1	FAD4		

Osph89G11970.1	<i>FAT1</i>	fatty acyl-ACP thioesterase	3.1.2.14; 3.1.2.21 / K10781
Osph219G10030.1	<i>FAT2</i>		
Osph13G23960.1	<i>FAT3Ψ</i>		
Osph17G26180.1 / OsphA2755:HAD1P	<i>HAD1 / HAD1Ψ</i>	3-hydroxyacyl-ACP dehydratase	4.2.1.59 / K02372
Osph6G35180.1	<i>HCD1</i>	3-hydroxyacyl-CoA dehydratase	4.2.1.134 / K10703
Osph15G26270.1 / OsphA1489:KAR1I	<i>KAR1Ψ / KAR1^I</i>	β-ketoacyl-ACP reductase	
Osph33G10850.1 / OsphA4052:KASii1	<i>KASii1Ψ / KASii1</i>	β-ketoacyl-ACP synthase II	
Osph400G10190.1	<i>KCR1</i>	β-ketoacyl-CoA reductase	
Osph385G10100.1	<i>KCR2Ψ</i>		
Osph10G16090.1	<i>KCR3Ψ</i>		1.1.1.330
Osph160G10230.1 / OsphA3894:KCS1P	<i>KCS1 / KCS1Ψ</i>	β-ketoacyl-CoA synthase	2.3.1.199 / K15397
Osph114G10860.1	<i>KCS2Ψ</i>		
Osph76G10840.1 / OsphA1338:KCS3P	<i>KCS3 / KCS3Ψ</i>		
Osph33G10370.1	<i>KCS4</i>		2.3.1.199 / K15397
Osph9G29660.1	<i>KCS5</i>		2.3.1.199 / K15397
Osph66G10710.1	<i>KCS6</i>		
Osph30G10460.1	<i>KCS7</i>		
Osph339G10090.1	<i>KCS8Ψ</i>		
Osph3G50760.1	<i>KCS9</i>		2.3.1.199 / K15397
Osph30G10690.1	<i>KCS10</i>		2.3.1.199 / K15397
Osph114G10750.1	<i>KCS11Ψ</i>		
- / OsphA120:KCS12	- / <i>KCS12</i>		
Osph36G11570.1 / OsphA61:KCS13	<i>KCS13Ψ / KCS13</i>		
Osph33G10680.1	<i>KCS14Ψ</i>		
Osph176G10770.1	<i>KCS15</i>		2.3.1.199 / K15397
Osph395G10070.1	<i>KCS16Ψ</i>		
Osph30G11470.1	<i>KCS17</i>		
Osph339G10150.1	<i>KCS18</i>		
Osph30G11040.1	<i>KCS19Ψ</i>		
Osph7G31430.1	<i>LACS1</i>	long-chain acyl-CoA synthetase	6.2.1.3 / K01897
Osph17G26020.1	<i>LACS2</i>		6.2.1.3 / K01897
- / OsphA358:LACS3	- / <i>LACS3</i>		
Osph18G18560.1	<i>LACS4Ψ</i>		
Osph18G18650.2	<i>LACS5Ψ</i>		
Osph1G61170.1	<i>LGPAT1</i>	lysophosphatidic acid/glycerol 3-phosphate acyltransferase	K13508
Osph1G61400.1	<i>LGPAT2Ψ</i>		
Osph332G10160.1	<i>LGPAT3</i>		
Osph9G41940.1	<i>LGPAT4Ψ</i>		
Osph4G59510.1 / OsphA1798:SAD1-Alt	<i>SAD1 / SAD1^{Alt}</i>	(soluble) stearoyl-ACP (or acyl-ACP) desaturase	→ <i>SAD1-B</i>
Osph4G59550.1	<i>SAD2</i>		1.14.19.2; 1.14.19.26 / K03921; → <i>SAD2-A</i>
Osph5G55220.1 / OsphA241:SAD3	<i>SAD3Ψ / SAD3</i>		→ <i>SAD3</i>
Osph75G10230.1	<i>SAD4</i>		1.14.19.2; 1.14.19.26 / K03921;

Osph250G10230.1	<i>SAD5^ψ</i>		→ <i>SAD4</i>
Osph210G10040.1 ^a	<i>SAD6</i>		→ <i>SAD5-A</i>
Osph4G59582.1 / OsphA1441: <i>SAD7</i>	<i>SAD7^ψ / SAD7</i>		→ <i>SAD6-B</i> ^c
Osph4G59580.1 / OsphA1441: <i>SAD8P</i>	<i>SAD8 / SAD8^ψ</i>		→ <i>SAD1-A</i>
Osph132G10080.1	<i>SAD9^ψ</i>		→ <i>SAD2-B</i>
Osph886G10040.1 ^b	<i>SAD10</i>		→ <i>SAD5-A</i>
Osph1397G10050.1	<i>SAD11I^ψ</i>		→ <i>SAD5-B</i> ^c
			→ <i>SAD6-B</i>
Transcription factors of interest			
Osph97G11060.1	<i>DEF1H1</i>	DEF/AP3 homologue (homologues of <i>Orchis italica</i> <i>DEF1-DEF4</i>) ^{62,63}	MADS→MADS- MIKC
Osph97G11190.1	<i>DEF1H2</i>		MADS→MADS- MIKC
Osph1456G10010.1	<i>DEF2H1</i>		
Osph1G46670.1	<i>DEF2H2^ψ</i>		
Osph3G26320.1	<i>DEF3H1</i>		MADS→MADS- MIKC
Osph3G26350.1	<i>DEF3H2^ψ</i>		
Osph18G15220.1	<i>DEF4H1^ψ</i>		
Osph17G28910.1	<i>GL3H1^ψ</i>	<i>Arabidopsis thaliana</i> GL3 homologue	bHLH
Osph17G29140.1	<i>GL3H2^ψ</i>		
Osph11G31450.1	<i>LFY1</i>	<i>LFY/FLO</i> ^{64,65}	
Osph11G31400.1	<i>LFY2^ψ</i>		
Osph59G10460.1	<i>ML1H1</i>	<i>Dendrobium crumenatum</i> MYB MIXTA-LIKE 1 (MYBML1) homologue ⁶⁶	
Osph59G10560.1	<i>ML1H2^ψ</i>		
Osph168G10600.1 / OsphA63:MYB12H1	<i>MYB11H1^ψ / MYB11H1</i>	<i>Arabidopsis thaliana</i> MYB11 homologue	MYB→MYB
Osph2G70710.1	<i>SPL8H1</i>	<i>Oryza sativa</i> SPL8 homologue	
Osph2G70890.1	<i>SPL8H2</i>		
Osph1G68420.1	<i>SPL9H1</i>	<i>Arabidopsis thaliana</i> SPL9 homologue	
Osph1G68550.1	<i>SPL9H2</i>		
Osph6G51030.1	<i>SPY1</i>	<i>Arabidopsis thaliana</i> SPY homologue	

^a gene model for *SAD6*, corresponding to a previously cloned sequence⁴¹, has exons 1 and 2 annotated on the + strand of contig 210 and exon 3 on the - strand; ^b gene model for *SAD10* spans two contigs (contigs 886 and 210), with break in the second intron and with first coding exon on contig 886; ^c These sequences are full-length and without stop codons, but do not appear to be enzymatically active and were hence previously categorised as non-functional^{41,61}.

Supplementary Table 8. Additional candidate genes identified from annotation data.

Additional genes of interest putatively involved in anthocyanin biosynthesis, carotenoid biosynthesis, or fatty acid, wax or hydrocarbon biosynthesis, identified from KEGG orthology (KO) and EC annotation.

Gene ID	Gene name	KO	EC
Anthocyanin biosynthesis			
<i>Anthocyanin 3-O-glucoside 2"-O-xylosyltransferase</i>			
Osph125G10550.1	AGT1	K17193	2.4.2.51
Anthocyanin synthase			
Osph3G27250.1	ANS1	K05277	1.14.20.4
Chalcone isomerase			
Osph9G23120.1	CHI1	K01859	5.5.1.6
Osph874G10050.1	CHI2	K01859	5.5.1.6
Chalcone synthase			
Osph12G18370.1	CHS1	K00660	2.3.1.74
Osph12G18400.1	CHS2	K00660	2.3.1.74
Osph15G18650.1	CHS3	K00660	2.3.1.74
Osph16G37380.1	CHS4	K00660	2.3.1.74
Osph23G12030.1	CHS5	K00660	2.3.1.74
Osph23G12360.1	CHS6	K00660	2.3.1.74
Osph77G10730.1	CHS7	K00660	2.3.1.74
Dihydroflavonol 4-reductase			
Osph14G43650.5	DFR1	K13082	1.1.1.219
Flavanone 3-hydroxylase			
Osph5G19490.1	F3H1	K00475	1.14.11.9
Osph5G19880.1	F3H2	K00475	1.14.11.9
Flavonoid 3'-hydroxylase			
Osph16G33680.1	F3'H1	K05280	1.14.14.82
Osph120G10490.3	F3'H2	K05280	1.14.14.82
Osph120G10890.1	F3'H3	K05280	1.14.14.82
Osph141G11050.1	F3'H4	K05280	1.14.14.82
Flavonol synthase			
Osph68G11120.1	FLS1	K05278	1.14.20.6
UDP-glucose:flavonol 3-O-glucosyltransferase			
Osph70G12020.1	UF3GT1	K10757	2.4.1.91
Carotenoid biosynthesis			
<i>ABA2 homologue (xanthoxin dehydrogenase)</i>			
Osph2G11980.1	ABA2H1	K09841	1.1.1.288
β-carotene isomerase			
Osph339G10010.1	BCI1	K17911	5.2.1.14
9-cis-β-carotene 9',10'-cleaving dioxygenase			
Osph1G54020.1	CCD7H1	K17912	1.13.11.68
Osph1G54070.1	CCD7H2	K17912	1.13.11.68
Osph80G10400.1	CCD7H3	K17912	1.13.11.68
Osph80G10450.1	CCD7H4	K17912	1.13.11.68
all-trans-10'-apo-β-carotenal 13,14-cleaving dioxygenase / carlactone synthase			
Osph4G21770.1	CCD8H1	K17913	1.13.11.69 / 1.13.11.70
Osph16G30340.1	CCD8H2	K17913	1.13.11.69 / 1.13.11.70
Osph16G30560.1	CCD8H3	K17913	1.13.11.69 / 1.13.11.70
β-carotene 3-hydroxylase			
Osph20G10110.1	CHY1	K15746	1.14.15.24
9-cis-epoxycarotenoid dioxygenase			
Osph1G77610.1	NCED1	K09840	1.13.11.51
Osph4G14880.1	NCED2	K09840	1.13.11.51
Osph18G13090.1	NCED3	K09840	1.13.11.51
Osph98G10650.1	NCED4	K09840	1.13.11.51

Osph98G10880.1	NCED5	K09840	1.13.11.51
15-cis-phytoene desaturase			
Osph219G10330.1	PDS1	K02293	1.3.5.5
Osph219G10360.1	PDS2	K02293	1.3.5.5
15-cis-phytoene synthase			
Osph91G10050.2	PSY1	K02291	2.5.1.32
violaxanthin de-epoxidase			
Osph10G16730.1	VDE1	K09839	1.23.5.1
zeaxanthin epoxidase			
Osph68G11190.1	ZEP1	K09838	1.14.15.21
ζ-carotene isomerase			
Osph10G20590.1	ZISO1	K15744	5.2.1.12
Fatty acid, wax and hydrocarbon biosynthesis			
diacylglycerol O-acyltransferase			
Osph1G69350.1	DGAT1	K22848	2.3.1.20
Osph16G32530.1	DGAT2	K22849	2.3.1.20
Osph16G32630.1	DGAT3	K22849	2.3.1.20
Osph56G12460.2	DGAT4	K11155	2.3.1.20 / 2.3.1.75 / 2.3.1.76
Osph116G10900.1	DGAT5	K22848	2.3.1.20
Osph157G11420.1	DGAT6	K22848	2.3.1.20
VLC enoyl-CoA reductase			
Osph9G12320.4	ECR3	K10258	1.3.1.93
fatty acid elongase (ELO-type)			
Osph154G10300.1	ELO1	K10245	2.3.1.199
alcohol-forming fatty acyl-CoA reductase			
Osph5G31110.1	FAR1	K13356	1.2.1.84
Osph5G51550.1	FAR2	K13356	1.2.1.84
Osph133G10230.1	FAR3	K13356	1.2.1.84
Osph279G10300.2	FAR4	K13356	1.2.1.84
Osph279G10310.1	FAR5	K13356	1.2.1.84
Osph578G10090.1	FAR6	K13356	1.2.1.84
Osph578G10170.1	FAR7	K13356	1.2.1.84
fatty acyl-ACP thioesterase			
Osph3G55170.2	FAT4	K10781	3.1.2.14 / 3.1.2.21
Osph5G55210.3	FAT5	K10782	3.1.2.14
fatty acyl-CoA thioesterase			
Osph3G76480.2	FCT1	K01068	3.1.2.2
VLC 3-hydroxyacyl-CoA dehydratase			
Osph57G11660.1	HCD2	K10703	4.2.1.134
β-ketoacyl-ACP synthases			
Osph16G37920.2	KASi1	K09458	2.3.1.179
Osph81G10950.1	KASi2	K00648	2.3.1.180
Osph97G10840.1	KASi3	K09458	2.3.1.179
Osph97G10700.1	KASiii1	K09458	2.3.1.179
lysophosphatidic acid or glycerol 3-phosphate acyltransferase			
Osph8G12500.1	LGPAT5	K00655	2.3.1.51
Osph50G12130.1	LGPAT6	K13519	2.3.1.51 / 2.3.1.23 / 2.3.1.-
Osph103G10680.1	LGPAT7	K13513	2.3.1.- / 2.3.1.51
Osph282G10060.1	LGPAT8	K13523	2.3.1.51 / 2.3.1.-

525
526

Supplementary Table 9. List of transcription factors identified

Family	Type	N	Genes
Alfin-like	TF	6	Osph3G80420.1; Osph5G27550.1; Osph9G38300.1; Osph14G44120.2; Osph93G11800.4; Osph170G10980.3
AP2/ERF-AP2	TF	12	Osph1G73530.4; Osph2G66470.1; Osph2G66920.2; Osph11G14450.3; Osph37G11040.4; Osph45G10630.1; Osph45G10780.1; Osph152G10120.2; Osph300G10140.1; Osph639G10090.1; Osph802G10070.1; Osph1057G10040.2
AP2/ERF-ERF	TF	85	Osph1G10850.1; Osph1G55610.1; Osph2G47170.1; Osph2G64430.1; Osph3G67220.1; Osph3G68770.1; Osph4G27470.2; Osph4G52140.1; Osph4G58900.1; Osph5G15890.1; Osph5G15990.1; Osph5G57030.1; Osph5G57340.1; Osph5G57680.1; Osph6G11020.1; Osph6G11030.1; Osph6G11230.1; Osph6G11250.1; Osph6G41240.1; Osph6G50480.1; Osph7G36620.1; Osph8G31780.1; Osph10G25260.1; Osph10G39700.1; Osph12G16290.1; Osph14G47500.1; Osph16G38520.1; Osph17G25290.1; Osph17G27760.1; Osph17G27820.1; Osph17G27850.1; Osph18G15950.1; Osph20G10440.1; Osph20G13630.1; Osph20G14350.1; Osph20G14380.1; Osph20G14780.1; Osph20G14820.1; Osph21G12250.1; Osph26G11120.1; Osph30G12100.1; Osph30G12360.1; Osph31G10070.1; Osph31G10120.1; Osph35G12510.1; Osph38G12090.1; Osph38G12880.1; Osph39G11090.1; Osph43G10560.1; Osph43G10760.1; Osph45G10600.1; Osph45G11360.1; Osph45G11840.1; Osph47G11140.1; Osph61G11290.1; Osph66G11740.1; Osph68G10310.1; Osph68G10560.1; Osph70G10950.1; Osph74G11090.1; Osph74G11460.1; Osph80G10420.1; Osph100G10040.1; Osph118G10170.1; Osph123G11120.1; Osph149G11100.1; Osph159G10460.1; Osph160G10800.1; Osph219G10060.1; Osph219G10130.1; Osph225G10540.1; Osph227G10340.1; Osph237G10120.1; Osph237G10430.1; Osph254G10640.1; Osph356G10100.1; Osph370G10100.1; Osph414G10010.1; Osph505G10020.1; Osph520G10040.1; Osph751G10030.1; Osph920G10140.1; Osph920G10170.1; Osph924G10170.1; Osph1022G10100.2
AP2/ERF-RAV	TF	3	Osph15G20590.1; Osph697G10120.1; Osph773G10040.1
ARID	TR	20	Osph10G14410.1; Osph10G14450.1; Osph10G14480.1; Osph10G14560.1; Osph10G14610.1; Osph10G14650.1; Osph13G10070.1; Osph13G10190.1; Osph34G11130.1; Osph34G11340.2; Osph41G10860.1; Osph42G10220.2; Osph42G10680.1; Osph53G11830.4; Osph53G12110.1; Osph74G12040.1; Osph74G12280.1; Osph87G10320.1; Osph505G10090.1; Osph557G10190.1
AUX/IAA	TR	21	Osph2G39240.1; Osph3G16840.1; Osph3G49760.1; Osph4G23510.1; Osph4G54250.1; Osph5G15830.1; Osph8G44580.1; Osph14G33810.1; Osph14G45130.5; Osph27G10050.2; Osph31G10740.1; Osph36G10680.1; Osph42G11200.2; Osph47G10890.2; Osph48G12040.1; Osph95G11590.1; Osph95G11680.2; Osph139G10730.1; Osph163G10020.1; Osph163G10140.1; Osph487G10010.6
B3	TF	21	Osph1G14570.2; Osph1G42960.1; Osph1G71110.1; Osph1G71200.1; Osph1G71640.1; Osph2G41480.2; Osph11G22330.3; Osph16G30730.1; Osph16G30800.1; Osph27G10990.3; Osph29G10320.1; Osph29G11390.1; Osph36G12160.1; Osph38G12340.1; Osph38G12670.2; Osph41G10110.1; Osph41G11130.2; Osph41G11620.1; Osph66G11910.1; Osph385G10020.1; Osph728G10120.1
B3-ARF	TF	19	Osph2G63210.3; Osph2G64190.2; Osph3G65370.2; Osph3G65580.1; Osph4G54330.1; Osph34G12260.1; Osph34G12880.1; Osph67G10150.1; Osph77G10370.4; Osph77G10380.2; Osph110G10440.2; Osph110G10650.3; Osph137G10650.1; Osph137G10720.1; Osph304G10030.2; Osph362G10190.2; Osph366G10100.1; Osph367G10390.1; Osph403G10220.1
BBR-BPC	TF	11	Osph4G21910.1; Osph6G28540.1; Osph6G28610.1; Osph7G42040.1; Osph8G40040.1; Osph8G40220.1; Osph12G12950.1;

			Osph12G13140.1; Osph12G23540.2; Osph17G24870.1; Osph80G10580.1
BES1	TF	6	Osph3G76510.1; Osph11G11170.1; Osph11G11400.1; Osph14G48050.1; Osph14G48210.1; Osph430G10010.1
bHLH	TF	94	Osph1G18750.2; Osph1G43560.1; Osph1G52050.2; Osph1G64700.2; Osph1G66550.4; Osph1G76330.1; Osph1G76670.1; Osph2G13380.3; Osph2G32530.3; Osph2G45470.2; Osph3G66610.1; Osph3G67020.2; Osph3G80070.1; Osph4G12130.3; Osph4G36780.2; Osph8G29600.1; Osph9G37520.1; Osph9G37600.1; Osph12G16770.1; Osph12G30690.1; Osph13G31730.1; Osph15G32350.1; Osph17G23710.1; Osph17G28910.1; Osph18G16690.1; Osph22G13260.1; Osph22G14270.1; Osph24G10720.1; Osph24G11230.1; Osph26G10610.1; Osph31G10090.1; Osph31G10140.1; Osph35G10660.1; Osph35G11010.1; Osph35G11780.1; Osph35G12220.1; Osph38G12230.1; Osph38G12770.1; Osph43G10320.1; Osph46G10730.1; Osph46G11740.1; Osph51G11770.2; Osph53G12150.2; Osph53G12560.1; Osph53G12630.2; Osph55G10310.1; Osph58G10690.1; Osph59G11850.1; Osph60G10090.1; Osph60G10300.1; Osph62G10210.3; Osph62G10590.1; Osph62G12190.1; Osph63G10840.1; Osph66G10900.1; Osph66G11050.1; Osph69G10350.1; Osph77G10100.1; Osph78G10530.1; Osph86G11180.2; Osph86G11230.1; Osph86G11310.2; Osph90G11290.1; Osph90G11590.1; Osph103G10020.1; Osph103G10340.1; Osph103G10830.1; Osph104G10840.1; Osph104G10870.1; Osph104G11460.1; Osph104G11480.1; Osph118G10910.1; Osph121G10530.1; Osph121G11210.1; Osph125G10090.2; Osph142G10280.1; Osph142G10600.1; Osph144G10890.1; Osph148G10330.1; Osph148G10520.1; Osph154G11110.1; Osph156G11050.3; Osph161G10290.4; Osph161G10550.1; Osph168G10250.1; Osph211G10140.1; Osph211G10280.1; Osph234G10060.1; Osph234G10220.1; Osph287G10110.1; Osph300G10100.1; Osph401G10130.2; Osph644G10020.3; Osph802G10020.1
bZIP	TF	52	Osph1G13300.1; Osph1G45510.2; Osph1G73380.1; Osph2G16410.1; Osph2G16540.1; Osph2G47930.4; Osph4G58830.1; Osph6G19320.4; Osph6G31700.1; Osph6G47560.1; Osph9G24330.1; Osph9G24430.2; Osph10G35970.1; Osph11G15160.1; Osph11G21360.1; Osph12G15550.1; Osph12G15630.1; Osph14G27560.1; Osph34G10670.1; Osph50G11990.1; Osph50G12080.1; Osph50G12100.1; Osph50G12110.1; Osph63G12010.4; Osph66G10600.2; Osph68G11170.1; Osph83G10100.1; Osph84G10120.1; Osph84G10480.1; Osph93G10120.2; Osph93G10280.1; Osph96G10600.1; Osph103G10130.1; Osph103G10410.1; Osph111G11170.4; Osph127G10630.1; Osph127G10890.1; Osph128G10100.1; Osph136G10410.1; Osph145G10330.1; Osph145G10560.1; Osph157G10190.1; Osph183G11440.1; Osph192G10190.1; Osph192G10330.1; Osph211G10060.1; Osph254G10110.1; Osph348G10090.1; Osph348G10210.1; Osph505G10130.1; Osph557G10170.1; Osph707G10060.1
C2C2-CO-like	TF	4	Osph9G26530.2; Osph32G11200.1; Osph32G11260.1; Osph38G10290.1
C2C2-Dof	TF	31	Osph1G71040.1; Osph1G71150.1; Osph1G71850.1; Osph3G56820.1; Osph3G69420.1; Osph3G78750.1; Osph6G10490.1; Osph6G10700.1; Osph6G23470.1; Osph6G35620.1; Osph7G35850.1; Osph9G40220.1; Osph9G44140.1; Osph14G43090.1; Osph15G15460.1; Osph18G12660.1; Osph43G10890.1; Osph43G11090.1; Osph50G10380.1; Osph59G10830.1; Osph82G10910.1; Osph94G11370.1; Osph94G11520.1; Osph121G10600.1; Osph121G11160.1; Osph127G10710.1; Osph127G11000.1; Osph300G10270.1; Osph319G10230.1; Osph512G10290.2; Osph988G10080.1
C2C2-GATA	TF	29	Osph1G79210.1; Osph2G34600.2; Osph10G11600.1; Osph10G11850.1; Osph12G32280.1; Osph13G10680.1; Osph13G15760.2; Osph15G22520.1; Osph19G17030.1; Osph24G12690.2; Osph24G12850.2; Osph28G10170.1;

			Osph28G10470.1; Osph28G12110.1; Osph28G12240.1; Osph30G10740.1; Osph30G11010.1; Osph37G10940.1; Osph40G10970.1; Osph51G12360.1; Osph56G12020.1; Osph80G11210.1; Osph121G10680.1; Osph124G10790.1; Osph133G11150.1; Osph137G10460.1; Osph137G10620.1; Osph197G10030.1; Osph257G10340.1
C2C2-LSD	TF	4	Osph14G30510.1; Osph33G12500.1; Osph52G10250.3; Osph572G10040.2
C2C2-YABBY	TF	15	Osph1G49910.1; Osph1G49940.1; Osph1G82340.1; Osph1G82610.1; Osph4G17370.3; Osph5G52750.1; Osph5G52760.1; Osph5G53310.1; Osph28G10760.2; Osph28G10990.1; Osph87G10600.2; Osph250G10030.3; Osph250G10090.1; Osph379G10190.2; Osph401G10290.1
C2H2	TF	118	Osph1G23800.1; Osph1G52550.1; Osph1G61090.1; Osph1G61320.1; Osph1G66410.1; Osph1G72120.1; Osph1G72300.1; Osph1G81000.2; Osph1G81110.1; Osph2G57770.1; Osph3G59760.1; Osph3G74200.1; Osph3G77610.1; Osph4G28140.1; Osph4G38080.3; Osph4G39610.1; Osph5G30960.1; Osph5G30970.1; Osph5G32060.1; Osph6G12320.1; Osph6G12590.1; Osph6G29110.1; Osph6G29230.1; Osph8G17360.1; Osph9G11940.1; Osph9G12040.1; Osph9G17270.1; Osph9G42870.1; Osph9G42980.1; Osph9G43040.1; Osph9G47080.1; Osph10G15380.1; Osph11G23340.1; Osph11G41360.1; Osph12G15740.1; Osph12G15830.1; Osph13G12410.1; Osph13G12540.1; Osph13G13590.1; Osph13G13730.1; Osph13G13740.1; Osph13G15550.1; Osph14G44060.1; Osph14G47800.4; Osph14G47930.3; Osph15G19910.1; Osph15G19920.1; Osph15G23170.1; Osph16G20880.1; Osph20G12170.1; Osph27G11530.1; Osph29G11060.1; Osph29G11170.1; Osph35G11910.1; Osph35G12100.1; Osph36G10800.1; Osph41G11170.1; Osph41G11660.1; Osph42G10940.1; Osph42G11080.1; Osph44G12630.2; Osph50G10110.1; Osph56G12350.1; Osph60G10150.1; Osph60G10230.1; Osph67G11840.1; Osph67G11880.3; Osph68G11530.1; Osph71G11440.1; Osph74G12490.1; Osph74G12530.1; Osph76G11330.1; Osph76G11740.1; Osph76G12000.1; Osph76G12150.1; Osph77G10040.2; Osph78G11800.1; Osph93G11790.1; Osph97G10290.1; Osph105G11030.1; Osph105G11350.1; Osph110G10960.1; Osph135G10290.1; Osph135G10410.1; Osph145G10320.1; Osph145G10540.1; Osph168G10670.2; Osph170G10880.1; Osph205G10120.1; Osph218G10230.1; Osph222G10060.1; Osph222G10310.1; Osph225G10520.1; Osph269G10010.1; Osph269G10340.1; Osph274G10630.1; Osph286G10290.1; Osph288G10260.1; Osph314G10120.2; Osph315G10050.1; Osph322G10170.1; Osph329G10230.1; Osph356G10130.1; Osph366G10020.3; Osph366G10230.1; Osph367G10240.1; Osph437G10010.1; Osph505G10230.1; Osph553G10170.2; Osph557G10100.1; Osph578G10260.1; Osph698G10080.1; Osph698G10130.1; Osph751G10180.1; Osph791G10250.1; Osph804G10070.1; Osph856G10150.1; Osph981G10180.2
C3H	TF	60	Osph1G16160.1; Osph1G36220.1; Osph1G76900.1; Osph1G77270.2; Osph2G33430.1; Osph3G17070.2; Osph3G20900.1; Osph3G68340.1; Osph3G68510.1; Osph4G21830.1; Osph4G46600.1; Osph4G46670.1; Osph5G33360.1; Osph5G49480.1; Osph5G55240.1; Osph5G56030.1; Osph5G56210.1; Osph6G50440.1; Osph10G11610.1; Osph10G11840.1; Osph10G14140.1; Osph10G38430.1; Osph10G38440.1; Osph10G38670.1; Osph11G28840.1; Osph12G38190.1; Osph14G12790.2; Osph14G47360.1; Osph15G40090.1; Osph18G12470.1; Osph18G13170.1; Osph18G26720.3; Osph20G14950.2; Osph23G10130.1; Osph38G10580.2; Osph44G10930.1; Osph44G11150.1; Osph45G11980.3; Osph56G10370.1; Osph56G10810.1; Osph61G11550.1; Osph61G11800.1; Osph71G10490.1; Osph71G10940.1; Osph74G12640.1; Osph88G10540.1; Osph88G10650.1; Osph95G10290.1; Osph102G10270.1; Osph118G10230.3; Osph120G10170.1; Osph159G10530.2; Osph160G10390.1; Osph160G10600.1; Osph172G10080.1;

			Osph185G10150.3; Osph232G10130.1; Osph488G10210.1; Osph502G10150.1; Osph681G10190.1
CAMTA	TF	4	Osph6G44800.2; Osph26G11230.1; Osph83G10300.2; Osph177G10810.1
Coactivator p15	TR	3	Osph3G14200.1; Osph3G14330.1; Osph8G15890.1
CPP	TF	6	Osph1G79630.1; Osph1G79670.1; Osph7G47310.1; Osph195G10220.1; Osph195G10570.1; Osph306G10100.1
CSD	TF	21	Osph2G35090.1; Osph2G35110.1; Osph2G65620.1; Osph3G53010.1; Osph3G77790.1; Osph3G77810.1; Osph10G25840.1; Osph17G26210.1; Osph59G10520.1; Osph94G11160.1; Osph118G10020.1; Osph159G10850.1; Osph344G10240.1; Osph344G10250.1; Osph344G10280.1; Osph344G10360.1; Osph344G10410.1; Osph598G10070.1; Osph598G10120.1; Osph598G10190.1; Osph987G10070.1
DBB	TF	3	Osph3G26180.5; Osph11G19140.3; Osph762G10110.2
DBP	TF	3	Osph69G11660.1; Osph69G11830.1; Osph507G10150.3
DDT	TF	8	Osph3G57590.1; Osph4G22380.1; Osph14G46400.1; Osph17G19220.1; Osph57G10390.1; Osph103G11260.1; Osph153G10200.1; Osph164G10480.1
E2F-DP	TF	13	Osph2G27900.1; Osph2G27910.1; Osph2G28000.1; Osph14G14480.2; Osph17G22410.1; Osph51G10610.2; Osph51G11070.1; Osph64G10950.1; Osph64G11050.1; Osph66G10090.2; Osph66G10200.2; Osph167G10770.1; Osph167G10910.1
EIL	TF	6	Osph1G22530.1; Osph8G34260.1; Osph15G21370.1; Osph145G10030.1; Osph274G10080.1; Osph567G10170.1
FAR1	TF	102	Osph1G16960.1; Osph1G29790.1; Osph1G62230.1; Osph1G78420.1; Osph1G82170.1; Osph2G17640.1; Osph2G19160.1; Osph2G20780.1; Osph2G20930.1; Osph2G27470.1; Osph2G50060.1; Osph2G53220.1; Osph2G57480.1; Osph3G46340.1; Osph3G66910.1; Osph4G30930.1; Osph4G48050.1; Osph4G51420.1; Osph4G51480.1; Osph5G35490.1; Osph5G50600.1; Osph5G55720.1; Osph6G16680.1; Osph6G39310.1; Osph6G47700.1; Osph6G47860.1; Osph7G15540.1; Osph7G23680.1; Osph8G19450.1; Osph8G26440.1; Osph9G11840.1; Osph9G37130.1; Osph9G39150.1; Osph9G39490.1; Osph10G14910.1; Osph10G26840.1; Osph10G31420.1; Osph11G10970.1; Osph11G30440.1; Osph11G41430.1; Osph12G11710.1; Osph13G15590.1; Osph13G26400.1; Osph14G10670.1; Osph14G22250.1; Osph14G42410.1; Osph15G24940.1; Osph15G32740.1; Osph15G33170.1; Osph16G30510.1; Osph17G23880.1; Osph17G25650.1; Osph18G22630.1; Osph18G23690.1; Osph24G13810.1; Osph27G11790.1; Osph29G10080.1; Osph29G10650.1; Osph33G13010.1; Osph33G13340.1; Osph40G11970.1; Osph40G12180.1; Osph45G12380.1; Osph49G11240.1; Osph54G10540.1; Osph54G10900.1; Osph60G10800.1; Osph62G11090.1; Osph74G12500.1; Osph80G11580.1; Osph83G10670.1; Osph87G10310.1; Osph89G11980.4; Osph99G11350.1; Osph114G10080.1; Osph123G10620.1; Osph123G10920.1; Osph129G10400.1; Osph129G11640.1; Osph146G10720.1; Osph170G11050.1; Osph174G10270.1; Osph179G10560.1; Osph184G10540.1; Osph191G10710.2; Osph195G10500.1; Osph195G10800.1; Osph199G10900.1; Osph199G10960.1; Osph200G10610.1; Osph232G10240.1; Osph242G10420.2; Osph305G10160.1; Osph311G10350.2; Osph381G10080.1; Osph431G10110.1; Osph521G10180.1; Osph717G10010.1; Osph756G10030.1; Osph913G10050.1; Osph945G10070.1; Osph1186G10050.1
GARP-ARR-B	TF	7	Osph6G36670.2; Osph31G11110.2; Osph70G11640.1; Osph70G11680.1; Osph70G11920.1; Osph70G11930.1; Osph110G11020.1
GARP-G2-like	TF	40	Osph1G71980.1; Osph5G28090.1; Osph6G23660.1; Osph7G22640.1; Osph7G41180.1; Osph8G38550.1; Osph8G38820.4; Osph9G43480.1; Osph10G30710.1; Osph11G43420.1; Osph12G24520.1; Osph13G10130.1; Osph13G10230.1; Osph13G10240.1; Osph19G14180.1; Osph21G11980.1; Osph22G10510.1; Osph24G13640.1; Osph28G13180.2; Osph32G10770.1; Osph34G10740.4; Osph34G12840.3; Osph67G11030.1

			Osph68G11390.1; Osph68G11490.2; Osph68G11820.2; Osph87G10240.1; Osph91G10070.2; Osph106G11650.4; Osph118G11250.2; Osph129G10350.1; Osph129G10900.1; Osph129G11060.1; Osph129G11710.1; Osph174G10340.1; Osph247G10350.1; Osph364G10220.1; Osph515G10240.1; Osph519G10090.1; Osph1410G10010.4
GeBP	TF	11	Osph2G21470.1; Osph5G30020.1; Osph6G13540.1; Osph6G13570.1; Osph6G13960.1; Osph6G14010.1; Osph6G14020.1; Osph10G22770.1; Osph26G12300.1; Osph261G10290.1; Osph284G10090.1
GNAT	TR	40	Osph1G66630.1; Osph1G66810.1; Osph1G67170.1; Osph1G72220.2; Osph1G72340.1; Osph1G79400.1; Osph1G79520.1; Osph2G10400.1; Osph3G58870.1; Osph3G73640.1; Osph3G77630.1; Osph4G10320.1; Osph6G12190.1; Osph13G12950.2; Osph13G13260.1; Osph14G26690.1; Osph14G38290.1; Osph14G45140.1; Osph14G48890.2; Osph14G48950.1; Osph24G11320.1; Osph24G11650.2; Osph30G11960.1; Osph30G13240.1; Osph32G10280.2; Osph32G10290.3; Osph34G12010.1; Osph60G10870.1; Osph60G11010.1; Osph64G10600.1; Osph64G10730.1; Osph93G11070.1; Osph123G10180.1; Osph144G10450.1; Osph155G10260.1; Osph189G10570.1; Osph545G10100.1; Osph649G10030.2; Osph737G10090.1; Osph789G10030.1
GRAS	TF	67	Osph1G17650.1; Osph1G45950.1; Osph1G51680.1; Osph1G64740.1; Osph1G64970.1; Osph2G39320.1; Osph3G12760.1; Osph3G54480.1; Osph4G10010.1; Osph4G23090.1; Osph7G20240.1; Osph9G13030.1; Osph9G43730.1; Osph11G12220.1; Osph11G12660.1; Osph12G38560.1; Osph13G13030.1; Osph13G13310.1; Osph14G47610.1; Osph20G10260.1; Osph26G11880.1; Osph29G12120.1; Osph29G12310.1; Osph33G10090.1; Osph35G10600.1; Osph35G10970.1; Osph35G10980.1; Osph38G11220.1; Osph38G12110.1; Osph38G12140.1; Osph38G12180.1; Osph38G12800.1; Osph38G12830.1; Osph38G12840.1; Osph38G12850.1; Osph44G10490.1; Osph44G10640.1; Osph50G10320.1; Osph78G11530.1; Osph78G11900.1; Osph83G11490.1; Osph91G10580.1; Osph91G10850.1; Osph97G11780.1; Osph113G10160.1; Osph113G10600.1; Osph122G10110.1; Osph123G10170.1; Osph129G11020.1; Osph132G10230.1; Osph140G11220.5; Osph154G10290.1; Osph250G10410.1; Osph264G10320.1; Osph322G10350.1; Osph390G10030.1; Osph390G10040.1; Osph390G10060.1; Osph390G10110.1; Osph390G10150.1; Osph390G10240.1; Osph390G10250.1; Osph664G10120.1; Osph743G10040.1; Osph823G10070.1; Osph828G10160.1; Osph1282G10020.1
GRF	TF	12	Osph1G73430.2; Osph36G10470.1; Osph50G11410.1; Osph50G11600.2; Osph76G11600.1; Osph76G12130.1; Osph136G10390.1; Osph199G10420.1; Osph372G10200.1; Osph396G10080.3; Osph396G10230.3; Osph691G10200.1
HB-BELL	TF	14	Osph1G59060.1; Osph1G59120.2; Osph2G53410.1; Osph3G81190.1; Osph4G39250.3; Osph7G44750.1; Osph16G29780.1; Osph16G29840.1; Osph33G12000.2; Osph33G12270.1; Osph102G10460.1; Osph297G10090.1; Osph361G10070.1; Osph567G10240.1
HB-HD-ZIP	TF	41	Osph1G73780.2; Osph1G74810.2; Osph1G76890.1; Osph1G77250.1; Osph3G25260.1; Osph3G78500.1; Osph4G13400.1; Osph12G19480.1; Osph12G19530.1; Osph14G40780.1; Osph14G40820.2; Osph14G40970.2; Osph26G13320.2; Osph34G10080.1; Osph37G10590.1; Osph38G10220.1; Osph47G10120.1; Osph47G10600.2; Osph53G10250.1; Osph53G10380.2; Osph55G11600.1; Osph59G12020.1; Osph66G11680.1; Osph70G10500.1; Osph70G10510.3; Osph70G10520.1; Osph88G11050.1; Osph93G11610.1; Osph121G10880.3; Osph122G10970.1; Osph157G10540.2; Osph281G10020.2; Osph284G10470.2; Osph344G10080.1; Osph418G10180.1; Osph418G10290.1; Osph633G10130.3; Osph633G10230.2; Osph818G10090.1; Osph1045G10050.1; Osph1119G10050.1

HB-KNOX	TF	9	Osph1G73740.1; Osph11G15030.1; Osph11G15320.1; Osph45G12420.2; Osph139G10330.4; Osph214G10120.1; Osph399G10190.1; Osph799G10090.1; Osph998G10040.1
HB-other	TF	11	Osph4G25880.1; Osph10G10130.1; Osph12G26110.1; Osph12G27000.1; Osph14G48030.1; Osph14G48200.1; Osph18G12450.1; Osph46G11340.1; Osph88G11450.1; Osph194G10490.1; Osph388G10030.1
HB-WOX	TF	9	Osph1G25160.2; Osph3G74350.1; Osph4G34850.2; Osph9G44260.1; Osph9G44390.1; Osph10G41340.1; Osph13G38260.1; Osph16G35150.1; Osph277G10140.1
HMG	TR	12	Osph1G57530.1; Osph3G78880.1; Osph9G36940.1; Osph41G12000.1; Osph41G12420.1; Osph50G10650.1; Osph50G10660.1; Osph116G11330.2; Osph116G11340.1; Osph692G10170.1; Osph804G10040.1; Osph1164G10050.1
HSF	TF	24	Osph1G11720.1; Osph1G11770.2; Osph3G78530.1; Osph4G22420.1; Osph7G34420.1; Osph13G20450.1; Osph13G20460.1; Osph15G34780.1; Osph22G12500.1; Osph42G11210.1; Osph43G12120.1; Osph47G11480.2; Osph52G11180.1; Osph52G11470.2; Osph80G10860.1; Osph161G10980.1; Osph198G10030.1; Osph211G10040.1; Osph211G10170.1; Osph446G10070.1; Osph463G10050.1; Osph728G10100.1; Osph926G10090.1; Osph1034G10040.1
IWS1	TR	12	Osph2G29110.2; Osph9G47500.1; Osph13G34540.2; Osph16G27840.1; Osph18G20970.1; Osph78G11460.1; Osph107G10140.1; Osph124G10330.1; Osph138G10170.1; Osph192G10630.1; Osph225G10120.1; Osph973G10090.2
Jumonji	TR	21	Osph3G74160.2; Osph6G52380.2; Osph10G15400.1; Osph26G11480.1; Osph33G11320.1; Osph33G11370.2; Osph33G12150.1; Osph41G11410.1; Osph54G11840.1; Osph54G12010.1; Osph54G12250.1; Osph54G12350.1; Osph54G12610.1; Osph54G12680.1; Osph88G10870.1; Osph88G11270.1; Osph96G10760.1; Osph101G11200.1; Osph364G10230.1; Osph515G10210.3; Osph581G10250.2
LIM	TF	6	Osph27G11560.2; Osph65G10820.2; Osph77G10240.1; Osph207G10140.3; Osph207G10300.3; Osph227G10490.1
LOB	TF	27	Osph1G37540.1; Osph1G79090.1; Osph2G34700.1; Osph2G69160.1; Osph2G69380.1; Osph2G70230.1; Osph2G70460.1; Osph3G11670.1; Osph8G43570.1; Osph8G43690.1; Osph11G21680.1; Osph13G12980.1; Osph13G13280.1; Osph16G36680.1; Osph23G11170.2; Osph56G12260.1; Osph56G12330.2; Osph61G10800.2; Osph61G11170.1; Osph113G10070.1; Osph113G10670.1; Osph114G10320.1; Osph286G10420.1; Osph471G10190.1; Osph530G10030.1; Osph756G10060.1; Osph945G10050.3
LUG	TR	3	Osph137G10860.1; Osph506G10320.1; Osph748G10160.3
MADS-MIKC	TF	30	Osph3G15430.3; Osph3G26320.1; Osph3G77030.1; Osph3G80460.5; Osph3G80660.1; Osph4G51190.1; Osph9G21470.1; Osph12G10230.1; Osph12G10360.1; Osph17G18960.1; Osph21G10020.1; Osph22G13690.5; Osph39G10250.1; Osph39G10410.2; Osph43G12090.1; Osph47G11250.2; Osph58G11490.2; Osph79G11230.1; Osph97G11060.1; Osph97G11190.1; Osph99G10780.1; Osph108G10730.1; Osph108G10960.1; Osph170G10490.1; Osph379G10130.1; Osph379G10140.1; Osph401G10200.1; Osph401G10210.1; Osph472G10030.1; Osph1456G10010.2
MADS-M-type	TF	88	Osph1G82860.2; Osph1G82970.1; Osph2G22970.1; Osph2G22990.1; Osph2G23000.1; Osph2G23030.1; Osph2G67370.1; Osph2G69460.1; Osph3G73380.1; Osph4G51340.1; Osph4G51350.1; Osph4G51410.1; Osph4G51520.1; Osph4G51610.1; Osph4G51650.1; Osph5G44370.1; Osph5G44390.1; Osph5G44560.1; Osph5G44580.1; Osph5G44620.1; Osph5G44700.1; Osph5G45550.1; Osph5G45940.1; Osph5G46020.1; Osph5G46050.1; Osph5G46720.1; Osph5G46850.1; Osph5G55860.1; Osph6G12180.1; Osph6G24380.1; Osph6G26170.1; Osph6G45680.1; Osph8G34720.1; Osph8G34860.1; Osph8G34950.1; Osph9G39650.1; Osph9G45670.1; Osph12G10440.1; Osph12G10460.1; Osph12G10710.1; Osph12G10740.1; Osph12G10800.1; Osph12G10860.1; Osph12G10890.1; Osph12G11160.1;

			Osph12G11180.1; Osph12G11190.1; Osph12G11200.1; Osph12G11240.1; Osph12G11270.1; Osph12G11360.1; Osph12G11370.1; Osph12G11470.1; Osph12G11750.1; Osph12G22790.1; Osph19G17210.1; Osph19G17430.1; Osph37G10050.1; Osph37G10110.1; Osph40G11200.1; Osph51G12120.1; Osph51G12140.1; Osph58G10700.1; Osph58G10850.1; Osph58G10870.1; Osph79G11310.1; Osph81G10780.1; Osph81G10800.1; Osph99G10630.1; Osph100G10760.1; Osph135G10870.1; Osph135G10890.1; Osph135G10920.1; Osph135G10970.1; Osph138G10500.1; Osph138G10510.1; Osph170G10320.1; Osph261G10230.1; Osph284G10170.1; Osph369G10400.1; Osph715G10160.1; Osph809G10150.1; Osph809G10160.1; Osph942G10050.1; Osph1285G10010.1; Osph1285G10030.1; Osph1285G10040.1; Osph1333G10030.1
MBF1	TR	3	Osph43G11610.1; Osph89G11040.1; Osph89G11360.1
MED7	TR	2	Osph736G10020.2; Osph935G10170.1
mTERF	TR	79	Osph1G24660.1; Osph1G74050.2; Osph1G75070.3; Osph6G35210.1; Osph6G45620.1; Osph9G12060.2; Osph12G29350.1; Osph12G31840.1; Osph13G22100.1; Osph14G33960.1; Osph14G35230.1; Osph14G36450.1; Osph14G37360.1; Osph14G37440.1; Osph14G37520.1; Osph14G37740.1; Osph14G37770.1; Osph14G38030.1; Osph14G38240.1; Osph14G38480.1; Osph14G38510.1; Osph14G38560.1; Osph14G38600.1; Osph14G38620.1; Osph14G38630.1; Osph14G38740.1; Osph14G38790.1; Osph14G38830.1; Osph14G38850.1; Osph14G38860.1; Osph14G38880.1; Osph14G38910.1; Osph14G38980.1; Osph14G39040.1; Osph14G39090.1; Osph14G39100.1; Osph14G39170.1; Osph14G39180.1; Osph14G39190.1; Osph14G39450.1; Osph14G39480.1; Osph14G39530.1; Osph14G39560.1; Osph14G40470.1; Osph14G41190.1; Osph14G41370.1; Osph14G41380.1; Osph14G41530.1; Osph14G41660.1; Osph14G41670.1; Osph14G43840.1; Osph14G43930.1; Osph14G45850.1; Osph14G46050.1; Osph14G46390.1; Osph15G33020.1; Osph22G10150.1; Osph22G10190.1; Osph22G10210.1; Osph22G10230.1; Osph22G10280.1; Osph32G12540.1; Osph37G11230.1; Osph37G11360.3; Osph45G12600.1; Osph66G11010.1; Osph88G11570.1; Osph88G11940.1; Osph107G10330.1; Osph249G10480.1; Osph366G10260.1; Osph367G10200.1; Osph393G10040.1; Osph545G10250.1; Osph656G10060.1; Osph656G10160.1; Osph672G10080.1; Osph827G10010.1; Osph1092G10050.1
MYB	TF	91	Osph1G11860.1; Osph1G25400.3; Osph1G74830.2; Osph2G15800.1; Osph2G15900.1; Osph2G19980.1; Osph2G34540.2; Osph3G14100.1; Osph3G20350.1; Osph3G46480.1; Osph3G77760.1; Osph3G79780.2; Osph3G80130.2; Osph6G10250.1; Osph6G23340.2; Osph6G23440.3; Osph6G38240.1; Osph7G43490.2; Osph7G43650.2; Osph8G37330.2; Osph8G37370.1; Osph8G43800.2; Osph8G43930.2; Osph9G22360.3; Osph10G38000.4; Osph11G12350.2; Osph12G32140.1; Osph15G19140.1; Osph17G23260.2; Osph17G24740.1; Osph18G28200.1; Osph20G11080.1; Osph20G11380.1; Osph27G12590.1; Osph36G10500.2; Osph37G10690.1; Osph50G12700.2; Osph51G10420.1; Osph51G10870.1; Osph55G10100.2; Osph55G10560.1; Osph59G10720.1; Osph61G10070.1; Osph61G10340.1; Osph63G11990.1; Osph65G10660.1; Osph67G10300.1; Osph69G10380.2; Osph70G11990.1; Osph74G10260.2; Osph75G11070.1; Osph77G10260.1; Osph78G11380.1; Osph78G11390.1; Osph80G10890.1; Osph81G10200.1; Osph81G10480.1; Osph81G10890.2; Osph82G10440.1; Osph84G11130.1; Osph91G11120.1; Osph102G10280.2; Osph102G10740.1; Osph105G11040.1; Osph105G11360.1; Osph108G10260.1; Osph137G10140.1; Osph149G10030.2; Osph149G10120.1; Osph150G10630.1; Osph168G10600.1; Osph176G10250.1; Osph186G10110.1; Osph189G10090.1; Osph189G10420.2; Osph214G10030.1; Osph215G10540.1; Osph217G10100.1;

			Osph232G10020.1; Osph237G10670.1; Osph266G10150.1; Osph266G10240.2; Osph306G10190.1; Osph457G10060.2; Osph528G10030.2; Osph783G10020.1; Osph809G10060.1; Osph910G10090.1; Osph926G10060.1; Osph943G10120.1; Osph1053G10080.1
MYB-related	TF	77	Osph1G11660.1; Osph1G68620.1; Osph2G24470.1; Osph2G55900.1; Osph2G57780.1; Osph3G20270.1; Osph3G44580.1; Osph3G50170.1; Osph3G77090.1; Osph4G28490.1; Osph4G34140.1; Osph4G57960.1; Osph6G10140.2; Osph6G10180.1; Osph7G14360.2; Osph7G47230.2; Osph8G15130.2; Osph8G16510.1; Osph8G33330.1; Osph9G19660.1; Osph9G22390.1; Osph10G25000.2; Osph10G29320.1; Osph10G37020.1; Osph11G18160.1; Osph11G34380.1; Osph12G11970.1; Osph12G12050.2; Osph12G35660.5; Osph13G13140.1; Osph16G25200.1; Osph16G29930.1; Osph17G17710.2; Osph20G10070.1; Osph23G11830.1; Osph24G12340.2; Osph24G12520.1; Osph36G10610.2; Osph39G11580.1; Osph40G11070.1; Osph51G10020.1; Osph57G12190.3; Osph57G12210.1; Osph57G12410.1; Osph61G10320.1; Osph62G10520.1; Osph62G11580.1; Osph73G10050.1; Osph74G10420.1; Osph74G12580.1; Osph76G11340.1; Osph82G10520.1; Osph82G10550.1; Osph82G10560.1; Osph82G11060.1; Osph88G11620.1; Osph88G11900.2; Osph91G10160.1; Osph124G10310.1; Osph145G10140.1; Osph150G10770.1; Osph159G10320.3; Osph165G10320.1; Osph175G10600.1; Osph203G10020.1; Osph237G10230.1; Osph240G10050.1; Osph254G10130.1; Osph293G10460.1; Osph319G10140.1; Osph383G10210.1; Osph394G10060.4; Osph480G10110.1; Osph572G10160.1; Osph588G10140.3; Osph615G10110.1; Osph887G10010.1
NAC	TF	80	Osph1G13330.1; Osph1G34040.1; Osph1G42160.1; Osph1G68920.1; Osph1G74380.1; Osph1G75410.1; Osph2G63430.1; Osph2G63470.2; Osph2G63500.1; Osph2G63530.1; Osph2G65860.1; Osph4G31700.1; Osph4G56370.1; Osph6G23090.2; Osph6G23150.1; Osph6G24170.1; Osph7G44640.1; Osph8G16320.2; Osph8G22710.1; Osph9G23010.1; Osph10G13440.1; Osph10G13730.1; Osph12G25720.1; Osph12G29930.2; Osph13G31150.1; Osph14G20900.1; Osph14G20980.1; Osph15G39880.1; Osph17G19460.2; Osph18G10010.1; Osph20G12320.1; Osph20G12510.5; Osph20G13380.4; Osph20G14220.1; Osph20G14520.2; Osph20G14540.2; Osph22G10390.1; Osph34G12090.1; Osph34G12120.2; Osph34G12130.1; Osph34G13000.1; Osph41G10990.1; Osph41G11530.1; Osph45G10040.1; Osph45G10260.1; Osph55G11140.1; Osph59G10490.1; Osph59G10500.1; Osph59G10580.1; Osph59G10610.3; Osph81G10570.1; Osph81G10730.1; Osph91G10730.1; Osph91G11030.1; Osph97G10540.1; Osph97G10560.1; Osph97G10590.2; Osph97G10730.1; Osph97G10750.2; Osph97G11630.2; Osph102G10520.2; Osph105G10150.1; Osph105G10270.1; Osph106G10950.1; Osph137G10360.1; Osph146G10030.1; Osph194G10180.1; Osph200G10340.2; Osph200G10530.2; Osph225G10470.4; Osph268G10220.3; Osph275G10070.2; Osph314G10380.2; Osph356G10180.1; Osph359G10010.1; Osph550G10170.2; Osph572G10010.1; Osph756G10080.2; Osph899G10030.3; Osph945G10040.4
NF-X1	TF	5	Osph11G30260.2; Osph13G11840.1; Osph13G11920.1; Osph151G10510.1; Osph151G10810.1
NF-YA	TF	5	Osph4G16720.1; Osph14G44220.1; Osph172G10230.5; Osph681G10050.1; Osph1007G10060.1
NF-YB	TF	8	Osph4G12480.1; Osph62G12110.1; Osph99G10570.1; Osph99G10710.1; Osph145G10270.1; Osph145G10520.1; Osph162G10510.1; Osph513G10310.1
NF-YC	TF	12	Osph2G38400.1; Osph10G18790.1; Osph13G26010.1; Osph18G28590.1; Osph35G10170.2; Osph39G12530.1; Osph83G12120.1; Osph92G11160.1; Osph92G11710.1; Osph106G11510.1; Osph154G11210.1; Osph247G10190.1
OFP	TF	23	Osph7G22670.1; Osph9G32300.1; Osph10G24060.1; Osph10G24090.1; Osph11G14480.1; Osph13G12340.1;

			Osph13G12650.1; Osph13G16310.1; Osph13G16350.1; Osph14G47750.1; Osph14G47890.1; Osph32G11790.1; Osph44G11380.1; Osph79G10800.1; Osph79G10990.1; Osph102G10620.1; Osph133G11190.1; Osph145G10090.1; Osph183G10590.1; Osph183G10760.1; Osph262G10070.1; Osph262G10100.1; Osph567G10100.1
Others	TR	57	Osph1G17990.1; Osph1G18630.1; Osph2G35660.1; Osph3G33150.1; Osph3G70060.1; Osph3G71140.1; Osph6G14700.1; Osph6G14730.1; Osph6G17350.1; Osph6G24970.2; Osph9G29990.2; Osph9G41140.1; Osph11G10960.1; Osph13G12280.2; Osph13G12690.2; Osph13G29460.1; Osph14G44390.1; Osph16G34880.1; Osph19G24530.1; Osph28G12550.1; Osph30G12810.2; Osph30G12920.1; Osph54G12400.1; Osph61G10200.1; Osph61G10500.1; Osph67G11100.1; Osph70G10470.1; Osph80G11170.2; Osph88G10160.1; Osph92G11400.2; Osph98G10520.1; Osph98G10740.1; Osph103G11420.1; Osph110G10050.1; Osph127G10050.1; Osph127G10520.3; Osph137G10180.1; Osph142G11210.1; Osph148G10610.1; Osph148G10830.1; Osph168G10920.1; Osph168G10930.1; Osph254G10250.1; Osph254G10390.4; Osph293G10260.1; Osph293G10550.1; Osph304G10110.1; Osph304G10260.1; Osph334G10130.2; Osph334G10220.2; Osph359G10320.1; Osph511G10010.2; Osph511G10240.1; Osph592G10020.1; Osph693G10080.1; Osph887G10080.1; Osph887G10350.1
PHD	TR	37	Osph1G65400.1; Osph1G66040.2; Osph3G52560.2; Osph4G58500.1; Osph5G37750.2; Osph7G16550.2; Osph11G12470.2; Osph11G12960.2; Osph14G47040.2; Osph16G23020.8; Osph18G11270.1; Osph24G11970.2; Osph33G12030.2; Osph33G12310.2; Osph41G10770.1; Osph41G11010.1; Osph41G11560.1; Osph45G12210.2; Osph45G12700.1; Osph47G12490.2; Osph56G11590.1; Osph61G11330.1; Osph61G11600.1; Osph70G10460.3; Osph93G10940.1; Osph94G11720.1; Osph95G10320.1; Osph103G11250.1; Osph117G10080.1; Osph168G10110.1; Osph218G10360.1; Osph220G10220.1; Osph239G10210.1; Osph239G10450.1; Osph511G10100.1; Osph511G10170.1; Osph653G10010.1
PLATZ	TF	13	Osph1G70810.3; Osph8G29380.2; Osph8G42610.1; Osph8G42770.1; Osph12G11840.1; Osph20G11520.1; Osph28G10690.1; Osph28G10900.1; Osph68G11950.1; Osph68G11970.1; Osph68G11980.1; Osph218G10530.1; Osph227G10600.2
Pseudo ARR-B	TR	4	Osph88G11720.3; Osph167G10190.1; Osph167G10350.2; Osph539G10170.1
Rcd1-like	TR	6	Osph42G10420.3; Osph42G10450.1; Osph88G11000.1; Osph88G11170.1; Osph281G10220.2; Osph527G10250.1
RWP-RK	TF	9	Osph5G50720.1; Osph5G50990.1; Osph5G51140.4; Osph5G51160.1; Osph9G29610.1; Osph14G29790.1; Osph16G37830.1; Osph114G10140.1; Osph162G11110.1
SAP	TF	3	Osph4G44380.1; Osph29G12790.1; Osph75G10510.1
SBP	TF	16	Osph1G53370.1; Osph1G69480.1; Osph1G69650.1; Osph2G37750.2; Osph2G70700.2; Osph2G70880.1; Osph3G61770.2; Osph5G40570.1; Osph7G11670.1; Osph7G11740.1; Osph17G31650.1; Osph28G12090.1; Osph101G10070.2; Osph190G10130.2; Osph249G10260.1; Osph715G10030.3
SET	TR	38	Osph2G34080.2; Osph2G47890.3; Osph3G54090.1; Osph4G20190.1; Osph5G32200.1; Osph7G21190.1; Osph8G44960.1; Osph9G24810.1; Osph11G35040.1; Osph13G13500.1; Osph15G12750.1; Osph15G12820.2; Osph17G14120.1; Osph18G15690.1; Osph21G12460.3; Osph26G12700.4; Osph26G13450.1; Osph39G10090.1; Osph44G11860.1; Osph51G12460.2; Osph59G10190.1; Osph59G10390.1; Osph61G10640.1; Osph63G10860.1; Osph68G11750.1; Osph77G10200.1; Osph88G10040.3; Osph88G10050.1; Osph88G10180.1; Osph110G10580.1; Osph110G10800.1; Osph124G10810.1; Osph133G10090.1; Osph139G10360.1; Osph217G10010.1; Osph217G10110.1; Osph297G10030.1; Osph1087G10010.1
SNF2	TR	45	Osph1G18330.1; Osph1G53810.1; Osph1G69530.1; Osph1G82850.1; Osph2G12690.1; Osph2G37180.1; Osph3G63550.4; Osph3G78370.1;

			Osph4G26060.1; Osph4G26100.1; Osph4G49510.1; Osph4G49660.1; Osph5G22910.1; Osph5G55870.1; Osph5G57360.1; Osph5G57700.1; Osph6G25350.1; Osph6G25390.2; Osph8G43230.1; Osph8G43360.1; Osph9G14120.1; Osph10G38800.1; Osph12G16890.1; Osph12G16950.2; Osph16G18530.1; Osph16G18590.1; Osph16G25350.1; Osph67G11850.1; Osph71G10460.2; Osph71G10840.1; Osph82G10350.1; Osph82G11080.1; Osph118G10740.1; Osph118G11050.1; Osph127G10340.1; Osph128G10040.1; Osph141G10700.1; Osph170G10100.1; Osph170G10820.1; Osph196G10330.1; Osph215G10610.1; Osph222G10030.1; Osph606G10040.1; Osph632G10120.1; Osph702G10010.1
SRS	TF	5	Osph2G23510.1; Osph10G35770.1; Osph80G11100.1; Osph172G10010.1; Osph1061G10050.1
SWI/SNF-BAF60b	TR	16	Osph1G65630.1; Osph1G65750.1; Osph1G71920.1; Osph3G48830.1; Osph8G20360.1; Osph16G14950.1; Osph16G15230.2; Osph17G15960.1; Osph18G13430.1; Osph57G10110.1; Osph60G11430.1; Osph60G11820.1; Osph97G11240.2; Osph97G11350.1; Osph164G10070.1; Osph164G10250.1
SWI/SNF-SWI3	TR	3	Osph92G10130.1; Osph92G10910.2; Osph539G10090.2
TAZ	TR	4	Osph11G35980.1; Osph15G35920.4; Osph128G10170.1; Osph341G10140.1
TCP	TF	26	Osph3G12020.1; Osph3G75590.1; Osph3G76020.1; Osph6G52620.1; Osph13G31900.1; Osph15G13700.1; Osph17G28030.1; Osph17G28120.1; Osph20G12690.1; Osph20G12850.1; Osph20G12990.1; Osph26G12840.1; Osph38G11160.1; Osph59G11670.1; Osph59G11780.1; Osph100G10540.1; Osph100G10560.1; Osph123G10050.1; Osph261G10190.1; Osph273G10090.1; Osph284G10200.1; Osph300G10210.1; Osph359G10260.1; Osph512G10330.1; Osph671G10210.1; Osph828G10070.1
Tify	TF	13	Osph1G77540.1; Osph30G10580.1; Osph30G11150.1; Osph33G10410.1; Osph33G10620.1; Osph65G10440.3; Osph81G10040.3; Osph81G11650.1; Osph87G10230.1; Osph147G11120.1; Osph181G10680.3; Osph191G10060.1; Osph191G10340.1
TRAF	TR	21	Osph4G23850.1; Osph4G24000.1; Osph5G36020.1; Osph5G51600.1; Osph7G21940.1; Osph7G21980.1; Osph7G34350.1; Osph7G34410.1; Osph9G48080.1; Osph13G19150.1; Osph57G10370.1; Osph58G10010.1; Osph58G10130.1; Osph87G11000.1; Osph187G10720.1; Osph209G10540.1; Osph209G10860.1; Osph217G10270.1; Osph222G10290.1; Osph222G10500.1; Osph686G10190.1
Trihelix	TF	25	Osph2G17040.1; Osph2G17600.1; Osph3G76980.1; Osph10G26560.1; Osph10G26670.1; Osph11G26990.1; Osph11G36600.1; Osph14G36410.1; Osph16G23380.2; Osph16G31080.1; Osph16G31250.1; Osph16G34650.1; Osph16G34740.1; Osph17G23050.1; Osph18G17100.2; Osph43G11600.1; Osph47G11860.1; Osph50G11740.1; Osph50G12180.1; Osph55G11520.1; Osph118G10440.1; Osph118G11330.1; Osph124G11570.1; Osph177G10460.1; Osph322G10140.1
TUB	TF	4	Osph7G24560.3; Osph43G11620.1; Osph580G10010.1; Osph861G10090.1
ULT	TF	3	Osph71G10070.3; Osph71G10350.2; Osph147G11080.1
Whirly	TF	3	Osph5G36760.1; Osph5G37360.3; Osph47G10810.1
WRKY	TF	59	Osph1G10730.1; Osph1G80950.1; Osph2G47630.1; Osph2G69120.1; Osph2G69420.1; Osph4G19440.1; Osph4G37050.1; Osph6G36750.3; Osph7G15630.1; Osph7G46840.1; Osph8G39640.1; Osph8G39740.1; Osph9G30090.2; Osph9G35030.2; Osph10G30440.3; Osph11G10560.1; Osph11G10610.1; Osph11G10730.1; Osph11G10780.1; Osph11G10790.1; Osph11G12230.1; Osph11G12650.1; Osph11G14100.1; Osph11G14230.1; Osph11G33780.1; Osph11G38500.1; Osph16G36620.2; Osph18G16150.1; Osph19G23000.1; Osph19G23030.1; Osph24G12620.1; Osph24G12800.1; Osph43G12570.1; Osph54G10930.1; Osph54G10950.1; Osph54G11170.1; Osph54G11200.1; Osph68G11450.2; Osph69G11260.1

			Osph69G11800.1; Osph70G10630.1; Osph82G11350.2; Osph89G10440.1; Osph90G11100.1; Osph90G11340.1; Osph90G11530.1; Osph90G11760.1; Osph110G10480.1; Osph110G10710.1; Osph110G11230.1; Osph139G10820.2; Osph139G11290.1; Osph142G10850.1; Osph514G10220.1; Osph592G10070.1; Osph632G10020.1; Osph665G10070.1; Osph693G10110.1; Osph721G10050.1
zf-HD	TF	11	Osph9G23380.1; Osph9G23400.1; Osph9G39880.2; Osph9G39890.1; Osph12G18580.1; Osph13G21040.1; Osph15G21530.1; Osph59G11090.1; Osph59G11420.1; Osph96G10820.1; Osph211G10710.1

528
529

Supplementary Table 10. List of protein kinases identified

Protein Kinase	N	Genes
AGC_MAST	11	Osph1G26870.1; Osph1G30610.1; Osph1G39780.1; Osph2G68540.1; Osph5G26410.4; Osph5G29330.1; Osph11G11940.1; Osph11G12090.1; Osph18G30460.1; Osph19G13150.1; Osph147G10920.1
AGC_NDR	10	Osph51G11950.2; Osph51G11960.1; Osph67G10540.2; Osph67G10550.2; Osph93G11000.1; Osph99G11860.2; Osph137G10990.3; Osph139G10940.1; Osph139G11090.3; Osph982G10010.1
AGC_PDK1	3	Osph12G33100.1; Osph105G10930.1; Osph105G11260.1
AGC_RSK-2	23	Osph1G14070.1; Osph1G69850.1; Osph1G70830.1; Osph1G77100.1; Osph1G77420.1; Osph3G43700.1; Osph5G31610.1; Osph14G25870.2; Osph32G11480.4; Osph32G11590.1; Osph61G11260.1; Osph66G11780.1; Osph88G11560.1; Osph88G11950.2; Osph139G10710.2; Osph195G10290.1; Osph195G10600.1; Osph227G10620.1; Osph238G10300.1; Osph249G10490.1; Osph270G10190.1; Osph393G10010.1; Osph393G10020.1
AGC-PI	2	Osph6G25300.1; Osph6G25480.1
Aur	3	Osph15G31070.1; Osph70G12110.1; Osph70G12120.2
BUB	2	Osph82G10750.1; Osph82G10840.1
CAMK_CAMKL-CHK1	22	Osph1G21800.1; Osph3G49250.4; Osph4G13230.4; Osph4G34800.1; Osph4G36720.1; Osph6G37390.1; Osph7G22030.1; Osph7G24540.2; Osph9G32910.1; Osph10G27390.1; Osph28G12880.2; Osph28G13150.2; Osph48G10650.1; Osph73G10160.2; Osph160G10900.1; Osph572G10240.3; Osph572G10250.1; Osph590G10110.2; Osph615G10030.2; Osph615G10040.1; Osph692G10040.2; Osph804G10150.2
CAMK_CAMKL-LKB	3	Osph17G11530.1; Osph31G11890.1; Osph31G12160.1
CAMK_CDPK	78	Osph1G70130.1; Osph1G78260.1; Osph1G78360.1; Osph1G78570.1; Osph2G55140.1; Osph2G55190.1; Osph2G55230.1; Osph3G34790.1; Osph3G79080.1; Osph3G80740.1; Osph4G17640.3; Osph5G23820.1; Osph5G41570.2; Osph5G50800.1; Osph5G51050.1; Osph5G51060.1; Osph5G51460.2; Osph7G45460.1; Osph7G45570.4; Osph8G23180.1; Osph8G23250.1; Osph9G33130.1; Osph9G38030.1; Osph10G39660.1; Osph13G10610.1; Osph14G47280.3; Osph16G36480.1; Osph19G10520.1; Osph19G10540.1; Osph19G10550.1; Osph19G10730.1; Osph19G10740.1; Osph19G10750.1; Osph19G10760.1; Osph19G10780.1; Osph19G10790.1; Osph19G10840.1; Osph19G10980.1; Osph19G11000.1; Osph19G11010.1; Osph19G11060.1; Osph28G10360.4; Osph28G10530.1; Osph38G12520.1; Osph42G11490.1; Osph55G12340.1; Osph61G10090.1; Osph61G10380.1; Osph67G10270.1; Osph73G11600.2; Osph73G11770.3; Osph74G10720.1; Osph74G10970.2; Osph77G10360.1; Osph88G11710.1; Osph97G11490.1; Osph104G10100.2; Osph105G10110.1; Osph105G10250.1; Osph151G10450.1; Osph151G10830.1; Osph165G10560.2; Osph165G10570.1; Osph165G10660.1; Osph170G10590.1; Osph172G10180.1; Osph196G10010.1; Osph597G10020.1; Osph681G10160.1; Osph1014G10010.1; Osph1014G10040.1; Osph1014G10050.1; Osph1014G10060.1; Osph1014G10070.1; Osph1014G10080.1; Osph1014G10090.1; Osph1014G10100.1; Osph1014G10110.1
CAMK_OST1L	12	Osph1G71890.4; Osph4G56440.3; Osph4G56450.1; Osph5G44360.1; Osph22G14460.1; Osph53G10930.1; Osph53G11150.1; Osph66G11980.1; Osph99G10180.1; Osph169G10060.1; Osph263G10050.1; Osph1379G10010.1
CK1_CK1	7	Osph63G10970.3; Osph137G11000.3; Osph139G11350.1; Osph139G11530.1; Osph169G11100.1; Osph314G10200.7; Osph553G10080.1
CK1_CK1-PI	4	Osph16G13890.1; Osph64G10190.2; Osph170G10070.1; Osph170G10850.1
CMGC_CDK-CCRK	2	Osph95G11570.1; Osph95G11660.1
CMGC_CDK-CDK7	2	Osph1G71340.1; Osph1G71490.1

CMGC_CDK-CRK7-CDK9	10	Osph1G69990.2; Osph3G51910.1; Osph4G60110.1; Osph13G12710.2; Osph18G25970.2; Osph39G10550.1; Osph39G11080.2; Osph47G11190.1; Osph218G10390.1; Osph509G10190.1
CMGC_CDK-PITSLRE	3	Osph14G35060.1; Osph16G34360.1; Osph16G34550.2
CMGC_CDK-PI	9	Osph1G21630.1; Osph1G69100.1; Osph1G69260.1; Osph2G37930.1; Osph13G37860.2; Osph42G11470.1; Osph74G12660.1; Osph257G10020.2; Osph924G10180.1
CMGC_CK2	4	Osph1G72350.1; Osph1G72500.1; Osph23G11400.1; Osph286G10330.1
CMGC_CLK	4	Osph3G80210.6; Osph9G46840.2; Osph9G46950.1; Osph12G38620.1
CMGC_DYRK-PRP4	4	Osph9G38510.1; Osph147G11180.1; Osph218G10190.1; Osph1045G10140.1
CMGC_GSK	9	Osph1G11690.1; Osph1G11790.2; Osph2G18070.3; Osph6G33050.1; Osph42G11280.2; Osph42G11310.1; Osph47G11560.1; Osph47G11700.2; Osph232G10150.2
CMGC_GSKL	2	Osph201G10100.1; Osph201G10180.1
CMGC_MAPK	14	Osph1G13280.1; Osph1G16520.1; Osph4G58680.1; Osph7G25530.2; Osph7G25820.1; Osph14G40870.1; Osph14G41000.1; Osph49G11340.1; Osph74G10320.1; Osph74G10360.2; Osph91G10090.1; Osph101G11460.1; Osph130G10530.1; Osph189G10870.7
CMGC_PI-Tthe	2	Osph2G54760.2; Osph2G54880.1
CMGC_RCK	2	Osph8G45160.1; Osph549G10050.1
CMGC_SRPK	3	Osph8G17430.1; Osph106G10590.2; Osph194G10600.2
Group-PI-2	2	Osph1G32930.4; Osph1G33490.1
IRE1	2	Osph67G11720.1; Osph500G10100.1
NAK	5	Osph18G26140.1; Osph23G10580.1; Osph23G10730.1; Osph103G11400.2; Osph121G10040.2
NEK	3	Osph1G79830.4; Osph15G11460.1; Osph15G11590.1
RLK-Pelle_C-LEC	2	Osph2G12620.2; Osph2G12630.1
RLK-Pelle_CR4L	6	Osph1G19710.1; Osph3G16630.1; Osph12G35980.1; Osph22G13900.1; Osph118G10420.1; Osph118G11340.1
RLK-Pelle_CrRLK1L-1	15	Osph2G70790.1; Osph2G70990.1; Osph6G32780.1; Osph10G32030.1; Osph11G13930.1; Osph18G24160.1; Osph28G11850.1; Osph30G10630.2; Osph30G11130.1; Osph71G10620.1; Osph71G11100.1; Osph74G10230.1; Osph74G10430.1; Osph181G10050.1; Osph431G10100.1
RLK-Pelle_DLSV	68	Osph1G64260.3; Osph2G51110.1; Osph2G59410.2; Osph2G59590.1; Osph2G63020.3; Osph3G73720.2; Osph3G74100.1; Osph3G74700.1; Osph3G74830.1; Osph10G17570.1; Osph10G17750.7; Osph10G29900.1; Osph11G14910.2; Osph11G14980.1; Osph11G15310.1; Osph11G32390.1; Osph14G41560.1; Osph14G41710.1; Osph15G21780.1; Osph17G19110.5; Osph18G18190.1; Osph20G10330.1; Osph33G12470.1; Osph40G10760.1; Osph44G12190.2; Osph44G12390.4; Osph44G12450.1; Osph66G10830.1; Osph75G10560.1; Osph75G10660.1; Osph75G10670.3; Osph79G10610.1; Osph79G10730.1; Osph88G10940.2; Osph88G11220.1; Osph92G11370.4; Osph92G11380.2; Osph104G10370.1; Osph104G10590.1; Osph134G10050.1; Osph134G10240.4; Osph134G10300.1; Osph134G10460.1; Osph134G10550.1; Osph134G10740.2; Osph142G10150.3; Osph142G10160.1; Osph142G10170.1; Osph142G10450.3; Osph142G10480.1; Osph142G10900.2; Osph147G10380.3; Osph147G10770.3; Osph204G10030.1; Osph204G10290.1; Osph204G10310.1; Osph204G10550.1; Osph204G10650.1; Osph204G10690.1; Osph661G10020.2; Osph661G10060.2; Osph661G10080.2; Osph661G10170.1; Osph661G10180.1; Osph739G10020.3; Osph913G10070.1; Osph981G10030.3; Osph981G10060.1
RLK-Pelle_Extensin	7	Osph8G44400.1; Osph8G44760.1; Osph56G12220.2; Osph191G10760.1; Osph304G10410.3; Osph314G10100.1; Osph553G10180.3
RLK-Pelle_L-LEC	11	Osph4G33610.1; Osph4G49410.1; Osph4G49750.1; Osph4G49910.1; Osph4G50000.1; Osph6G34350.1; Osph9G43170.2; Osph11G15000.1; Osph122G11110.1; Osph122G11490.1; Osph227G10580.1
RLK-Pelle_LRR-I-1	4	Osph5G50590.1; Osph5G50850.1; Osph7G15400.1; Osph20G10590.1

RLK-Pelle_LRR-II	12	Osph5G53630.1; Osph26G11270.1; Osph32G12660.1; Osph43G11750.3; Osph83G10340.1; Osph115G10910.5; Osph122G11190.1; Osph122G11210.1; Osph122G11570.1; Osph152G10250.1; Osph214G10220.1; Osph388G10130.1
RLK-Pelle_LRR-III	40	Osph1G18270.1; Osph1G57110.1; Osph1G57240.1; Osph1G76860.2; Osph1G77230.1; Osph3G54720.1; Osph3G54740.1; Osph3G73600.2; Osph6G20070.2; Osph6G20150.1; Osph7G10280.1; Osph7G17250.3; Osph10G14690.1; Osph14G43330.1; Osph14G44310.1; Osph15G33660.1; Osph43G10800.1; Osph57G10040.2; Osph69G11690.1; Osph69G11840.1; Osph74G11320.1; Osph74G11820.1; Osph76G11110.1; Osph83G10160.1; Osph87G10180.2; Osph88G10480.1; Osph88G10700.1; Osph118G10830.2; Osph118G10960.1; Osph124G10980.1; Osph124G11350.1; Osph170G10430.1; Osph170G10440.1; Osph181G10190.1; Osph181G10510.1; Osph181G10520.1; Osph225G10050.2; Osph572G10090.1; Osph775G10090.1; Osph775G10160.1
RLK-Pelle_LRR-IX	6	Osph9G47690.1; Osph15G35440.1; Osph15G35750.1; Osph68G10330.1; Osph68G10580.1; Osph218G10400.1
RLK-Pelle_LRR-V	10	Osph2G31250.1; Osph22G12240.1; Osph64G11800.2; Osph64G11820.1; Osph71G10500.1; Osph71G10960.2; Osph87G10470.1; Osph128G10110.1; Osph162G10990.1; Osph1017G10050.1
RLK-Pelle_LRR-VI-1	3	Osph15G31670.1; Osph73G11140.1; Osph779G10010.1
RLK-Pelle_LRR-VI-2	11	Osph7G16690.1; Osph7G16770.2; Osph9G37290.3; Osph41G10730.1; Osph41G10810.2; Osph41G10850.1; Osph47G11580.1; Osph47G11740.1; Osph92G10340.2; Osph92G10720.1; Osph1271G10020.1
RLK-Pelle_LRR-VII-1	4	Osph6G26430.1; Osph7G20120.4; Osph15G27690.1; Osph80G10550.1
RLK-Pelle_LRR-VII-2	3	Osph47G11630.1; Osph54G10880.1; Osph54G11110.1
RLK-Pelle_LRR-VIII-1	5	Osph1G26060.4; Osph4G40170.3; Osph7G10340.1; Osph74G11200.4; Osph74G11590.1
RLK-Pelle_LRR-Xa	4	Osph10G10440.1; Osph10G10560.1; Osph278G10150.1; Osph311G10010.1
RLK-Pelle_LRR-Xb-1	13	Osph2G37970.1; Osph12G13010.1; Osph12G13100.1; Osph13G31440.1; Osph13G31590.1; Osph16G18880.1; Osph47G10230.1; Osph81G10160.1; Osph81G10400.1; Osph93G10250.1; Osph362G10310.1; Osph1101G10050.1; Osph1119G10080.1
RLK-Pelle_LRR-Xb-2	2	Osph179G10280.1; Osph179G10570.1
RLK-Pelle_LRR-XI-1	54	Osph1G73590.1; Osph1G73940.1; Osph1G75000.1; Osph2G63340.1; Osph2G63560.1; Osph2G69510.1; Osph3G77300.1; Osph5G43770.1; Osph5G53770.1; Osph5G54050.1; Osph11G15700.1; Osph11G15720.1; Osph11G16320.1; Osph12G24670.1; Osph13G31480.1; Osph15G10640.1; Osph17G26220.1; Osph34G12590.1; Osph34G12740.2; Osph39G11870.1; Osph39G12020.1; Osph39G12480.1; Osph39G12490.1; Osph40G11250.1; Osph40G11280.1; Osph42G11020.1; Osph42G11180.1; Osph47G10860.1; Osph49G12050.1; Osph50G11930.1; Osph50G12660.2; Osph67G11460.1; Osph67G11470.1; Osph67G11510.1; Osph67G11590.1; Osph67G11600.1; Osph67G11610.1; Osph67G11630.1; Osph74G11340.1; Osph74G11840.1; Osph95G10910.2; Osph95G11170.1; Osph97G10610.1; Osph97G10760.1; Osph104G10790.1; Osph132G10470.1; Osph132G10890.1; Osph169G10310.1; Osph173G10290.1; Osph196G10100.1; Osph196G10150.1; Osph196G10160.1; Osph987G10020.1; Osph1322G10030.1
RLK-Pelle_LRR-XI-2	2	Osph7G41090.1; Osph110G10200.1
RLK-Pelle_LRR-XII-1	24	Osph11G20460.1; Osph11G30630.1; Osph16G36330.2; Osph47G10390.1; Osph47G10400.1; Osph47G10430.1; Osph47G10450.1; Osph47G10460.1; Osph47G10490.1; Osph47G10540.1; Osph47G10660.4; Osph47G10670.1; Osph47G10690.1; Osph47G10700.1; Osph56G10430.1; Osph56G10870.1; Osph89G10050.1; Osph89G10340.1; Osph89G10460.1; Osph89G10730.1; Osph104G10800.1; Osph104G11510.1; Osph160G10300.1; Osph1218G10060.1
RLK-Pelle_LRR-XIIIa	3	Osph1G70260.1; Osph2G58220.1; Osph4G59210.1

RLK-Pelle_LRR-XIIb	2	Osph36G10290.2; Osph66G10650.1
RLK-Pelle_LRR-XIV	5	Osph86G10860.1; Osph99G11530.2; Osph265G10080.1; Osph939G10060.3; Osph1203G10050.3
RLK-Pelle_LRR-XV	6	Osph71G11580.1; Osph89G10320.1; Osph89G10680.1; Osph120G10320.1; Osph120G10720.1; Osph791G10080.1
RLK-Pelle_LysM	10	Osph3G65060.1; Osph8G43180.1; Osph16G10990.1; Osph20G10220.1; Osph40G11780.1; Osph41G11320.1; Osph41G11780.1; Osph115G11030.1; Osph143G10600.1; Osph143G10700.1
RLK-Pelle_PERK-1	10	Osph7G46320.1; Osph71G10530.2; Osph71G10990.1; Osph86G10520.2; Osph86G10630.1; Osph102G10580.1; Osph141G10230.1; Osph141G10320.1; Osph268G10010.3; Osph275G10270.3
RLK-Pelle_PERK-2	3	Osph102G10490.1; Osph191G10150.1; Osph191G10260.2
RLK-Pelle_RKF3	3	Osph13G14440.1; Osph13G14780.1; Osph244G10070.1
RLK-Pelle_RLCK-IV	5	Osph1G60360.1; Osph1G60460.1; Osph115G10310.1; Osph115G10480.1; Osph143G10440.1
RLK-Pelle_RLCK-IXb	9	Osph2G67130.1; Osph2G67240.1; Osph3G18020.1; Osph7G41480.5; Osph8G30620.1; Osph33G10390.2; Osph33G10640.1; Osph56G11890.1; Osph156G10780.1
RLK-Pelle_RLCK-V	8	Osph5G28010.3; Osph13G30230.3; Osph15G17270.1; Osph20G10030.2; Osph48G10390.2; Osph86G10750.3; Osph86G10810.3; Osph457G10090.1
RLK-Pelle_RLCK-VI	16	Osph2G70740.1; Osph2G70920.3; Osph3G34990.1; Osph4G56720.1; Osph4G57920.3; Osph7G17960.1; Osph17G14640.1; Osph20G10980.2; Osph20G10990.1; Osph20G11290.2; Osph28G11690.2; Osph68G11720.2; Osph74G11960.1; Osph74G12170.1; Osph136G10490.1; Osph156G10770.4
RLK-Pelle_RLCK-VIIa-1	15	Osph3G44740.8; Osph5G17620.2; Osph7G15640.1; Osph11G25570.1; Osph11G40020.1; Osph35G11090.1; Osph74G12590.1; Osph95G10280.3; Osph95G10360.1; Osph153G10470.1; Osph153G10500.1; Osph359G10130.1; Osph407G10080.1; Osph407G10090.1; Osph671G10030.1
RLK-Pelle_RLCK-VIIa-2	36	Osph1G32630.1; Osph3G29490.1; Osph3G29510.2; Osph3G76930.1; Osph3G78870.1; Osph4G43520.1; Osph4G43670.4; Osph6G33620.1; Osph8G45170.1; Osph9G22690.3; Osph9G45320.3; Osph9G45530.4; Osph13G15860.1; Osph13G16020.1; Osph27G10340.5; Osph27G10770.1; Osph36G10060.1; Osph36G11650.1; Osph66G12260.1; Osph89G10770.2; Osph114G10760.1; Osph114G10880.4; Osph114G10960.1; Osph122G11000.3; Osph123G10350.4; Osph133G11170.3; Osph156G10370.3; Osph181G10270.1; Osph181G10370.2; Osph277G10070.1; Osph277G10310.3; Osph281G10490.2; Osph372G10170.1; Osph539G10190.1; Osph548G10010.1; Osph691G10230.1
RLK-Pelle_RLCK-VIIb	2	Osph146G10070.1; Osph899G10070.1
RLK-Pelle_RLCK-VIII	12	Osph2G21030.1; Osph2G21060.1; Osph2G21090.1; Osph2G21130.2; Osph3G18840.4; Osph5G57930.1; Osph11G18090.1; Osph11G18310.3; Osph11G18320.1; Osph17G25080.3; Osph45G10910.1; Osph45G10990.2
RLK-Pelle_RLCK-X	2	Osph239G10250.1; Osph399G10030.1
RLK-Pelle_RLCK-XI	2	Osph13G10320.1; Osph13G10520.1
RLK-Pelle_RLCK-XII-1	6	Osph1G63030.1; Osph13G160.1; Osph52G10790.1; Osph74G12450.2; Osph155G10360.1; Osph471G10180.1
RLK-Pelle_RLCK-XV	11	Osph2G25860.3; Osph3G26920.1; Osph3G73770.2; Osph3G74640.1; Osph18G22340.3; Osph18G28830.1; Osph54G12410.3; Osph175G10730.1; Osph175G10900.1; Osph536G10010.1; Osph881G10030.1
RLK-Pelle_RLCK-XVI	2	Osph41G10600.1; Osph41G10670.2
RLK-Pelle_SD-2b	30	Osph1G60320.1; Osph1G60550.1; Osph5G42430.1; Osph11G28990.1; Osph39G12430.1; Osph47G11530.1; Osph47G11600.1; Osph47G11670.1; Osph75G10330.1; Osph92G11510.1; Osph93G10770.1; Osph93G10790.1; Osph93G10800.1; Osph93G11050.1; Osph93G11060.1; Osph104G10330.1; Osph104G10560.1; Osph106G10900.1; Osph142G11140.1; Osph185G10310.2; Osph194G10280.1; Osph240G10090.1; Osph240G10110.1; Osph240G10270.1; Osph279G10240.1;

		Osph347G10040.1; Osph347G10090.1; Osph347G10130.1; Osph578G10220.1; Osph1016G10020.1
RLK-Pelle_URK-1	4	Osph4G57060.1; Osph4G58010.1; Osph98G11020.1; Osph98G11730.2
RLK-Pelle_WAK_LRK10L-1	6	Osph5G45850.1; Osph5G45970.2; Osph78G12340.1; Osph103G10880.3; Osph155G10770.1; Osph592G10340.1
STE_STE11	31	Osph1G72000.4; Osph2G10500.5; Osph2G37170.1; Osph3G14250.1; Osph3G14310.1; Osph6G12440.1; Osph12G16930.1; Osph12G16980.1; Osph12G28240.1; Osph14G23640.1; Osph17G29890.1; Osph17G29980.1; Osph19G24510.1; Osph23G13000.2; Osph23G13170.1; Osph26G12540.2; Osph40G10160.1; Osph40G10510.1; Osph84G10290.1; Osph97G11470.1; Osph100G10840.1; Osph118G10470.1; Osph118G11310.1; Osph139G10860.1; Osph139G11230.1; Osph154G11300.2; Osph154G11330.1; Osph177G10720.1; Osph213G10070.2; Osph213G10080.1; Osph737G10220.1
STE_STE20-Fray	5	Osph1G11590.1; Osph1G11920.1; Osph2G69430.1; Osph15G18730.1; Osph152G10300.1
STE_STE7	8	Osph1G77460.1; Osph5G55070.1; Osph5G55180.1; Osph7G41420.4; Osph22G12480.2; Osph26G10550.1; Osph68G10430.1; Osph68G10650.1
STE_STE-PI	3	Osph24G10160.1; Osph24G10290.3; Osph66G11390.1
TKL_CTR1-DRK-2	13	Osph2G16650.1; Osph11G41190.1; Osph12G34240.1; Osph13G16780.2; Osph13G16890.2; Osph22G11400.1; Osph22G11460.1; Osph121G10810.1; Osph133G11240.1; Osph288G10010.1; Osph329G10440.1; Osph499G10010.1; Osph499G10060.1
TKL-PI-2	2	Osph13G19500.1; Osph13G19570.1
TKL-PI-4	24	Osph1G79730.3; Osph2G10520.1; Osph4G21350.3; Osph6G37400.1; Osph10G13880.1; Osph39G10540.1; Osph39G11070.1; Osph43G12280.1; Osph74G11210.1; Osph79G10810.1; Osph79G11010.2; Osph101G10760.1; Osph101G10930.1; Osph148G10150.1; Osph225G10630.1; Osph263G10200.1; Osph263G10280.1; Osph356G10010.1; Osph456G10010.1; Osph474G10130.1; Osph530G10240.1; Osph530G10400.2; Osph834G10010.1; Osph1016G10030.1
TKL-PI-5	3	Osph10G23670.2; Osph13G40340.1; Osph16G16550.1
TKL-PI-6	6	Osph1G56370.1; Osph1G56450.1; Osph2G61330.1; Osph124G11620.1; Osph139G10340.1; Osph641G10040.1
TKL-PI-7	2	Osph118G10650.1; Osph118G11180.2
TLK	2	Osph106G10480.3; Osph194G10680.1
TTK	6	Osph44G10190.1; Osph44G10220.1; Osph44G10300.4; Osph213G10640.1; Osph412G10050.1; Osph412G10290.1
ULK_Fused	3	Osph383G10030.2; Osph918G10090.1; Osph918G10110.1
WEE	2	Osph62G12260.2; Osph62G12400.1
WNK_NRBP	6	Osph6G52960.3; Osph13G11250.1; Osph93G10840.2; Osph279G10040.3; Osph319G10330.3; Osph859G10050.1

532
533
534
535

Supplementary Table 11. Repetitive content in the *Ophrys sphegodes* genome

Type	Code	Length (bp)	% in the genome ^a
Class I			
LTR/Gypsy	RLG	2 265 079 734	43.39
LTR/Copia	RLC	1 593 298 782	30.52
LTR/LTR	RLX	34 618 888	0.66
LTR/Bel-Pao	RLB	0	0.00
LTR/Retrovirus	RLR	0	0.00
LINE/R2	RIR	12 898 808	0.25
LINE/L1	RIL	17 816 225	0.34
LINE/RTE	RIT	18 947 522	0.36
Class II			
TIR/CACTA	DTC	28 268 745	0.54
TIR/Pif-Harbinger	DTH	27 869 444	0.53
TIR/hAT	DTA	11 632 404	0.22
TIR/Mutator	DTM	8 833 173	0.17
TIR/Tc1-Mariner	DTT	105 801	0.00
Helitron	DHH	20 226 278	0.39
Simple repeats		34 196 269	0.66
Total		4 073 792 073	78.06

^a Total genome length 5 219 753 694 bp

Supplementary Table 12. Repetitive elements in orchids

Type	<i>Pgu</i> (%)	<i>Pzi</i> (%)	<i>Ash</i> (%)	<i>Dca</i> (%)	<i>Peq</i> (%)	<i>Gel</i> (%)
LTR/Bhikhari	0.00	0.00	0.00	0.00	0.00	0.00
LTR/Cassandra	0.00	0.02	0.00	0.00	0.00	1.49
LTR/Caulimoviru	0.02	0.04	0.06	0.04	0.07	0.00
LTR/Caulimovirus	0.04	0.07	0.33	0.02	0.07	0.04
LTR/Copia	35.75	33.98	4.97	24.70	7.31	6.42
LTR/Copia(Xen1)	0.00	0.00	0.00	0.00	0.00	0.00
LTR/Delta	0.00	0.00	0.00	0.00	0.00	0.00
LTR/DIRS	0.01	0.02	0.00	0.03	0.00	0.01
LTR/ERV	0.00	0.00	0.00	0.00	0.00	0.00
LTR/ERV-Foamy	0.00	0.00	0.00	0.00	0.00	0.00
LTR/ERV-Lenti	0.00	0.00	0.00	0.00	0.00	0.00
LTR/ERV1	0.20	0.21	0.10	0.14	0.20	0.02
LTR/ERV4	0.00	0.00	0.00	0.00	0.00	0.00
LTR/ERVK	0.09	0.08	0.07	0.01	0.03	0.00
LTR/ERVL	0.03	0.02	0.00	0.00	0.00	0.00
LTR/ERVL-MaLR	0.00	0.00	0.00	0.00	0.00	0.00
LTR/Foamy	0.01	0.01	0.00	0.00	0.00	0.00
LTR/Ginger	0.00	0.00	0.00	0.00	0.00	0.00
LTR/Gypsy	37.48	37.80	11.84	15.18	36.88	48.33
LTR/Gypsy-Cigr	0.06	0.04	0.02	0.01	0.02	0.00
LTR/Gypsy-Troyk	0.00	0.00	0.00	0.00	0.00	0.00
LTR/Gypsy-Troyka	0.00	0.00	0.00	0.00	0.00	0.00
LTR/Lenti	0.00	0.00	0.00	0.00	0.00	0.00
LTR/LTR	3.49	5.82	5.58	4.17	4.23	4.72
LTR/Ngaro	0.00	0.00	0.04	0.00	0.04	0.30
LTR/Pao	0.01	0.03	0.03	0.01	0.02	0.02
LTR/TATE	0.00	0.00	0.00	0.00	0.00	0.00
LTR/Viper	0.00	0.00	0.00	0.00	0.00	0.00
Total Gypsy/Copia	73.23	71.78	16.81	39.88	44.19	54.75

543

544

545

546

547

548

549

550

Repetitive elements in six orchid species (table adapted from Li, MH. et al. 2022. *Nat. Plants*. 8, 373-388, Supplementary Table 8). Repetitive elements percentage for *V. planifolia* are from Hasing, T. et al, 2020. *Nat. Food*. 1, 811-819, Supplementary Table 2. The listed species are: *Pgu* = *Platanthera guangdongensis*; *Pzi* = *Platanthera zijinensis*; *Ash* = *Apostasia shenzhenica*; *Dca* = *Dendrobium catenatum*; *Peq* = *Phalaenopsis equestris*; *Gel* = *Gastrodia elata*.

Supplementary Table 13. Statistics of sequence variants

Data set ^a	Filtering ^b	Individuals	Genotype calls
WGS	None	8	191 200 928
	Biallelic	8	174 192 143
	biallelic & SNP	8	146 090 706
	biallelic & indel	8	11 010 748
	DP≥10	8	9 567 686
	DP≥10 & biallelic	8	7 961 327
	DP≥20	8	3 485 576
	DP≥20 & biallelic	8	2 874 985
RNA-Seq	None	37	13 177 461
	Biallelic ^c	32	12 535 528
	DP≥5 & biallelic ^c	32	439 708
	DP≥10 & biallelic ^c	32	330 449
	DP≥20 & biallelic ^c	32	233 053
GBS ^d	none	126	1 680 249
	biallelic	126	1 621 433
	biallelic & SNP	126	1 356 864
	biallelic & indel	126	120 557
	DP≥5 & biallelic	126	15
	biallelic & ($N_A \geq 6 \times 4$ pops)	126	18 862
	biallelic & ($N_A \geq 6 \times 4$ pops) & $F_{ST} \geq 0.001$	126	8 541

^a WGS: whole-genome (re)sequencing data set from this study (2 individuals $\times$ 4 species); RNA-Seq, raw data from ref.²⁴; GBS, raw data from ref.⁵².

^b Filtering applied across all individuals; DP here refers to the VCF SAMPLE/FORMAT DP tag, not the INFO DP tag. N_A refers to the number of allele calls counted from a given population; F_{ST} as calculated by SPA software. ‘&’ denotes a logical AND operator.

^c applied to data set excluding bud samples pooled from different individuals and samples with unclear population assignment.

^d using the data set after exclusion of putatively mis-assigned individual 527I.

Supplementary Table 14. Distance matrix for comparing genotype calls.

Pairwise distance matrix for diploid unphased single-locus genotype calls, including partial genotype calls with an allele call missing. Here, 0 and 1 refer to two different alleles (e.g. A and T in a SNP call), whereas ? denotes a missing allele.

<i>Distance</i>	0/0	0/1	1/1	0/?	1/?
0/0	0				
0/1	0.5	0			
1/1	1	0.5	0		
0/?	0	0.5	1	0	
1/?	1	0.5	0	1	0

Supplementary Table 15. Point mutation transition matrices

RNA-Seq data set (32 × DP≥5); N= 389 561				
	A	C	G	T
A	0.0000	0.1112	0.2940	0.1052
C	0.1112	0.0000	0.0840	0.2926
G	0.2940	0.0840	0.0000	0.1131
T	0.1052	0.2926	0.1131	0.0000
GBS data set; N= 1 356 793				
	A	C	G	T
A	0.0000	0.1271	0.2897	0.0963
C	0.1271	0.0000	0.0707	0.2897
G	0.2897	0.0707	0.0000	0.1265
T	0.0963	0.2897	0.1265	0.0000
WGS data set (8 × DP≥10); N= 7 260 426				
	A	C	G	T
A	0.0000	0.1314	0.2966	0.0989
C	0.1314	0.0000	0.0459	0.2963
G	0.2966	0.0459	0.0000	0.1309
T	0.0989	0.2963	0.1309	0.0000

574 **Supplementary Table 16. Statistics of syntenic analysis of *O. sphegodes* (21 462**
575 **genes) compared with other orchid genomes**

576

Orchid species 2	N genes species 2	N collinear gene pairs	N collinear blocks	Mean N genes per block
<i>Cymbidium goeringii</i>	29 556	1767	153	11.55
<i>Dendrobium chrysotoxum</i>	27 094	4230	208	20.34
<i>Phalaenopsis aphrodite</i>	19 666	3970	191	20.79
<i>Platanthera zjzensis</i>	24 513	5624	148	38.00
<i>Platanthera guangdongensis</i>	22 559	5216	135	38.64
<i>Vanilla planifolia</i>	29 167	3441	337	10.21

Supplementary Table 17. Gene family expansion and contraction across 21 plant species with sequenced genome.

Species	Code	Expanded	Contracted
<i>Amborella trichopoda</i>	<i>Atr</i>	188	2 744
<i>Ananas comosus</i>	<i>Aco</i>	1 136	1 635
<i>Apostasia shenzhenica</i>	<i>Ash</i>	672	2 482
<i>Arabidopsis thaliana</i>	<i>Ath</i>	2 026	2 029
<i>Asparagus officinalis</i>	<i>Aof</i>	1 668	2 335
<i>Brachypodium distachyon</i>	<i>Bdi</i>	721	1 025
<i>Cymbidium goeringii</i>	<i>Cgo</i>	431	6 228
<i>Dendrobium catenatum</i>	<i>Dca</i>	1 817	391
<i>Gastrodia elata</i>	<i>Gel</i>	765	2 994
<i>Musa acuminata</i>	<i>Mac</i>	4 115	829
<i>Ophrys sphegodes</i>	<i>Osph</i>	3 712	756
<i>Oryza sativa</i>	<i>Osa</i>	1 046	504
<i>Phalaenopsis equestris</i>	<i>Peq</i>	1 714	529
<i>Phoenix dactylifera</i>	<i>Pda</i>	3 653	610
<i>Platanthera guangdongensis</i>	<i>Pgu</i>	841	1 969
<i>Platanthera zijinensis</i>	<i>Pzi</i>	732	1 368
<i>Populus trichocarpa</i>	<i>Ptr</i>	4 640	429
<i>Sorghum bicolor</i>	<i>Sbi</i>	945	712
<i>Spirodela polyrhiza</i>	<i>Spo</i>	866	3 301
<i>Vanilla planifolia</i>	<i>Vpl</i>	3 248	1 568
<i>Vitis vinifera</i>	<i>Vvi</i>	907	1 409

Supplementary Table 18. Accession number of orchid genome assemblies used in this study

Species	Assembly type	Assembly size	Annotated genes	Source
<i>Cymbidium goeringii</i>	chromosomes	3.99 Gb	29 556	http://orchidgenome.com/
<i>Dendrobium catenatum</i>	scaffolds	1.11 Gb	29 257	Requested from authors
<i>Dendrobium chrysotoxum</i>	chromosomes	1.37 Gb	30 044	PRJNA664445
<i>Gastrodia elata</i>	scaffolds	1.06 Gb	18 969	PRJCA000931
<i>Gastrodia elata</i>	chromosomes	1.09 Gb	21 115	PRJCA005619
<i>Phalaenopsis equestris</i>	scaffolds	1.13 Gb	29 545	PRJNA262478
<i>Platanthera guangdongensis</i>	chromosomes	4.27 Gb	22 559	Requested from authors
<i>Platanthera zijinensis</i>	chromosomes	4.19 Gb	24 513	Requested from authors
<i>Vanilla planifolia</i>	chromosome-phased	736 Mb ^a	22 026 ^a	PRJNA668740

Source lists NCBI identifiers. ^a haplotype A.

6 References

1. Wick, R. R., et al. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 1–10 (2019).
2. De Coster, et al. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
3. Li, H. Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).
4. Nagarajan, N., et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
5. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
6. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics*. doi: 10.6084/M9.FIGSHARE.963153.V1 (2013).
7. Walker, B. J. et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
8. Nagarajan, N. & Pop, M. Sequence assembly demystified. *Nat. Rev. Genet.* **14**, 157–167 (2013).
9. Simpson, J. T. & Pop, M. The theory and practice of genome sequence assembly. *Annu. Rev. Genomics Hum. Genet.* **16**, 153–172 (2015).
10. Li, F. W. & Harkess, A. A guide to sequence your favorite plant genomes. *Applications in Plant Sciences* **6**, e1030 (2018).

- 611 11. Lantz, H. et al. Ten steps to get started in genome assembly and annotation.
612 *F1000Research* **7**, 148 (2018).
- 613 12. Roach, M. J., et al. Purge Haplotigs: allelic contig reassignment for third-gen diploid
614 genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).
- 615 13. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome
616 assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
- 617 14. Ghurye, J., et al. Scaffolding of long read assemblies using long range contact
618 information. *BMC Genomics* **18**, 1–11 (2017).
- 619 15. Pryszcz, L. P. & Gabaldón, T. Redundans: An assembly pipeline for highly
620 heterozygous genomes. *Nucleic Acids Res.* **44**, e113 (2016).
- 621 16. Grob, S., et al. Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with
622 similarities to the flamenco locus of *Drosophila*. *Mol. Cell* **55**, 678–693 (2014).
- 623 17. Liu, C. In Situ Hi-C library preparation for plants to study their three-dimensional
624 chromatin interactions on a genome-wide scale. *Methods Mol. Biol.* **1629**, 155–166
625 (2017).
- 626 18. Ghurye, J. et al. Integrating Hi-C links with assembly graphs for chromosome-scale
627 assembly. *PLOS Comput. Biol.* **15**, e1007273 (2019).
- 628 19. Durand, N. C. et al. Juicebox provides a visualization system for hi-c contact maps
629 with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
- 630 20. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C
631 yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).

- 632 21. Nakabayashi, R. & Morishita, S. HiC-Hiker: A probabilistic model to determine contig
633 orientation in chromosome-length scaffolds with Hi-C. *Bioinformatics* **36**, 3966–3974
634 (2020).
- 635 22. Simão, F. A., et al. BUSCO: Assessing genome assembly and annotation completeness
636 with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
- 637 23. Manni, M., et al. BUSCO update: novel and streamlined workflows along with broader
638 and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral
639 genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
- 640 24. Piñeiro Fernández, L. et al. A phylogenomic analysis of the floral transcriptomes of
641 sexually deceptive and rewarding european orchids, *Ophrys* and *Gymnadenia*. *Front.*
642 *Plant Sci.* **10**, 1–13 (2019).
- 643 25. Wicker, T. et al. A unified classification system for eukaryotic transposable elements.
644 *Nat. Rev. Genet.* **8**, 973–982 (2007).
- 645 26. Smit, A., et al. RepeatMasker Open-4.0. <http://www.repeatmasker.org> (2015).
- 646 27. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–
647 2948 (2007).
- 648 28. Hoff, K. J., et al. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**,
649 65–95 (2019).
- 650 29. Hoff, K. J. & Stanke, M. Predicting Genes in Single Genomes with AUGUSTUS.
651 *Curr. Protoc. Bioinforma.* **65**, e57 (2019).
- 652 30. Brůna, T., et al. GeneMark-EP+: eukaryotic gene prediction with self-training in the

- 653 space of genes and proteins. *NAR Genomics Bioinforma.* **2**, lqaa026 (2020).
- 654 31. Buchfink, B., et al. Fast and sensitive protein alignment using DIAMOND. *Nat.*
655 *Methods* **12**, 59–60 (2014).
- 656 32. Keilwagen, J. et al. Using intron position conservation for homology-based gene
657 prediction. *Nucleic Acids Res.* **44**, e89–e89 (2016).
- 658 33. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal
659 transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654 (2003).
- 660 34. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15
661 (2013).
- 662 35. Zhang, G. Q. et al. The *Apostasia* genome and the evolution of orchids. *Nature* **549**,
663 379–383 (2017).
- 664 36. Chung, O. et al. A chromosome-scale genome assembly and annotation of the spring
665 orchid (*Cymbidium goeringii*). *Mol. Ecol. Resour.* **22**, 1168–1177 (2021).
- 666 37. Kawahara, Y. et al. Improvement of the *Oryza sativa* Nipponbare reference genome
667 using next generation sequence and optical map data. *Rice* **6**, 3–10 (2013).
- 668 38. Harkess, A. et al. The asparagus genome sheds light on the origin and evolution of a
669 young Y chromosome. *Nat. Commun.* **8**, 1279 (2017).
- 670 39. Schlüter, P. M. & Schiestl, F. P. Molecular mechanisms of floral mimicry in orchids.
671 *Trends Plant Sci.* **13**, 228–235 (2008).
- 672 40. Schlüter, P. M. et al. Stearoyl-acyl carrier protein desaturases are associated with floral

673 isolation in sexually deceptive orchids. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5696–5701
674 (2011).

675 41. Xu, S., et al. The genetic basis of pollinator adaptation in a sexually deceptive orchid.
676 *PLoS Genet.* **8**, e1002889 (2012).

677 42. Sedeek, K. E. M. et al. Transcriptome and proteome data reveal candidate genes for
678 pollinator attraction in sexually deceptive orchids. *PLoS One* **8**, e64621 (2013).

679 43. Rice, P., et al. EMBOSS: The European Molecular Biology Open Software Suite.
680 *Trends Genet.* **16**, 276–277 (2000).

681 44. Dagona, A. G. BioEdit: a user-friendly biological sequence alignment editor and
682 analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).

683 45. Buccini, F. et al. TRAPID 2.0: a web application for taxonomic and functional
684 analysis of de novo transcriptomes. *Nucleic Acids Res.* **49**, e101–e101 (2021).

685 46. Van Bel, M. et al. PLAZA 4.0: an integrative resource for functional, evolutionary and
686 comparative plant genomics. *Nucleic Acids Res.* **46**, D1190–D1196 (2018).

687 47. Zheng, Y. et al. iTAK: A program for genome-wide prediction and classification of
688 plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* **9**,
689 1667–1670 (2016).

690 48. Chan, P. P., et al. tRNAscan-SE 2.0: improved detection and functional classification
691 of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).

692 49. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches.
693 *Bioinformatics* **29**, 2933–2935 (2013).

- 694 50. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing.
695 *Nat. Methods* **14**, 407–410 (2017).
- 696 51. Lee, I. et al. Simultaneous profiling of chromatin accessibility and methylation on
697 human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199 (2020).
- 698 52. Sedeek, K. E. M. et al. Genic rather than genome-wide differences between sexually
699 deceptive *Ophrys* orchids with different pollinators. *Mol. Ecol.* **23**, 6192–6205 (2014).
- 700 53. Devey, D. S., et al. Friends or relatives? Phylogenetics and species delimitation in the
701 controversial European orchid genus *Ophrys*. *Ann. Bot.* **101**, 385–402 (2008).
- 702 54. Paulus, H. & Gack, C. Pollinators as prepollinating isolation factors: evolution and
703 speciation in *Ophrys* (Orchidaceae). *Isr. J. Bot.* **39**, 43–79 (1990).
- 704 55. Xu, S. et al. Floral isolation is the main reproductive barrier among closely related
705 sexually deceptive orchids. *Evolution* **65**, 2606–2620 (2011).
- 706 56. Kosman, E. & Leonard, K. J. Similarity coefficients for molecular markers in studies
707 of genetic relationships between individuals for haploid, diploid, and polyploid
708 species. *Mol. Ecol.* **14**, 415–424 (2005).
- 709 57. Gower, J. C. Some distance properties of latent root and vector methods used in
710 multivariate analysis. *Biometrika* **53**, 325 (1966).
- 711 58. Schlüter, P. M. & Harris, S. A. Analysis of multilocus fingerprinting data sets
712 containing missing data. *Mol. Ecol. Notes* **6**, 569–572 (2006).
- 713 59. Cavalli-Sforza, L. L. & Edwards, A. W. F. Phylogenetic analysis models and
714 estimation procedures. *Am. J. Hum. Genet.* **19**, (1967).

- 715 60. Takezaki, N. & Nei, M. Genetic distances and reconstruction of phylogenetic trees
716 from microsatellite DNA. *Genetics* **144**, 389–99 (1996).
- 717 61. Sedeek, K. E. M. et al. Amino acid change in an orchid desaturase enables mimicry of
718 the pollinator's sex pheromone. *Curr. Biol.* **26**, 1505–1511 (2016).
- 719 62. Aceto, S. et al. The analysis of the inflorescence miRNome of the orchid *Orchis italica*
720 reveals a *DEF*-like MADS-box gene as a new miRNA target. *PLoS One* **9**, e97839
721 (2014).
- 722 63. Valoroso, M. C., et al. The MADS-box genes expressed in the inflorescence of *Orchis*
723 *italica* (Orchidaceae). *PLoS One* **14**, e0213185 (2019).
- 724 64. Montieri, S., et al. Isolation of the LFY/FLO homologue in *Orchis italica* and
725 evolutionary analysis in some European orchids. *Gene* **333**, 101–109 (2004).
- 726 65. Schlüter, P. M., et al. A screen of low-copy nuclear genes reveals the *LFY* gene as
727 phylogenetically informative in closely related species of orchids (*Ophrys*). *Taxon* **56**,
728 493–504 (2007).
- 729 66. Gilding, E. K. & Marks, M. D. Analysis of purified glabra3-shapeshifter trichomes
730 reveals a role for NOECK in regulating early trichome morphogenic events. *Plant J.*
731 **64**, 304–317 (2010).

732