# Chromosome-Scale Genome Assembly of the Resurrection Plant *Acanthochlamys bracteata* (Velloziaceae)

Zhi-Yuan Gao[1,2], Zhang-Hai Li[3], Dong-Liang Lin[1,2], and Xiao-Hua Jin[1,*]

[1]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
[3]Institute of Pharmaceutical Biology and Biotechnology, University of Marburg, Marburg, Germany

*Corresponding author: E-mail: xiaohuajin@ibcas.ac.cn.

**Accepted:** 21 June 2021

## Abstract

*Acanthochlamys bracteata* (Velloziaceae) is a resurrection plant with cold tolerance. Herein, a chromosome-level reference genome of *A. bracteata* based on Nanopore, Illumina, and Hi-C data is reported. The high-quality assembled genome was 197.97 Mb, with a scaffold N50 value of 8.64 Mb and a contig N50 value of 6.96 Mb. We annotated 23,509 protein-coding genes. Eight contracted gene families and three expanded gene families were detected. Repeat sequences accounted for approximately 28.63% of the genome. The LEA1 and Dehydrin gene families, which are involved in desiccation resistance, expanded in *A. bracteata*. We identified genes involved in chilling tolerance, *COLD1*.

**Key words:** resurrection plant, *Acanthochlamys bracteata*, cold tolerance.

## Significance

We sequenced the genome of the resurrection plant *Acanthochlamys bracteata*. Based on its phylogenetic position and ecological properties, the species is well-suited to analyze the molecular basis of cold and desiccation tolerance.

## Introduction

Drought resistance is a common trait in land plants and plants able to withstand extreme drought are called resurrection plants. Resurrection is an intriguing trait because it arose independently many times during the course of land plant evolution (Oliver et al. 2000). Plants belonging to Velloziaceae are characterized by drought resistance (Costa et al. 2017). Most members of Velloziaceae grow in tropical and subtropical regions (Mello-Silva 2005), only one species of Velloziaceae, *Acanthochlamys bracteata*, grows in a dry-hot river valley with an elevation of 2,700–3,500 m in Hengduan Mountains, western China (Kao 1987). The alpine region of dry-hot river valley in Hengduan Mountains has a special climate characterized by very dry and hot conditions during the daytime with chilling during the night (Kao 1987). Accordingly, *A. bracteata* is not only a resurrection plant (supplementary fig. S1, Supplementary Material online) but is also a cold-tolerant species. Here, we present a chromosome-scale genome assembly of *A. bracteata*, determined by a combination of long-read sequencing and Hi-C scaffolding technologies.

## Results and Discussion

### Genome Assembly and Annotation

Genome survey sequencing was performed and 67.80 G (∼278.61× coverage) of raw data were obtained. The genome of *A. bracteata* was sequenced using Nanopore (∼129× coverage) and paired-end Illumina (∼149× coverage) technologies. *A. bracteata* genome was 204.24 Mb with a k-mer size of 17. We used wtdbg2 (Ruan and Li 2020) for genome assembly with the raw data generated by PromethION (Eid et al. 2009; Goodwin et al. 2015)

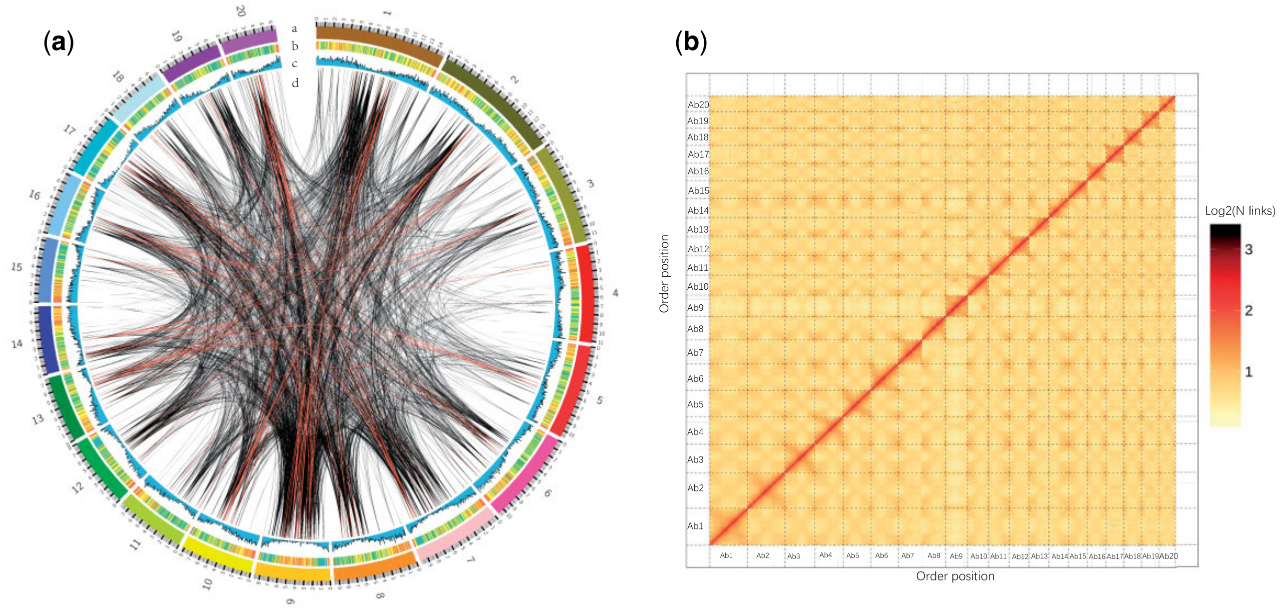

**Fig. 1.**—Genome assembly of *Acanthochlamys bracteata* (*a*) Circos map with features of the chromosomes of the *A. bracteata* genome. a, chromosomes; b, heatmap representing the average GC content per 5 kb, from blue to red means low GC content to high GC content; c, gene density per 5 kb; d, links representing the repeats in genome, repeat links above 10 kb in red, and repeat links of more than 5 kb but less than 10 kb in black. (*b*) Hi-C chromatin interaction map for the 20 pseudochromosomes of the *A. bracteata* genome.

combined with the assembly generated from Illumina data. The final assembled genome was 197.97 Mb with a contig N50 value of 6.69 Mb.

To assemble the scaffolds into pseudochromosomes, high-throughput chromosome conformation capture (Hi-C) technology was adopted to assess the chromosome-level diploid genome. *Acanthochlamys bracteata* had 38 chromosomes (2*n* = 38) (Kao et al. 1993). We obtained 40 chromosomes with lengths ranging from 6.43 to 14.91 Mb and a scaffold N50 of 8.64 Mb (contigs <100 bp were discarded) (fig. 1*b*, supplementary table S1, Supplementary Material online). The BUSCO (Simao et al. 2015) and CEGMA (Parra et al. 2007) assessment scores were 90.9% and 97.18%, respectively, suggesting that the assembly was complete and of high quality. The genome scaffold N50 was 6.96 Mb, which was much better than those of the five other sequenced genomes for resurrection species, further supporting the quality of the assembly (table 1). The GC content was 35.06%. We obtained 23,509 predicted protein-coding genes in *A. bracteata* (supplementary table S2, Supplementary Material online), comparable to the number in the genome of *Xerophyta viscosa* (Costa et al. 2017). In total, 93.3% of predicted genes were annotated and considered functional based on searches against the InterPro, Swiss-Port, NR, and KEGG databases.

**Table 1**

Properties of five resurrection species genomes

| Species | Assembly | | | | | Annotation | |
|---|---|---|---|---|---|---|---|
| | Size | Chromosomes | ContigN50 | Contigs | GC content | Genes | TEs |
| *Acanthochlamys bracteata* | 197.97 Mb | 2*n* = 40 | 6.96 Mb | 873 | 35.06% | 23,509 | 28.63% |
| *Boea hygrometrica* | 1,548 Mb | Unclear | 110 kb | 520,969 | 42.30% | 49,374 | 75.16% |
| *Oropetium thomaeum* | 236 Mb | 2*n* = 218 | 2.0 Mb | 436 | 34.86% | 28,835 | 43% (V1) |
| *Selaginella lepidophylla* | 122 Mb | 2*n* = 20 | 163 Kb | 1,149 | unknown | 27,204 | 24.61% |
| *Selaginella tamariscina* | 301 Mb | 2*n* = 20 | 407 Kb | 1,391 | 37.44% | 27,761 | 60.58% |
| *Xerophyta viscosa* | 295.5 Mb | 8× = 48 | 1.67 Mb | 1,811 contigs + 896 scaffolds | 36.51% | 25,425 | 36.50% |

Repetitive elements accounted for a particularly low proportion of the *A. bracteata* genome (i.e., 28.63%) compared with 36.5% in *X. viscosa*, which belongs to the same family as *A. bracteata*. Long terminal repeats (LTRs) were the most abundant type of interspersed repeats, occupying the majority (54.9%) of the repeat sequences, followed by DNA transposons at 12.3%. With regard to noncoding RNA, 231 microRNAs (miRNAs), 1030 transfer RNAs (tRNAs), 333 small nucleolar RNAs (snRNAs), and 432 ribosomal RNAs (rRNAs) were predicted in the *A. bracteata* genome (supplementary table S3, Supplementary Material online).

### Evolution of Gene Families

The protein-coding sequences of nine vascular plants were clustered, yielding 23,593 orthologous groups that cover 265,123 genes. From these orthologous groups, 6,016 gene families were found in all nine plant taxa and represent evolutionarily conserved ancestral gene families. In *A. bracteata*, 1,579 genes assigned to 1,020 gene families were unique. *Acanthochlamys bracteata* and *X. viscosa* diverged approximately 81.8 Ma (supplementary fig. S2, Supplementary Material online). We identified three gene families that expanded and eight gene families that contracted in the *A. bracteata* genome (supplementary tables S4 and S5, Supplementary Material online). Seven expanded genes of one gene family coding self-incompatibility protein S1 (PF05938) were identified in *A. bracteata* genome (see supplementary table S4, Supplementary Material online), suggesting that the species might be self-incompatible.

### Genes Involved in Cold and Drought Tolerance

LEA (late embryogenesis abundant) proteins accumulate late in plant seed development and contribute to the response to abiotic stress conditions (Alpert 2006). We identified LEA gene families in the genomes of *A. bracteata*, *X. viscosa*, and other species using HMM (Hidden Markov models) (Majoros et al. 2004). We found 118 putative LEAs divided into eight families (supplementary fig. S3, Supplementary Material online). There were substantially more genes encoding LEAs in Velloziaceae than in other species (supplementary tables S6 and S7, Supplementary Material online). In a comparative analysis of these nine species, we found that the LEA1 and Dehydrin families expanded in the *A. bracteata* genome and the LEA4, 5, and 6 families expanded in the *X. viscosa* genome.

The COLD1 protein is involved in chilling tolerance (Ma et al. 2015). We identified the gene encoding COLD1 in *A. bracteata* and *X. viscosa*; however, the *COLD1* homologs in these two species showed substantial divergence in structure (supplementary fig. S4, Supplementary Material online). We analyzed the motif structures of *COLD1* of *A. bracteata* and *X. viscosa* and found that all essential motifs were present (supplementary fig. S5, Supplementary Material online). Additionally, the early light-induced protein (ELIP) family plays important roles in protection against photooxidative damage under high light conditions in resurrection plants. We identified seven genes encoding ELIP proteins in *A. bracteata*, five of which were tandemly duplicated.

## Materials and Methods

### DNA Sequencing and De Novo Assembly

Samples for whole-genome sequencing were collected from a living *A. bracteata* specimen obtained from Maili Mountain in Sichuan Province (30°57′56″N, 101°7′1″E), China. High-quality genomic DNA was extracted using the Qiagen DNeasy Plant Mini Kit from the leaves of *A. bracteata* (Hilden, Germany).

We built four libraries based on the genomic DNA of *A. bracteata*. For Nanopore sequencing, genomic DNA was sheared to a size range of 15–40 kb using a Covaris g-TUBE device (Covaris). The large fragments were enriched and enzymatically repaired and converted into one ONT template library. These fragments were ligated with hairpin adapters and cleaned up. The PromethION system was used to sequence the genomic DNA of *A. bracteata*, yielding 31.45 Gb of data (~129.24×). For Illumina sequencing, three sequencing libraries were generated using Truseq Nano DNA HT Sample preparation Kit mixed with 1.5 $\mu$g DNA per library. The DNA sample was fragmented by Covaris micro-TUBE-50 device (Covaris) to a size of 350 bp, then DNA fragments were end polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. Finally, these libraries constructed above were sequenced by Illumina Hiseq2000 platform and 150 bp paired-end reads were generated. Among those three libraries, one library is Hi-C Library. Using the Illumina HiSeq2000 platform, approximately 36.35 Gb of data (~149.37×) were generated in all. The raw reads were further processed using in-house Perl scripts to remove reads containing adapters, reads containing ploy-N, and low-quality reads.

We also built another four Illumina libraries based on genomic RNA to further annotate the genome of *A. bracteata*. The genomic RNA from different tissues, such as fruits, flowers, leaves, and bracts. For RNA library construction, a total of 1.5 µg RNA was prepared, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. And libraries were generated using the NEBNextUltra RNA Library Prep Kit for Illumina (NEB, USA). Subsequently, the libraries were sequenced on an Illumina HiSeq platform to produce paired-end reads. The raw reads were further processed using in-house Perl scripts to remove reads containing adapters, reads containing ploy-N, and low-quality reads.

K-mer was set to 17 for a survey analysis, and the genome size was approximately 204.24 Mb. The clean data were assembled into contigs using SOAPdenovo2 with a k-mer of 41.

Raw data generated using PromethION were also assembled using wtdbg2 v2.5, combined with the SOAPdenovo results. We used Megablast v2.2.26 and BlobTools v1.1 (Laetsch and Blaxter 2017) to remove the contamination of genome assembly. Firstly, we used Megablast to align the assembly contigs to the NT database with parameters "megablast -p 0.8 -v 5 -b 5 -e 1e-5 -m 8 -a 12," and generated file for further analysis. Then we used BlobTools with default parameters to remove the contaminations of contigs.

We used Hi-C technology (Lieberman-Aiden et al. 2009) to assist in the assembly. The high-quality reads in the Hi-C library were mapped to the draft scaffolds using a fast and accurate short-read alignment with a Burrows–Wheeler transform (Li and Durbin 2009), and then the duplicated mapping reads and unmapped reads were removed using SAMtools v0.1.19 (Li et al. 2009). Based on chromatin interactions, contigs were clustered by using LACHESIS v1.0 (Burton et al. 2013). To assess the quality of the genome assembly, core gene annotation was performed using the BUSCO (Benchmarking Universal Single-Copy Orthologs; http://busco.ezlab.org/, last accessed August 13, 2019) method and the CEGMA (Core Eukaryotic Genes Mapping Approach; http://korflab.ucdavis.edu/datasets/cegma/, last accessed August 13, 2019) method. Illumina reads were mapped to the draft genome using BWA v0.7.17 (http://bio-bwa.sourceforge.net/, last accessed August 13, 2019), and then the integrity of the assembly and the uniformity of sequencing were evaluated (Li and Durbin 2009; Cock et al. 2010).

### Protein-Coding Gene Prediction and Functional Annotation

Augustus v3.2.3 (Stanke et al. 2004), Geneid v1.4 (Parra et al. 2000), Genescan v1.0 (Aggarwal and Ramaswamy 2002), GlimmerHMM v3.0.4 (Majoros et al. 2004), and SNAP v2013.11.29 (Korf 2004) were used for ab initio gene prediction. For homolog prediction, reference proteins were downloaded from Ensemble and NCBI. Tblastn v2.2.28 was used to align protein sequences to the genome with an *E*-value cutoff of 1e−5, and the matching proteins were accurately spliced using GeneWise v2.4.1 (Birney et al. 2004). The RNA-Seq data from different tissues of *A. bracteata* (flower, fruit, leaf, and root), and the transcriptome read assemblies were generated using Trinity v2.1.1 (Grabherr et al. 2011) for genome annotation.

Genes predicted by the three methods were merged to generate a nonredundant reference gene set using the PASA pipeline (Program to Assemble Spliced Alignment) (Haas et al. 2003) and EvidenceModeler v1.1.1 (Haas et al. 2008). For functional annotations, Blastp v2.2.28 (with a threshold *E*-value of <1e−5) was used to align the protein sequences to the Swiss-Prot and NR databases, and only the best-matched targets were collected. InterProScan v5.31-70.0 (Mitchell et al. 2015) was used to annotate motifs and domains by searching against ProDom (Corpet et al. 1998), PRINTS, PFAM, SMRT, PANTTHER, and PROSITE databases. Gene Ontology (Harris et al. 2004) annotations were obtained based on the InterPro entries (Mitchell et al. 2015). KEGG (Ogata et al. 1999) pathway analysis of the gene set was also performed and the best match for each gene was identified.

### Repeat and Noncoding RNA Annotation

A combined strategy based on homology alignment and de novo searches to identify whole-genome repeats was applied in our repeat annotation pipeline. TRF v4.09 (Benson 1999) was used to identify the tandem repeats in the genome of *A. bracteata*. RepeatMasker v4.07 (Smit et al. 2017) was employed to the repeat homolog prediction by searching against Repbase (Bao et al. 2015). The repeat regions were extracted by using an in-house script (RepeatProteinMask) with default parameters. The de novo repetitive elements were identified using LTR_FINDER v1.0.7, RepeatScout v1.0.5 (Smit et al. 2017), and RepeatModeler v1.0.3 with default parameters, and all repeat sequences with lengths >100 bp and gap "N" less than 5% were included in the raw transposable element (TE) library.

A nonredundant library was generated by combining the repeat database and our TE library. RepeatMasker v4.0.7 was used for further DNA-level repeat identification. For noncoding RNA annotations, TRNAscan-SE v1.4 (Lowe and Eddy 1997) was used to predict the tRNAs. BLAST and INFERNAL v1.1.2 (Nawrocki and Eddy 2013) were used to identify ncRNAs, including miRNAs and snRNAs, by searching against the RFAM database (Nawrocki et al. 2015).

### Phylogenetic Tree Construction

The genomes of *A. bracteata* and eight other species (*Arabidopsis thaliana*, *X. viscosa*, *Ananas comosus*, *Oryza sativa*, *Oropetium thomaeum*, *Asparagus officinalis*, *Boea hygrometrica*, and *Amborella trichopoda*) were used to identify orthologs. To remove redundancy caused by alternative splicing, only the longest predicted transcript at each gene locus was retained. To exclude putative fragmented genes, all genes encoding protein sequences shorter than 50 aa were filtered out. Then, OthoMCL v1.4 was used setting the inflation parameter to 1.5 ( Li et al. 2003) to analyze the filtered proteins from all nine species. These single-copy orthologs were used for phylogenetic tree construction. The alignment matrix was generated using MUSCLE v3.8.31 (Edgar 2004), RAxML v8.2.12 (Stamatakis 2014) (http://sco.h-its.org/exelixis/software.html, last accessed March 13, 2020) was used to build a maximum likelihood tree. Divergence times between species were calculated using MCMCtree v4.9 (http://abacus.gene.ucl.ac.uk/software/paml.html, last accessed March 13, 2020) implemented in PAML v4.9 (Yang 2007) with default settings. Based on divergence time estimates in the TimeTree database (http://www.timetree.org/, last accessed March 13, 2020), calibration points were applied. The pairwise divergence times were as follows: *O. sativa–O. thomaeum* (42–52 Ma), *A. comosus–O. thomaeum* (102–120 Ma), *A. comosus–A. officinalis* (104–125 Ma), *A. bracteata–A. comosus* (116–144 Ma), *A. thaliana–B. hygrometrica* (111–131 Ma), *O. sativa–B. hygrometrica* (111–131 Ma), *A. comosus–A. trichopoda* (173–199 Ma).

### Expansion and Contraction of Gene Families

Gene family evolution was evaluated as a stochastic birth and death process using eight plant species. Orthologous groups were constructed using OrthoMCL v1.4 (Li et al. 2003) and the number of genes of spices used in supplementary table S8, Supplementary Material online. Expansions and contractions of orthologous gene families were determined using CAFE v4.2 (Han et al. 2013) with a *P*-value cutoff of 0.05. The phylogenetic tree topology and branch lengths were input to infer the significance of changes in gene family size along each branch.

### Identification of LEA, ELIPs, and Homologs of Cold Tolerance Proteins

The Hmmsearch script from the HMMER3.1 package (Finn, et al. 2011) was used to identify LEA homologs in the genome of *A. bracteata*. Mafft v7.402 (Katoh et al. 2002) was used to build a matrix of all full-length amino acid sequences. Maximum likelihood trees were constructed using IQtree v1.6.10 (Nguyen et al. 2015) with the best-fit WAG+I+G4 model and 1000 bootstrap replicates (Costa et al. 2017). Tblastn v2.2.28 was used to identify ELIPs with *A. thaliana* ELIP1 (P93735) and ELIP2 (Q94K66) as queries and an *E*-value cutoff of 1e−5. Blastp v2.2.28 was used to identify homologs of COLD1. Hmmpress v3.1 and Hmmscan v 3.1 were also used for searches. Only the results generated by both Hmmscan v3.1 and Blastp v2.2.28 were considered homologs of COLD1.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19050201) and the National Natural Science Foundation of China (31870195 and 31670194).

## Author Contributions

X.-H.J. conceived and supervised the study. Z.-Y.G., Z.-H.L., and D.-L.L. performed the experiments. Z.-Y.G. analyzed the data. Z.-Y.G. and X.-H.J. wrote the manuscript. X.-H.J. and Z.-H.L revised the manuscript. All authors read and approved last manuscript.

## Data Availability

The data have been deposited in the NCBI GenBank database under the BioProject Number PRJNA703828.

## Literature Cited

Aggarwal G, Ramaswamy R. 2002. Ab initio gene identification: prokaryote genome annotation with GeneScan and GLIMMER. J Biosci. 27(1 Suppl 1):7–14.

Alpert P. 2006. Constraints of tolerance: why are desiccation-tolerant organisms so small or rare? J Exp Biol. 209(9):1575–1584.

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6:11.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27(2):573–580.

Birney E, Clamp M, Durbin R. 2004. GeneWise and genomewise. Genome Res. 14(5):988–995.

Burton JN, et al. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 31(12):1119–1125.

Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 38(6):1767–1771.

Corpet F, Gouzy J, Kahn D. 1998. The ProDom database of protein domain families. Nucleic Acids Res. 26(1):323–326.

Costa MD, et al. 2017. A footprint of desiccation tolerance in the genome of *Xerophyta viscosa*. Nat Plants. 3:17038.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Eid J, et al. 2009. Real-time DNA sequencing from single polymerase molecules. Science. 323(5910):133–138.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39(Web Server Issue):W29–W37.

Goodwin S, et al. 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res. 25(11):1750–1756.

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652.

Haas BJ, et al. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31(19):5654–5666.

Haas BJ, et al. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9(1):R7–R22.

Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol. 30(8):1987–1997.

Harris MA, et al.; Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32(Database Issue):D258–D261.

Kao B. 1987. Plant community and pollen morphology of *Acanthochlamys*. Acta Bot Yunnanica. 9:401–405.

Kao B, Tang Y, Guo W. 1993. A cytological study on *Acanthochlamys bracteata* P. C. Kao (Acanthochlamyaceae). J Syst Evol. 31:42–44.

Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14):3059–3066.

Korf L. 2004. Gene finding in novel genomes. Bioinformatics 14:1–9.

Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. F1000Research 6:1287–1218.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760.

Li H, et al.; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13(9):2178–2189.

Lieberman-Aiden E, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5):955–964.

Ma Y, et al. 2015. COLD1 confers chilling tolerance in rice. Cell 160(6):1209–1221.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20(16):2878–2879.

Mello-Silva RD. 2005. Morphological analysis, phylogenies and classification in Velloziaceae. Biol J Linn Soc Lond. 148:157–173.

Mitchell A, et al. 2015. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 43(Database Issue):D213–D221.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22):2933–2935.

Nawrocki EP, et al. 2015. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. 43(Database Issue):D130–D137.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.

Ogata H, et al. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27(1):29–34.

Oliver MJ, Tuba Z, Mishler BD. 2000. The evolution of vegetative desiccation tolerance in land plants. Plant Ecol. 151(1):85–100.

Parra G, Blanco E, Guigo R. 2000. GeneID in Drosophila. Genome Res. 10(4):511–515.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23(9):1061–1067.

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 17(2):155–158.

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.

Smit AFA, Hubley R, Green P. 2017. Repeat Masker Open-4.0. Available from: http://www.repeatmasker.org. Accessed September 13, 2019.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.

Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32(Web Server Issue):W309–W312.

Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.

**Associate editor:** Tanja Slotte