Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests

Pan Chen[1], Ju-Yu Lian[2,3,4*], Bin Wu[1], Hong-Lin Cao[2,3,4], Zhi-Hong Li[1] and Zheng-Feng Wang[2,3,4*]

Abstract

Objectives *Castanopsis* is the third largest genus in the Fagaceae family and is essentially tropical or subtropical in origin. The species in this genus are mainly canopy-dominant trees, and the key components of evergreen broadleaved forests play a crucial role in the maintenance of local biodiversity. *Castanopsis chinensis*, distributed from South China to Vietnam, is a representative species. It currently suffers from a high disturbance of human activity and climate change. Here, we present its assembled genome to facilitate its preliminary conservation and breeding on the genome level.

Data description The *C. chinensis* genome was assembled and annotated by Nanopore and MGI whole-genome sequencing and RNA-seq reads using leaf tissues. The assembly was 888,699,661 bp in length, consisting of 133 contigs and a contig N50 of 23,395,510 bp. A completeness assessment of the assembly with Benchmarking Universal Single-Copy Orthologs (BUSCO) indicated a score of 98.3%. Repetitive elements comprised 471,006,885 bp, accounting for 55.9% of the assembled sequences. A total of 51,406 genes that coded for 54,310 proteins were predicted. Multiple databases were used to functionally annotate the protein sequences.

Keywords Gene structures, Genome feature, Genome annotation, Repetitive elements

*Correspondence:
Ju-Yu Lian
lianjy@scib.ac.cn
Zheng-Feng Wang
wzf@scib.ac.cn
[1]Guangdong Forestry Survey and Planning Institute, Guangzhou 510520, China
[2]Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
[3]Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
[4]South China National Botanical Garden, Guangzhou 510650, China

Objective

Castanopsis is the third largest genus in the Fagaceae family [1]. It includes about 120–130 species in the genus [1–3]. Fossil evidence indicates that *Castanopsis* is widely distributed in both the Northern and Southern Hemisphere through the Eocene to the Pliocene in history [2, 4], but currently, it is mainly distributed in the subtropics and tropics of East and Southeast Asia [2, 4]. *Castanopsis* species are mainly canopy-dominant trees and can grow up to 25–40 m in height [5]. Therefore, they are the main components of evergreen broadleaved forests, safeguarding local biodiversity [2, 6]. *Castanopsis* species are good timber trees, and their seeds are edible [2, 3, 7]. They also contain many polyphenols and are used in traditional medicines [3]. Climate change is the main threat to

Castanopsis species due to its restricted migration ability [3, 7]. By studying 32 dominant *Castanopsis* species in East Asia that grow from 5°N to 38°N, it has been predicted that their present high richness distribution range will be reduced by 94.5%, on average, by 2070 [3].

China is a central region of the *Castanopsis* distribution and includes approximately 60 species [3], half of which are endemic to China [1]. *Castanopsis chinensis* is distributed from South China to Vietnam. It is a pioneer-dominant canopy tree in evergreen broad-leaved forests and plays a key role in ecosystems [3, 8]. As a fast-growing and soil erosion-controlling species, it is also widely used in reforestation [7]. Because the *C. chinensis* distribution area suffers from disturbances of high human activity, most forests have been converted or degraded. Therefore, we present here the *C. chinensis* genome to better understand its evolution and adaptation and enhance its conservation, management and utility in the future.

Data description

We collected leaf samples of *C. chinensis* from a individual planted in the South China Botanical Garden, Guangzhou China. To perform genome assembly and annotation, three sequencing libraries were constructed from genomic DNA or RNA extracted from the leaf tissues. The first library was constructed by long read whole genome sequencing using a Nanopore PromethION sequencer, which generated about 139.5 GB of data (Data files 1–3) [9–11]. The second was generated by short read whole genome sequencing using an MGI DNBSEQ-T7 sequencer, which generated about 149.6 GB data (Data file 4) [12], and the third was generated by RNA sequencing (RNA-seq) using an MGI DNBSEQ-T7 sequencer, which generated about 29.7 GB data (Data file 5) [13]. All sequencing using the MGI platform was applied using 150 bp paired-end mode.

The genome size of *C. chinensis* was estimated by KmerGenie v1.7044 [14] (under the parameter of "-k 141 --diploid") and GenomeScope 2.0 [15] (under the *k*-mer of 21) with short whole genome reads, which were trimmed using Sickle v1.33 [16] with the parameter "-q 30 -l 80". The genome sizes estimated by KmerGenie and GenomeScope were 1,143,475,699 and 744,772,109 bp, respectively. Nanopore long reads were quality trimmed by Porchop v0.2.4 [17] and ontbc v1.1 [18] to remove adapters, and the reads had quality scores<7 and lengths<5000 bp. NextDenovo v2.3.1 [19] was then used to assemble the genome with the filtered reads. Pseudo-haploid [20] and Purge_Dups v1.2.6 [21] (running twice) were used to remove duplicated contigs. Racon v1.5.0 [22], hapo-G v1.3.2 [23] and polypolish v0.5.0 [24] were further used to improve the assembly, in which racon and hapo-G were each run twice. The final assembly consisted of 133 contigs of 888,699,661 bp in length and a contig N50 of 23,395,510 bp (Data file 6) [25]. BUSCO v5.4.6 [26] assessed 98.3% completeness using the Eudicots odb10-2020-09-10 database (Data file 7) [27].

RED v2.0 [28] and EDTA v2.1.0 [29] were used to predict repetitive elements in the assembly, which revealed 400,198,509 (Data file 8) [30] and 410,582,904 bp (Data file 9) [31] of the sequences, respectively. Combining the RED and EDTA results resulted in 496,557,194 bp sequences (Data file 10) [32], accounting for 55.9% of the genome assembly. After soft-masking the repetitive elements in the assembly, braker v.2.0 [33] was used to predict the gene structures. Braker is an automated gene annotation pipeline that uses transcriptome data and reference protein sequences (Data file 11) [34]. The braker results were then inputted into Funannotate pipeline v1.8.16 [35] to obtain integrated gene sets. The pipeline included "train", "predict" and "update" steps. In all steps, the parameter of "--max_intronlen 1000000" was used. In the "predict" step, the parameters of "--busco_seed_species arabidopsis --organism other --busco_db embryophyta" were added. The Funannotate pipeline finally produced 51,406 genes that coded 54,310 protein sequences (Data files 12–14) [36–38]. After gene structure prediction, functional annotation of the genes was performed using the "funannotate annotate" command in the Funannotate pipeline (Data files 15–16) [39, 40].

Limitations

Although the genome size estimators of KmerGenie and GenomeScope were highly discrepant, yielding 1,143,475,699 and 744,772,109 bp, the final assembly size of 888,699,661 bp was comparable to previously reported genome sizes for *Castanopsis* species, including 878.6 Mb for *C. tibetana* [1] and 882.6 Mb for *C. hystrix* [41], and higher than 785.5 Mb for *C. mollissima* [42]. It has been reported that accurate genome size estimation with short reads is challenging [43, 44]. Therefore, long HiFi sequencing data may be further needed to obtain an accurate size estimation [44]. Due to their long length and high accuracy, HiFi date could character genome size reliably both in small and large *k*-mers [43, 45–48], which help determining the true result when the discrepancy happening in genome size estimation by short reads [43].

Currently, the assembled genome in this report is still fragmented. Therefore, it is not suitable for complete genome structure analysis, hindering the complex regions digging in its conservation and breeding [49–51]. Further high-quality genome assemblies (preferably complete and gapless) using ultra-long read, Hi-C, and other sequencing technologies are needed [52].

Table 1 Overview of all data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Raw long whole genome sequencing reads	Fastq file (.fastq)	NCBI Sequence Read Archive, https://identifiers.org/ncbi/insdc.sra:SRR26081294 [9]
Data file 2	Raw long whole genome sequencing reads	Fastq file (.fastq)	NCBI Sequence Read Archive, https://identifiers.org/ncbi/insdc.sra:SRR26081295 [10]
Data file 3	Raw long whole genome sequencing reads	Fastq file (.fastq)	NCBI Sequence Read Archive, https://identifiers.org/ncbi/insdc.sra:SRR26081296 [11]
Data file 4	Raw short whole genome sequencing reads	Fastq file (.fastq)	NCBI Sequence Read Archive, https://identifiers.org/ncbi/insdc.sra:SRR26081292 [12]
Data file 5	Raw RNA reads of leaf tissues	Fastq file (.fastq)	NCBI Sequence Read Archive, https://identifiers.org/ncbi/insdc.sra:SRR26075029 [13]
Data file 6	Assembled genome	Fasta file (.fasta)	NCBI Nucleotide, https://identifiers.org/nucleotide:JAVQMG000000000.1 [25]
Data file 7	BUSCO assessment of the assembly	Text (.txt)	Figshare, https://doi.org/10.6084/m9.figshare.24417850.v2 [27]
Data file 8	Repetitive sequences predicted by RED	Text file (.bed)	Figshare, https://doi.org/10.6084/m9.figshare.24417889.v1 [30]
Data file 9	Repetitive sequences predicted by EDTA	Gff3 file (.gff3)	Figshare, https://doi.org/10.6084/m9.figshare.24417895.v1 [31]
Data file 10	Repetitive sequences combined by RED and EDTA	Text file (.bed)	Figshare, https://doi.org/10.6084/m9.figshare.24417910.v1 [32]
Data file 11	Table 1 Species with their protein sequences used for gene prediction	Table (.xlsx)	Figshare, https://doi.org/10.6084/m9.figshare.24417970.v1 [34]
Data file 12	Predicted gene	Gff3 file (.gff3)	Figshare, https://doi.org/10.6084/m9.figshare.24417985.v1 [36]
Data file 13	Predicted genes - nucleotide sequences	Fasta file (.fasta)	Figshare, https://doi.org/10.6084/m9.figshare.24417991.v1 [37]
Data file 14	Predicted genes - translated sequences	Fasta file (.fasta)	Figshare, https://doi.org/10.6084/m9.figshare.24418003.v1 [38]
Data file 15	Gene annotation using GO, Pfam, interPro, Uni-Prot, dbCAN, MEROPS and SignalP databases	Text (.txt)	Figshare, https://doi.org/10.6084/m9.figshare.24418012.v1 [39]
Data file 16	Gene annotation from eggNOG-mapper analysis	Text (.txt)	Figshare, https://doi.org/10.6084/m9.figshare.24418015.v1 [40]

Data Availability
Raw sequenced reads have been uploaded to the NCBI Sequence Read Archive under accession number SRR26081294, SRR26081295 and SRR26081296 for long whole genome sequencing reads [9–11], SRR26081292 for short whole genome sequencing reads [12], SRR26075029 for RNA-seq reads [13], and JAVQMG000000000 for the assembled genome [25]. Please further see Table 1 for details and references [30–32, 34, 36–40] of the results of the annotations submitted to figshare.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Sun Y, Guo J, Zeng X, Chen R, Feng Y, Chen S, Yang K. Chromosome-scale genome assembly of *Castanopsis Tibetana* provides a powerful comparative framework to study the evolution and adaptation of Fagaceae trees. Mol Ecol Resour. 2022;22:1178–89. https://doi.org/10.1111/1755-0998.13539.
2. Tang CQ, Matsui T, Ohashi H, Nualart N, Herrando-Moraira S, Dong YF, Grote PJ, Ngoc NV, Sam HV, Li S, Han P-B, Shen L-Q, Huang D-S, Peng M-C, Wang C-Y, Li X-S, Yan H-Z, Zhu M-Y, Lu X, Wen J-R, Yao S-Q, Du M-R, Shi Y-C, Xiao S-L, Zeng J-L, Wang H-C, López-Pujol J. Identifying long-term stable refugia for dominant *Castanopsis* species of evergreen broad-leaved forests in East Asia: a tool for ensuring their conservation. Biol Conserv. 2022;273:109663. https://doi.org/10.1016/j.biocon.2022.109663.
3. Wang Y-F, Lin P, Huang Y-L, He R-J, Yang B-Y, Liu Z-B. Isolation of two new phenolic glycosides from *Castanopsis Chinensis* Hance by combined multistep CC and HSCCC separation and evaluation of their antioxidant activity. Molecules. 2023;28(8):3331. https://doi.org/10.3390/molecules28083331.
4. Wilf P, Nixon KC, Gandolfo MA, Cúneo NR. Eocene Fagaceae from Patagonia and Gondwanan legacy in Asian rainforests. Science. 2019;364:6444. https://doi.org/10.1126/science.aaw513.
5. Wang Z, Wu X, Sun B, Yin S, Quan C, Shi G. First fossil record of *Castanopsis* (Fagaceae) from the middle Miocene Fotan Group of Fujian, southeastern China. Rev Palaeobot Palyno. 2022;305:104729. https://doi.org/10.1016/j.revpalbo.2022.104729.
6. Ashton P, Zhu H. The tropical-subtropical evergreen forest transition in East Asia: an exploration. Plant Divers. 2020;42(4):255–80.

7.   Cheuk ML, Fischer GA. The impact of climate change on the distribution of *Castanopsis* (Fagaceae) species in south China and Indo-China region. Glob Ecol Conserv. 2021;26:e01388. https://doi.org/10.1016/j.gecco.2020.e01388.

8.   Wang Z-F, Lian J-Y, Ye W-H, Cao H-L, Wang Z-M. The spatial genetic pattern of *Castanopsis chinensis* in a large forest plot with complex topography. For Ecol Manag. 2014;318:318–25. https://doi.org/10.1016/j.foreco.2014.01.042.

9.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broad-leaved evergreen forest. NCBI Seq Read Archive. 2023. https://identifiers.org/ncbi/insdc.sra:SRR26081294.

10.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. 2023. NCBI Sequence Read Archive. https://identifiers.org/ncbi/insdc.sra:SRR26081295.

11.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broad-leaved evergreen forests. NCBI Seq Read Archive. 2023. https://identifiers.org/ncbi/insdc.sra:SRR26081296.

12.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broad-leaved evergreen forests. NCBI Seq Read Archive. 2023. https://identifiers.org/ncbi/insdc.sra:SRR26081292.

13.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broad-leaved evergreen forests. NCBI Seq Read Archive. 2023. https://identifiers.org/ncbi/insdc.sra:SRR26075029.

14.   Chikhi R, Medvedev P. Informed and automated *k*-mer size selection for genome assembly. Bioinformatics. 2014;30:31–7. https://doi.org/10.1093/bioinformatics/btt310.

15.   Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33:2202–4. https://doi.org/10.1093/bioinformatics/btx153.

16.   Joshi NA, Fass JN, Sickle. A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. (2011) Available at: https://github.com/najoshi/sickle. Accessed 24 Aug 2022.

17.   Porchop v0.2.4. Available at: https://github.com/rrwick/Porechop. Accessed 4 November 2022.

18.   Ontbc v1.1. : Pipeline for oxford nanopore barcoding. Available at: https://github.com/FlyPythons/ontbc. Accessed 26 Aug 2022.

19.   NextDenovo v2. 3.1: Fast and accurate de novo assembler for long reads. Available at: https://github.com/Nextomics/NextDenovo. Accessed 24 January 2023.

20.   Pseudohaploid. Create a pseudohaploid assembly from a partially resolved diploid assembly. Available at:https://github.com/schatzlab/pseudohaploid. Accessed 26 January 2023.

21.   Guan DF, McCarthy SA, Wood J, Howe K, Wang YD. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36:2896–8. https://doi.org/10.1093/bioinformatics/btaa025.

22.   Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737–46. https://doi.org/10.1101/gr.214270.116.

23.   Aury JM, Istace B. Hapo-G, haplotype-aware polishing of genome assemblies with accurate reads. NAR Genom Bioinform. 2021;3(2):lqab034. https://doi.org/10.1093/nargab/lqab034.

24.   Wick RR, Holt KE, Polypolish. Short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol. 2022;18(1):e1009802. https://doi.org/10.1371/journal.pcbi.1009802.

25.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broad-leaved evergreen forests. NCBI Nucleotide. 2023. https://identifiers.org/nucleotide:JAVQMG000000000.1.

26.   Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol. 2019;1962:227–45. https://doi.org/10.1007/978-1-4939-9173-0_14.

27.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24417850.v2.

28.   Girgis HZ. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. BMC Bioinform. 2015;16:227. https://doi.org/10.1186/s12859-015-0654-5.

29.   Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, Jiang N, Hirsch CN, Hufford MB. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20:275. https://doi.org/10.1186/s13059-019-1905-y.

30.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24417889.v1.

31.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24417895.v1.

32.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24417910.v1.

33.   Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP + and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 2021;3(1):lqaa108. https://doi.org/10.1093/nargab/lqaa108.

34.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24417970.v1.

35.   Palmer J, Funannotate. Eukaryotic Genome Annotation Pipeline. Available at:https://github.com/nextgenusfs/funannotate. Accessed 20 Sep 2022.

36.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24417985.v1.

37.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24417991.v1.

38.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24418003.v1.

39.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24418012.v1.

40.   Chen P, Lian J-Y, Wu B, Cao H-L, Li Z-H, Wang Z-F. Draft genome of *Castanopsis chinensis*, a dominant species safeguarding biodiversity in subtropical broadleaved evergreen forests. Figshare. 2023. https://doi.org/10.6084/m9.figshare.24418015.v1.

41.   Huang W-C, Liao B, Liu H, Liang Y-Y, Chen X-Y, Wang B, Xia H. A chromosome-scale genome assembly of *Castanopsis hystrix* provides new insights into the evolution and adaptation of Fagaceae species. Front. Plant Sci. 2023;14:1174972. https://doi.org/10.3389/fpls.2023.1174972.

42.   Xing Y, Liu Y, Zhang Q, Nie X, Sun Y, Zhang Z, Li H, Fang K, Wang G, Huang H, Bisseling T, Cao Q, Qin L. Hybrid de novo genome assembly of Chinese chestnut (*Castanea mollissima*), GigaScience. 2019;8(9):giz112. https://doi.org/10.1093/gigascience/giz112.

43.   Wang Z-F, Rouard M, Droc G, Heslop-Harrison P, Ge X-J. Genome assembly of *Musa beccarii* shows extensive chromosomal rearrangements and genome expansion during evolution of Musaceae genomes. GigaScience. 2023;12:giad005. https://doi.org/10.1093/gigascience/giad005.

44.   Gao Y, Liao H-B, Liu T-H, Wu J-M, Wang Z-F, Cao H-L. Draft genome and transcriptome of *Nepenthes mirabilis*, a carnivorous plant in China. BMC Genom Data. 2023;24:21. https://doi.org/10.1186/s12863-023-01126-5.

45.   Hai DM, Yen DT, Liem PT, Tam BM, Huong DTT, Hang BTB, Hieu DQ, Garigliany M-M, Coppieters W, Kestemont P, Phuong NT, Farnir F. A high-quality genome assembly of striped catfish (*Pangasianodon Hypophthalmus*) based on highly accurate long-read HiFi sequencing data. Genes. 2022;13:923. https://doi.org/10.3390/genes13050923.

46.   Navrátilová P, Toegelová H, Tulpová Z, Kuo Y-T, Stein N, Doležel J, Houben A, Šimková H, Mascher M. Prospects of telomere-to-telomere assembly in barley: analysis of sequence gaps in the MorexV3 reference genome. Plant Biotechnol J. 2022;20:1373–86. https://doi.org/10.1111/pbi.13816.

47. Jayakodi M, Golicz AA, Kreplak J, Fechete LI, Angra D, Bednář P, Bornhofen E, Zhang H, Boussageon R, Kaur S, Cheung K, Čížková J, Gundlach H, Hallab A, Imbert B, Keeble-Gagnère G, Koblížková A, Kobrlová L, Krejčí P, Mouritzen TW, Neumann P, Nadzieja M, Nielsen LK, Novák P, Orabi J, Padmarasu S, Robertson-Shersby-Harvie T, Robledillo LÁ, Schiemann A, Tanskanen J, Törönen P, Warsame AO, Wittenberg AH, Himmelbach A, Aubert G, Courty P, Doležel J, Holm LU, Janss LL, Khazaei H, Macas J, Mascher M, Smýkal P, Snowdon RJ, Stein N, Stoddard FL, Tayeh N, Torres AM, Usadel B, Schubert I, O'Sullivan DM, Schulman AH, Andersen SU. The giant diploid faba genome unlocks variation in a global protein crop. Nature. 2023;615:652–9. https://doi.org/10.1038/s41586-023-05791-5.
48. Tolman ER, Beatty CD, Bush J, Kohli M, Moreno CM, Ware JL, Weber KS, Khan R, Maheshwari C, Weisz D, Dudchenko O, Aiden EL, Frandsen PB. A chromosome-length assembly of the black petaltail (*Tanypteryx hageni*) dragonfly. Genome Biol Evol. 2023;15(3):evad024. https://doi.org/10.1093/gbe/evad024.
49. European Reference Genome Atlas (ERGA) Consortium. The era of reference genomes in conservation genomics. Trends Ecol Evol. 2022;37(3):197–202. https://doi.org/10.1016/j.tree.2021.11.008.
50. Luo H, Jiang X, Li B, Wu J, Shen J, Xu Z, Zhou X, Hou M, Huang Z, Ou X, Xu L. A high-quality genome assembly highlights the evolutionary history of the great bustard (*Otis tarda*, Otidiformes). Commun Biol. 2023;6:746. https://doi.org/10.1038/s42003-023-05137-x.
51. Shi X, Cao S, Wang X, Huang S, Wang Y, Liu Z, Liu W, Leng X, Peng Y, Wang N, Wang Y, Ma Z, Xu X, Zhang F, Xue H, Zhong H, Wang Y, Zhang K, Velt A, Avia K, Holtgräwe D, Grimplet J, Matus JT, Ware D, Wu X, Wang H, Liu C, Fang Y, Rustenholz C, Cheng Z, Xiao H, Zhou Y. The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. Hort Res. 2023;10(5):uhad061. https://doi.org/10.1093/hr/uhad061.
52. Luo J, Wei Y, Lyu M, Wu Z, Liu X, Luo H, Yan C. A comprehensive review of scaffolding methods in genome assembly. Brief Bioinform. 2021;22(5):bbab033. https://doi.org/10.1093/bib/bbab033.

Publisher's Note