Supplementary Information

The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defenses

Supplementary Text

Protein-coding gene annotation. To search for homologous genes, the protein sequences from all ferns and lycophytes transcriptomes in the OneKP project[1] were retrieved and aligned to the *A. capillus-veneris* genome, using GeneWise[2]. For transcriptome-based prediction, nineteen transcriptomes covering the entire life cycle of *A. capillus-veneris* were generated in this study (Supplementary Table 8). RNA was extracted using the Qiagen RNeasy protocol and sequenced on an Illumina HiSeq 4000 with a 300 bp insert size. For transcriptome-based prediction, the HISAT2[3] and StringTie[4] programs were used for transcript assembly[5]. The program PASA (http://pasapipeline.github.io) was used to align spliced transcripts and annotate candidate genes. *Ab initio* prediction was performed with AUGUSTUS[6], GlimmerHMM[7], and SNAP[8]. Finally, nonredundant gene models were obtained with EVidenceModeler (version 1.1.0)[9] to integrate the gene models developed by different datasets.

To validate the assembly quality, RNA-seq reads from nineteen tissues (Supplementary Table 8), together with publicly available EST sequences from the NCBI database (downloaded from http://togodb.dbcls.jp/library), were mapped to the *A. capillus-veneris* genome using HISAT2[3] and BLAT[10] with default parameters, respectively. The BLAT results were filtered with an identity and coverage cutoff of 0.9.

Identification of noncoding RNAs. We used tRNAscan-SE (version 2.0rc2)[11], with default parameters, to search for tRNAs in the *A. capillus-veneris* genome. A total of 1,624 tRNAs were found. Moreover, the Rfam14.0 database[12], including 3,445 noncoding RNA families, was used to annotate additional noncoding RNAs (ncRNAs), including miRNAs, snRNAs, and tRNAs, using INFERNAL (version 1.1.2)[13] program.

We predicted rRNA (5S, 5.8S, 28S, 18S) by using HMM searching based rRNA predicator Barrnap (version 0.9, https://github.com/tseemann/barrnap#barrnap), with default parameters. We finally identified 145 5S, 75 5.8S, 155 28S, and 165 18S sequences and their locations within the genome assembly of *A. capillus-veneris*.

Identification of long noncoding RNAs (lncRNAs). In contrast to other ncRNAs, lncRNAs are defined as transcripts longer than 200 bp that do not encode proteins. These lncRNAs were predicted using the RNA-seq datasets from nineteen tissues shown in Supplementary Table 8. The RNA-seq reads from the samples were aligned to the genome of *A. capillus-veneris* using HISAT2 (version 2.1.0)[3], and unique aligned reads were used for further analysis. Transcripts representing all isoforms were identified using StringTie (version 1.3.5)[4]. Transcripts with lengths shorter than 200 bp and open reading frames (ORFs) longer than 100 amino acids likely encoding proteins were filtered. Subsequently, three tests were used to filter lncRNAs. Firstly, all transcripts were aligned to the Swiss-Prot[14] database using BLASTX (version 2.5.0+)[15] with the E-value threshold at 0.001. Secondly, the ORFs and protein sequences were obtained using the transeq program in the EMBOSS package[16]. All predicted proteins were aligned to the Pfam database (Pfam-A)[17] using HMMER with a model E-value threshold of 0.001 and a domain E-value threshold of 0.001. Third, transcripts were predicted by the Coding Potential Calculator (CPC)[18] with default parameters. All sequences aligned to the above three databases were considered protein-coding genes and removed. As a result, 7,058 candidate lncRNAs were obtained. To further identify high-confidence lncRNAs, the obtained candidate lncRNAs were aligned with the ncRNA database[12], by BLASTn[15] with an identity threshold of 90%. Finally, we predicted 6,541 high-confidence lncRNAs.

Gene family gain and loss, expansion and contraction analysis. We used sixty-three low-copy gene families (one or two gene copies of each orthogroup for each genome), shared by the above eighteen species, to construct a phylogenetic species tree. MAFFT (version 7.471)[19] multiple sequence alignment tool was used to align these low-copy genes. The gap regions in the alignment were trimmed with trimAL (version 1.4.rev15)[20]. Poor quality alignments were filtered out by Gblocks (version 0.91b)[21], and only conserved regions were retained. The gap-trimmed orthologous sequences were concatenated into one supersequence with 6,853 amino acid sites for species tree construction. A maximum-likelihood phylogenic species tree was estimated by IQ-TREE (version 2.0.3)[22] with 1,000 bootstrap replicates.

We used the CAFE 3 (version 4.2.1) software tool[23] to count gene numbers at the nodes in the species tree and infer gene families that had undergone expansions or contractions. Following the instructions given in the CAFE 3 manual, we first transformed the above species tree into an ultrametric tree via r8s (version 1.81)[24]. The divergence times between *Arabidopsis* and *V. vinifera* (117 Mya), *O. sativa* and *Z. mays* (48 Mya), *G. biloba* and *P. taeda* (271~310 Mya), and *M. polymorpha* and *P. patens* (395~541 Mya) were provided to r8s for scaling branch lengths into time units. All the divergence times were estimated by TimeTree (http://timetree.org). The birth/death parameter (λ) was inferred for the entire tree. Families were retained if estimated to have undergone a significant expansion or contraction at the corresponding most recent common ancestor, with a *P*-value < 0.01.

We carried out the DOLLOP program from the PHYLIP package (version 3.696) to determine gene family gain and loss evolutions. DOLLOP is based on the Dollo parsimony principle, which assumes that genes arise exactly once on the evolutionary tree and can be lost independently in different evolutionary lineages. For Dollo inference, the supermatrix tree and a binary matrix derived from the matrix used for CAFE analysis were provided. The number of orthogroup gain and loss, in each branch and node, was further extracted using in-house Perl scripts. Finally, GO enrichment analyses were performed against the orthogroups that gained or lost, expanded or contracted at each key evolutionary node within euphyllophytes.

Expression profile of JA biosynthesis genes. The JA biosynthesis genes of *A. capillus-veneris* were identified via BLASTP *Arabidopsis* JA genes based on the *Arabidopsis* hormone database, version 2.0 (http://ahd.cbi.pku.edu.cn)[25] against *A. capillus-veneris* peptide sequences. The domains of the best hits were further identified, using HMM tools. The *A. capillus-veneris* genes with the same conserved domain as their *Arabidopsis* homologs were regarded as *A. capillus-veneris* JA biosynthesis genes.

The RNA-seq data from wounding treatment samples were aligned to the *A. capillus-veneris* genome by HISAT2[3]. Gene expression levels were represented by FPKM values based on the above-described methods. The expression values were further transformed by the function $\log_2$ (FPKM + 1) for the heatmap plot. Expression levels

of JA biosynthesis genes in *A. capillus-veneris* in leaves at specific time intervals after wounding treatment were extracted using a Perl script.

Jasmonate content determination. The sample metabolites were extracted using a modified Wolfender method[26]. First, the samples stored in a -80°C refrigerator were ground into powder in liquid nitrogen. Then, 100 mg of powder was weighed accurately and transferred into a 2 mL centrifuge tube. To correct loss that occurred during sample preparation, 10 μL of stable isotope internal standard (10 μg/mL d_5-JA, Cayman Chemical Co.) was added to each sample before extraction. Then, 1.5 mL of extraction buffer (isopropanol: formic acid at 99.5 : 0.5, v/v) was added, followed by 2 min of vortexing, three times, in a mixer mill at 900 rpm. After a 15-minute centrifugation at 14,000 × g, the supernatants were dried in a Labconco CentriVap vacuum centrifugal concentrator (Labconco, MO, USA) and then resuspended in 1 mL of methanol solvent (85 : 15, v/v). A Waters Sep-Pak C18 SPE column was further used for more readily detected and more accurately quantitated detection, and then a total of 2 mL of SPE column eluent was collected per sample. The eluent was dried in a Labconco CentriVap vacuum centrifugal concentrator and resuspended in 100 μL of methanol solvent (60 : 40, v/v) for LC-MS analysis.

For the determination of jasmonates, the prepared samples were analyzed with an AB Sciex 4500 QTRAP triple quadrupole mass spectrometer (AB SCIEX, MA, USA). We achieved separation of all targeted compounds on an ACQUITY UPLC™ BEH C18 column (Waters, Eschborn, Germany) (50 × 2.1 mm, 1.7 μm) based on an ACQUITY UPLC I-Class (Waters, Eschborn, Germany). The injection volume was set at 10 μL, and the flow rate was set at 200 μL/min. Solvents for the mobile phase were 0.1% formic acid in acetonitrile (A) and 0.1% formic acid in acetonitrile (B). The gradient elution was 0-10 min, linear gradient 50%-100% A, and then the column was washed with 100% B and 100% A before the next injection. The autosampler was set at 10°C to protect samples during analysis.

We acquired mass spectrometry in negative electrospray ionization mode, with [M - H] of target analytes selected as the precursor ion. Multiple reaction monitoring (MRM) mode was applied for quantification using the mass transitions (precursor ions/product

ions, Supplementary Table 13). General electrospray ionization (ESI) conditions were as follows: the temperature of the ESI ion source was 500°C, the curtain gas flow was 25 psi, collisionally activated dissociation (CAD) gas medium, and the ion spray voltage was (-) 4500 V for negative ionization mode, with ion gas 1 setting as 250 psi. Calibration standards were used to construct calibration curves of target analytes ranging with the same stable isotope internal standard added to the samples. Finally, we acquired mass spectra data and processed concentration calculations using AB SCIEX analyst 1.6.3 software (AB SCIEX, MA, USA).

Metabolite detection. The samples stored at -80°C were freeze dried and fully ground in a SPEX Geno2010 (SPEX SamplePrep, LLC., USA). Then, a 50 mg aliquot of freeze-dried leaf powder was weighed and transferred into a 2 mL centrifuge tube at liquid nitrogen temperature. To extract hydrophilic metabolites from samples, one milliliter of extraction buffer (MeOH : H_2O at 80 : 20, v/v, containing 0.1% formic acid) with internal standard (mixer of d_5-coumarin, d_3-DL-nicotine, d_5-L-phenylalanine, d_3-homovanillic acid, and nimodipine, with a final concentration of 1 μg/mL for each compound) was added, vortexed and subjected to a 1 h ultrasonic treatment in an ice bath for efficient resuspension of powder. Samples were then stored at 4°C, overnight. To more accurately compare metabolite differences, 800 μl of supernatant after a 15-minute centrifugation at 14,000 × g was pipetted and dried in a Labconco CentriVap vacuum centrifugal concentrator. One hundred microliters of reconstitution buffer (MeOH : H_2O at 80 : 20, v/v) with internal standard (mixer of d_5-L-tryptophan, d_5-succinic acid, d_3-nortriptyline, norethindrone, and diclofenac with a final concentration of 1 μg/mL for each compound) was used to resuspend each sample. Subsequently, a 14,000 × g centrifugation was applied twice to remove residuals before liquid chromatography–mass spectrometry/mass spectrometry (LC-MS/MS) analysis. A 10-μL aliquot of the supernatant for each sample was pipetted and mixed as QC samples, and a 1-μL aliquot was injected for LC-MS/MS analysis.

We performed a nontarget metabolomic analysis in a UPLC system coupled to a Q-Exactive orbitrap mass spectrometer (Thermo Fisher, CA, USA) equipped with an HESI probe under both positive and negative (ESI+ and ESI-) modes. Extracts were

separated by a Kinetex® Biphenyl column (Phenomenex, CA, USA) (150 × 2.1 mm, 2.6 μm).

The injection volume was 1 μL, and the flow rate was set at 250 μL/min. A 20-minute gradient was used, and solvents for the mobile phase were 2 mM ammonium acetate in water (A) and 100% acetonitrile (B). The column chamber and sample tray were held at 40°C and 10°C, respectively. The ESI source parameters were as follows: discharge current 6 μA, capillary temperature 320°C, heater temperature 250°C, sheath gas flow rate 35 Arb, auxiliary gas flow rate 10 Arb. Data with mass ranges of m/z 80-1200 and m/z 80-1200 were acquired in both positive ion mode and negative ion mode separately, with data-dependent MS/MS acquisition. The full scan and fragment spectra were collected with resolutions of 70,000 and 17,500, respectively.

We used TraceFinder analysis software (version 3.3, Thermo Fisher, MA, USA) to analyze MS and MS/MS data acquired and achieved more than 10,000 metabolite profiles for each sample, with data processing parameter settings as follows: minimum peak width of 10 s, maximum peak width of 60 s, mzwid of 0.015, minfrac of 0.5, bw of 5, and signal/noise threshold of 6. Then, we further used TraceFinder software to annotate metabolites from mass spectra data based on commercial databases and in-house databases via a high-resolution MS-associated method. First, the candidate chemical formulae of metabolites were identified using accurate high-resolution m/z values and MS/MS fragmentation patterns with a mass accuracy of 2 ppm based on in-house databases. Finally, the annotated metabolites were checked and verified again by manually comparing both fragment patterns and isotope ratios. Peak detection, retention time correction, chromatogram alignment, and statistical evaluation were performed. The ion intensity for identified compounds were extracted and calculated (i.e., EIC or extracted ion chromatography). After mean-centering and Pareto scaling, *t*-tests, ANOVA, and other statistical analyses for significantly changed metabolites were carried out based on the normalized data matrix. The results output contained a peak list with metabolite names, retention times, m/z values, *t*-test and ANOVA results, statistical significance values (*P*-values), and mean ion abundances with standard deviations (SDs). These results were used for subsequent multivariate statistical

analysis, and the compounds were classified according to the KEGG compound database[27] and PubChem database[28].

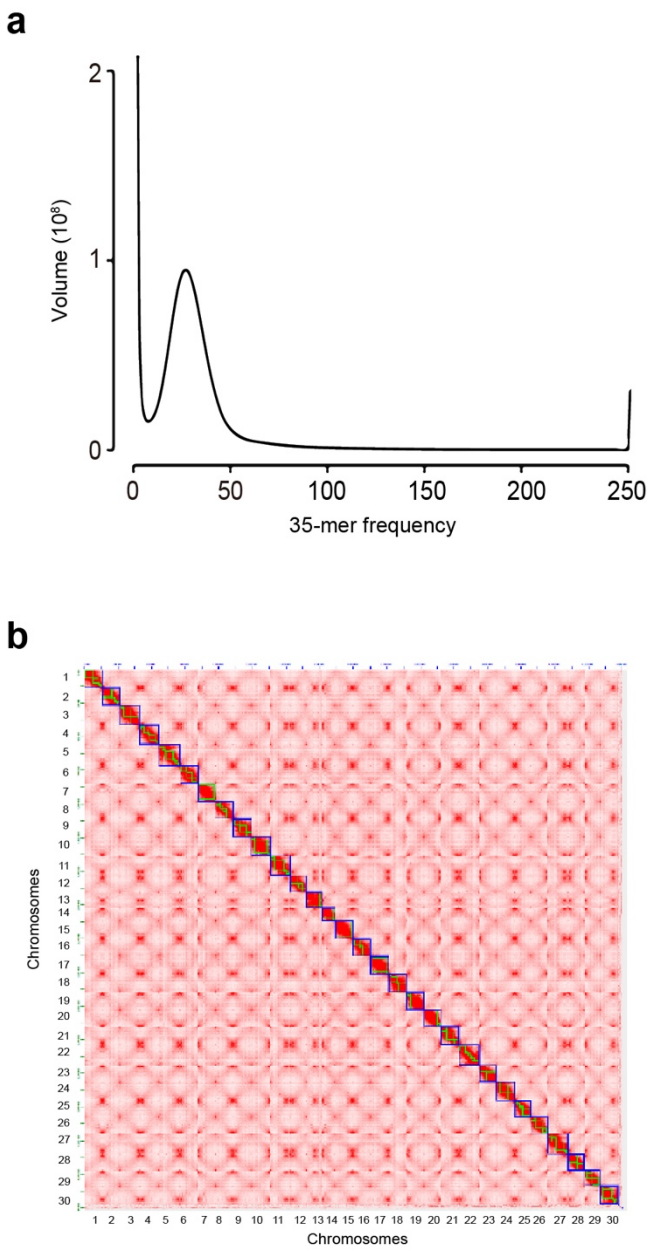
References

1. Leebens-Mack, J. H. *et al.* One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
2. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).
3. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
4. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
5. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
6. Stanke, M. *et al.* AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
7. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
8. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 1–9 (2004).
9. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
10. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
11. Lowe, T. M. & Chan, P. P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **44**, W54–W57 (2016).
12. Kalvari, I. *et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).

13. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
14. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
15. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
16. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The european molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
17. Finn, R. D., Miller, B. L., Clements, J. & Bateman, A. iPfam: a database of protein family and domain interactions found in the Protein Data Bank. *Nucleic Acids Res.* **42**, D364–D373 (2014).
18. Kang, Y. *et al.* CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).
19. Katoh, K. & Toh, H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26**, 1899–1900 (2010).
20. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
21. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
22. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
23. Han, M. V, Thomas, G. W., Lugo-Martinez, J. & Hahn, M. W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
24. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
25. Jiang, Z. *et al.* AHD2.0: an update version of Arabidopsis Hormone Database for plant systematic studies. *Nucleic Acids Res.* **39**, D1123–D1129 (2011).

26. Wolfender, J. L., Marti, G., Thomas, A. & Bertrand, S. Current approaches and challenges for the metabolite profiling of complex natural extracts. *J. Chromatogr A* **1382**, 136–164 (2015).
27. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
28. Kim, S. *et al.* PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **47**, D1102–D1109 (2019).

Supplementary Figures



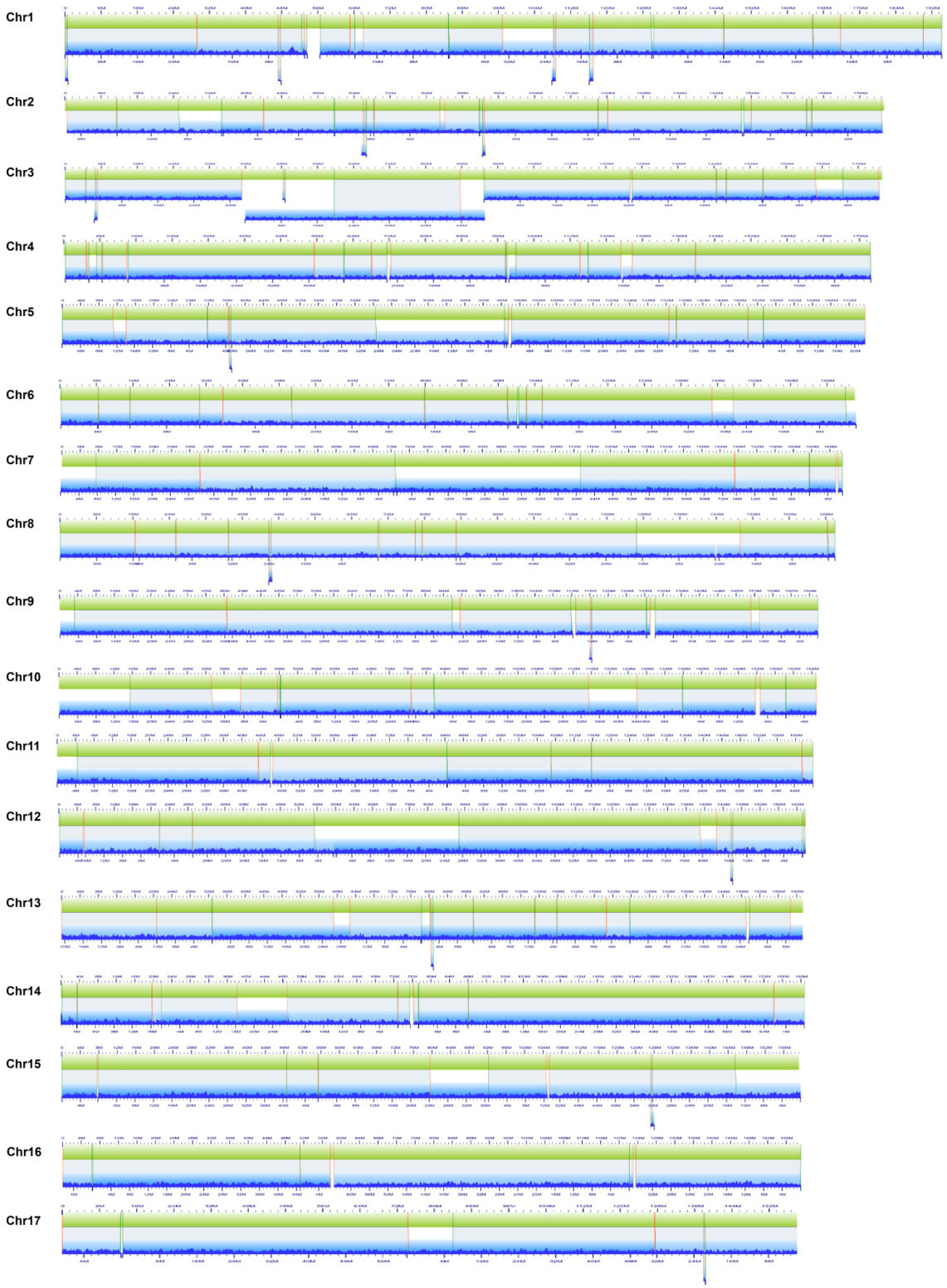
Supplementary Fig. 1 | Assembly of an *A. capillus-veneris* genome.

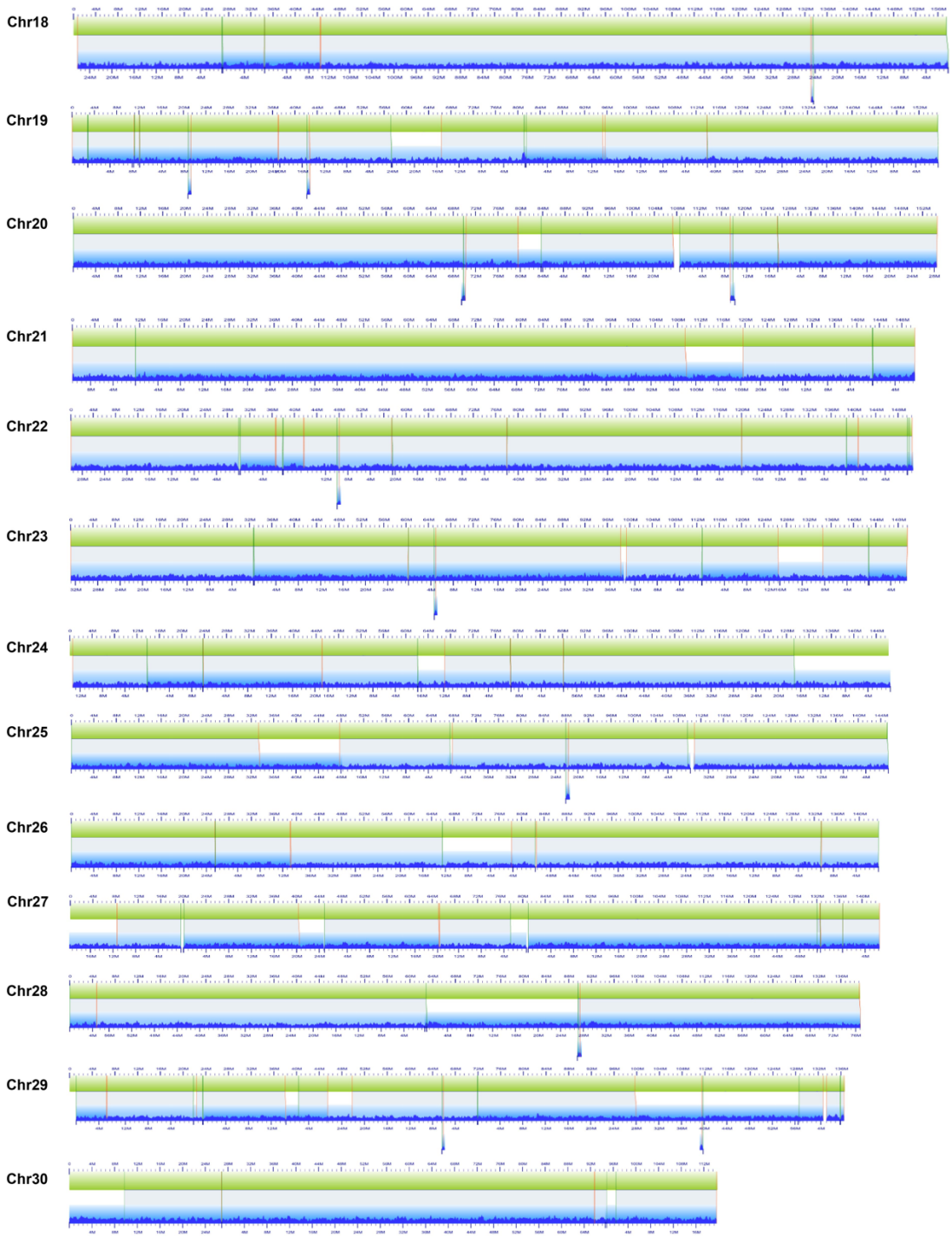
a, Genome complexity of *A. capillus-veneris* estimated using 35-mer analysis by Jellyfish. The genome size was estimated at approx. 4.95 Gb.

b, Hi-C chromatin interaction map for the 30 pseudochromosomes of the *A. capillus-veneris* genome. Bin = 500 Kb.

Supplementary Fig. 2 | Whole genome view of repeat elements distribution. The top to bottom tracks for each chromosome, red for gene density, blue for *Gypsy* superfamily density, green for *Copia* superfamily density, and orange for rRNA

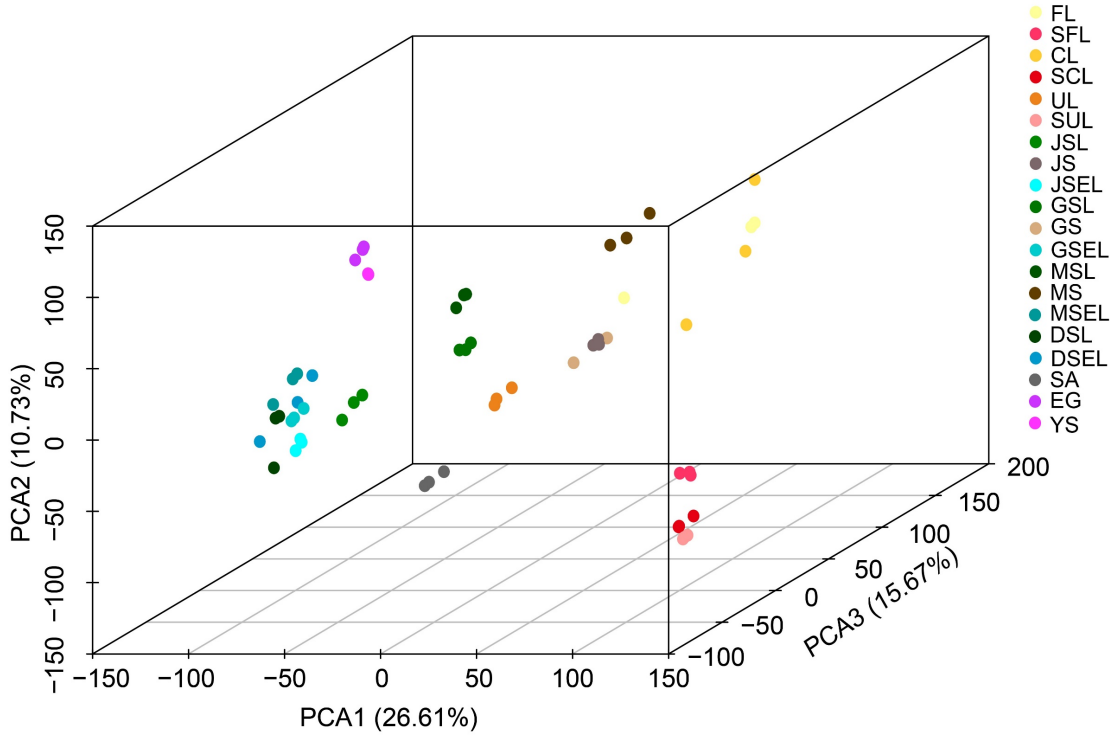
density in 500 Kb windows. The color gradation indicates the values of the density, the darker color indicates the higher density.


Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17



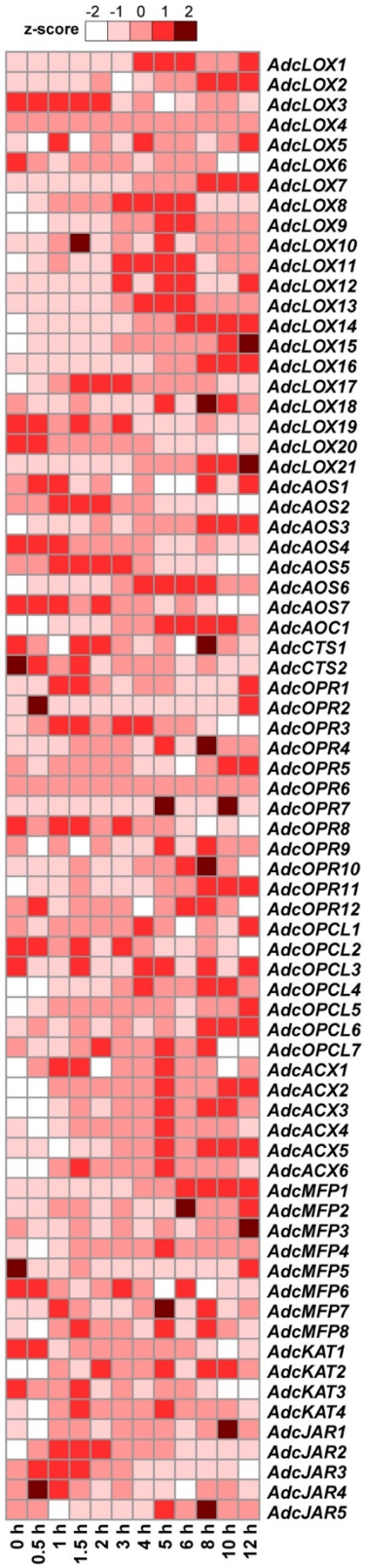
Supplementary Fig. 3 | Overview of the comparison of all pseudomolecules to genome maps.

Each green horizontal bar represents a genome map. Each blue horizontal bar represents an assembled pseudomolecule. The vertical lines represent their matching boundaries.



Supplementary Fig. 4 | Principal component analysis (PCA) of fifty-eight RNA-seq datasets from twenty samples collected from *A. capillus-veneris*.

Twenty tissue samples, covering both sporophyte and gametophyte generation of *A. capillus-veneris*, were collected for RNA-seq. All twenty samples had three replicates, except for mature sporangium (MS) and young sporophyte (YS) samples, which had two replicates, totally generating fifty-eight transcriptome datasets. Please see Extended Data Table 1 for the summary of these fifty-eight transcriptome datasets. Stem (SA, shoot apex), leaf stalk (SFL, stalk of the fist leaf; SCL, stalk of circinate leaf; SUL, stalk of unfurled leaf), frond (FL, the fist leaf; CL, circinate leaf; UL, unfurled leaf; JSEL, juvenile-sporangium-excised leaf; GSEL, green-sporangium-excised leaf; MSEL, mature-sporangium-excised leaf; DSEL, dehiscent-sporangium-excised leaf), sporangium (JS, juvenile sporangium; GS, green sporangium; MS, mature sporangium) and gametophyte (EG, embryo gametophyte; YS, young sporophyte).



Supplementary Fig. 5 | An expression heatmap of JA biosynthesis homologs in *A. capillus-veneris.*

The wounding treatment time points are indicated beneath the heatmap.

Supplementary Tables

Supplementary Table 1. Statistics of 5 cell continuous long reads (CLR) sequencing from PacBio Sequel II platform.

Library	Subreads number	Subreads base (bp)	Average subreads length (bp)	Subreads length N50 (bp)
TYD19180301-1	9,228,327	109,940,412,903	11,913.36	18,561
TYD19180301-2	9,062,643	121,239,155,066	13,377.90	20,222
TYD19180301-3	9,378,011	112,653,904,198	12,012.56	18,767
TYD19180302-1	9,147,217	118,520,938,982	12,957.05	18,682
TYD19180302-2	8,998,539	117,028,839,458	13,005.32	18,814

Supplementary Table 2. Valid BioNano sequencing used in this study.

Parameters	Valid data
Enzyme	DLE-1
Total length (Gb)	716.33
Average length (kb)	341.275
Molecule N50 (kb)	346.657
Label density (/100kb)	13.127
Raw coverage of reference (X)	141.32

Supplementary Table 3. Summary of Illumina sequencing data for *A. capillus-veneris* genome assembly.

Library type	Library insert size (bp)	Sequence type	Data (Gb)	Read length (bp)	GC (%)		Q30 base	
					Read1	Read2	Read1	Read2
WGS	500	PE250	21.07	250	41.17	41.15	93.51	86.49
	500		25.03	250	41.16	41.12	93.75	87.89
	500		4.44	250	41.16	41.12	92.56	85.61
	500		53.92	250	41.17	41.14	88.3	79.38
	500		54.54	250	41.15	41.08	87.71	78.76
	500		181.31	250	41.2	41.16	92.55	85.92
		Total	340.31					

Supplementary Table 4. Summary of the Hi-C data.

Enzyme	Type	Reads number	Valid reads rate (%)
Mbo I	Clean data	1,523,459,633	-
	Valid reads	551,738,568	36.22

Supplementary Table 5. Assembly statistics.

Assembly	Contigs	Scaffolds	Contig N50	Scaffold N50	Sum (Gb)
Canu	7,180	-	1.29 Mb	-	4.81
HERA	2,081		16.22 Mb		
BioNano map	-	285	-	32.86 Mb	4.80
BioNano map + HERA		1,653		99.90 Mb	4.83
BioNano map + HiC + HERA	-	762	-	159.32 Mb	4.83

Supplementary Table 6. Statistics of chromosomes of *A. capillus-veneris*.

Chromosome	Length (bp)	Gap count	Gap length (bp)	Gaps ratio
Chr1	194,202,920	77	997,674	0.51%
Chr2	181,470,100	60	1,726,062	0.95%
Chr3	181,400,794	34	165,187	0.09%
Chr4	178,378,862	38	297,815	0.17%
Chr5	175,538,835	30	239,791	0.14%
Chr6	174,067,070	34	796,375	0.46%
Chr7	170,764,101	25	958,805	0.56%
Chr8	170,047,294	35	270,754	0.16%
Chr9	165,893,482	43	597,442	0.36%
Chr10	165,055,666	32	413,798	0.25%
Chr11	163,754,988	21	89,423	0.05%
Chr12	161,980,389	22	194,201	0.12%
Chr13	161,428,364	25	290,220	0.18%
Chr14	160,835,734	30	217,437	0.14%
Chr15	159,320,382	27	216,994	0.14%
Chr16	159,151,851	26	274,991	0.17%
Chr17	158,492,396	24	164,698	0.10%
Chr18	157,573,187	25	426,672	0.27%
Chr19	155,384,207	19	345,379	0.22%
Chr20	154,628,473	30	43,048	0.03%
Chr21	150,368,723	15	226,700	0.15%
Chr22	150,505,716	21	160,270	0.11%
Chr23	149,639,048	41	736,276	0.49%
Chr24	146,224,257	19	208,531	0.14%
Chr25	145,199,038	26	113,858	0.08%
Chr26	143,415,753	9	798,565	0.56%
Chr27	143,092,580	26	480,472	0.34%
Chr28	139,379,317	17	218,586	0.16%
Chr29	136,679,751	37	386,765	0.28%
Chr30	114,394,066	60	158,391	0.14%
Total	**4,768,267,344**	**928**	**12,215,180**	**0.26%**

Supplementary Table 7. Statistics of repeat elements in *A. capillus-veneris* genome.

Classification	Number	Length (bp)	Percent of repeats (%)	Percent of genome (%)
Class I: Retroelement	1,717,000	1,986,250,463	48.27	41.14
LTR Retrotransposon	1,462,398	1,798,437,311	43.71	37.26
Ty1/*Copia*	344,148	427,385,079	10.39	8.85
Ty3/*Gypsy*	1,089,079	1,357,032,780	32.98	28.12
Other	29,171	14,019,452	0.34	0.29
Non-LTR Retrotransposon	254,602	187,813,152	4.56	3.88
LINE	251,919	186,741,433	4.54	3.87
SINE	2,683	1,071,719	0.03	0.01
Class II: DNA elements	258,361	208,165,955	5.06	4.29
CMC	136,355	131,874,557	3.21	2.73
hAT	41,978	21,983,828	0.53	0.45
PIF/Harbinger	4,809	2,598,785	0.06	0.05
MuLE-MuDR	7,743	5,419,435	0.13	0.11
Sola	15,619	11,788,469	0.29	0.24
Helitron	23113	18,966,878	0.46	0.39
other	28,744	15,534,003	0.38	0.32
Low complexity	29,484	1,615,958	0.04	0.03
Tandem	6,243	4,640,667	0.11	0.10
simple repeats	1,165,019	73,992,430	1.80	1.53
Small RNA	4,854	2,281,155	0.06	0.05
Unknown	3,492,312	1,837,521,588	44.66	38.07
Total repeat fraction	**6,673,273**	**4,114,468,216**	**100.00**	**85.25**

Supplementary Table 8. Nineteen transcriptome datasets for protein-coding gene prediction.

Tissue	Read Numbers	Mapping Rate
Stem apical	44,322,734	95.52%
Fist leaf	85,963,698	96.14%
Circinate leaf	49,232,036	93.25%
Unfurled leaf	44,067,166	94.04%
Juvenile sporangium	52,199,304	96.13%
Juvenile sporangium leaf	41,221,548	93.58%
Juvenile-sporangium-excised leaf	44,575,322	93.91%
Green sporangium	47,480,764	96.01%
Green sporangium leaf	55,780,806	95.98%
Green-sporangium-excised leaf	46,683,410	96.30%
Transition sporangium	43,745,288	96.35%
Transition sporangium leaf	40,715,872	95.83%
Transition-sporangium-excised leaf	46,985,826	94.45%
Mature sporangium	38,311,852	96.03%
Mature sporangium leaf	38,776,992	96.11%
Mature-sporangium-excised leaf	45,990,992	96.20%
Dehiscent sporangium leaf	89,068,992	95.45%
Dehiscent-sporangium-excised leaf	50,412,978	95.74%
Gametophyte	44,310,318	96.25%
Total	**949,845,898**	**95.44%**

Supplementary Table 9. Coding genes annotated through different databases.

Database	Number of genes	Percentage
InterProScan	20,920	64.79%
Swiss-Port	17,491	54.17%
Pfam	19,832	61.42%
EggNOG	19,832	61.42%
TAIR	21,161	65.54%
Total	**22,255**	**68.93%**

Supplementary Table 10. Completeness evaluation of genome assembly. The mapping rate results with identity rate ⩾ 95% and E-value < 1e-5 were retained.

Type	Numbers (reads/contigs)	Mapping rate (%)
EST	30,540	97.90
BioNano coverage	-	97.05
Chimera contigs	2	-
Indels (> 5 Kb)	0	-

Supplementary Table 11. Component analysis of repeat elements in the intron of *A. capillus-veneris*.

Classification	Length (bp)	Percent (%)
LTR Retrotransposon	293,493,168	52.47
LINE	98,081,634	16.94
SINE	94,295	0.02
Simple repeats	185,901	0.03
Small RNA	252,788	0.04
Unknown	129,296,254	22.33
Other	57,733,548	9.97
Total repeat fraction	579,137,588	100.00

Supplementary Table 12. Component analysis of repeat elements in the intron of *G. biloba*.

Repeat elements	Length in introns (bp)	% in introns	Length in the genome (bp)	Percent of the genome (%)
SINE	410,604	0.05%	687,956	0.01
Helitron	279,997	0.03%	29,761,942	0.30
rRNA	490,499	0.06%	4,419,060	0.04
Simple_repeat	1,533,040	0.18%	141,160,562	1.43
DNA	6,242,025	0.73%	13,406,472	0.14
LINE	9,037,992	1.06%	386,434,051	3.91
LTR	167,244,644	19.69%	5,127,893,868	51.96
Unknown	295,703,244	34.81%	1,889,095,839	19.15
Total	**480,942,045**	**56.62%**	**7,592,859,750**	**76.97**

Supplementary Table 13. The collision energies for different MRM pairs.

Name	Molecular formula	$(M-H)^-$	Fragments	Collision energies (V)
JA	$C_{12}H_{18}O_3$	209.12	59.01	-20
JA-Ile	$C_{18}H_{29}NO_4$	322.20	130.09	-27
OPDA	$C_{18}H_{28}O_3$	291.20	165.13	-28
d_5-JA	$C_{12}H_{13}D_5O_3$	214.15	62.03	-20

Supplementary Table 14. The jasmonate content in wounding-treated *A. capillus-veneris* pinnae. Results are expressed as mean ± SD, with five independent biological replications carried out for each treatment.

Time (h)	OPDA (nmol gFW^{-1})	Jasmonic Acid (JA, nmol gFW^{-1})	JA-Ile (nmol gFW^{-1})
0	1.12 ± 0.04	0.24 ± 0.18	0.06 ± 0.01
0.5	4.92 ± 0.26	1.00 ± 0.05	0.15 ± 0.00
1.0	6.88 ± 0.51	0.85 ± 0.37	0.17 ± 0.01
1.5	6.67 ± 0.21	0.76 ± 0.05	0.23 ± 0.01
2.0	6.37 ± 0.47	0.72 ± 0.02	0.25 ± 0.02
3.0	9.16 ± 0.03	1.21 ± 0.13	0.46 ± 0.02
4.0	9.69 ± 2.02	1.45 ± 0.35	0.63 ± 0.01
5.0	13.16 ± 0.28	1.68 ± 0.19	0.88 ± 0.02
6.0	15.67 ± 2.74	1.86 ± 0.24	1.09 ± 0.02
8.0	18.00 ± 1.81	1.40 ± 0.34	0.91 ± 0.02
10.0	22.00 ± 1.98	1.45 ± 0.01	0.97 ± 0.03
12.0	12.69 ± 1.63	1.75 ± 0.00	1.12 ± 0.01