

DATA NOTE

Open Access



# The de novo, chromosome-level genome assembly of the sweet chestnut (*Castanea sativa* Mill.) Cv. Marrone Di Chiusa Pesio

Luca Bianco<sup>1†</sup>, Paolo Fontana<sup>1†</sup>, Alexis Marchesini<sup>2,3†</sup>, Sara Torre<sup>4</sup>, Mirko Moser<sup>1</sup>, Stefano Piazza<sup>1</sup>, Sara Alessandri<sup>5</sup>, Vera Pavese<sup>6</sup>, Paola Pollegioni<sup>2,3</sup>, Cristiano Vernesi<sup>1</sup>, Mickael Malnoy<sup>1</sup>, Daniela Torello Marinoni<sup>6</sup>, Sergio Murolo<sup>7</sup>, Luca Dondini<sup>5</sup>, Claudia Mattioni<sup>2,3</sup>, Roberto Botta<sup>6</sup>, Federico Sebastiani<sup>4</sup>, Diego Micheletti<sup>1</sup> and Luisa Palmieri<sup>1\*</sup>

## Abstract

**Objectives** The sweet chestnut *Castanea sativa* Mill. is the only native *Castanea* species in Europe, and it is a tree of high economic value that provides appreciated fruits and valuable wood. In this study, we assembled a high-quality nuclear genome of the ancient Italian chestnut variety 'Marrone di Chiusa Pesio' using a combination of Oxford Nanopore Technologies long reads, whole-genome and Omni-C Illumina short reads.

**Data description** The genome was assembled into 238 scaffolds with an N50 size of 21.8 Mb and an N80 size of 7.1 Mb for a total assembled sequence of 750 Mb. The BUSCO assessment revealed that 98.6% of the genome matched the embryophyte dataset, highlighting good completeness of the genetic space. After chromosome-level scaffolding, 12 chromosomes with a total length of 715.8 and 713.0 Mb were constructed for haplotype 1 and haplotype 2, respectively. The repetitive elements represented 37.3% and 37.4% of the total assembled genome in haplotype 1 and haplotype 2, respectively. A total of 57,653 and 58,146 genes were predicted in the two haplotypes, and approximately 73% of the genes were functionally annotated using the EggNOG-mapper. The assembled genome will be a valuable resource and reference for future chestnut breeding and genetic improvement.

**Keywords** European chestnut, Genome sequence assembly, Chromosome-scale

<sup>†</sup>Luca Bianco, Paolo Fontana and Alexis Marchesini contributed equally to this work.

\*Correspondence:

Luisa Palmieri

luisa.palmieri@fmach.it

<sup>1</sup>Research and Innovation Center, Edmund Mach Foundation, via Mach 1, San Michele all'Adige, TN 38098, Italy

<sup>2</sup>Research Institute on Terrestrial Ecosystem, National Research Council, Via Marconi 2, Porano, TR 05010, Italy

<sup>3</sup>NBFC, National Biodiversity Future Center, Palermo, Italy

<sup>4</sup>Institute for Sustainable Plant Protection, National Research Council, Via Madonna del Piano 10, 50019, Sesto Fiorentino FI, Italy

<sup>5</sup>Dept. of Agricultural and Food Sciences, University of Bologna, Via Zamboni 33, Bologna, BO 40126, Italy

<sup>6</sup>Dept. of Agricultural, Forest and Food Sci, University of Turin, L.go P. Braccini 2, Grugliasco, TO 10095, Italy

<sup>7</sup>Dep. of Agricultural, Food and Env.Sci, Marche Polytechnic University, via Breccia Bianche, Ancona, AN 60131, Italy



## Objective

*Castanea* Mill. ( $2n=2x=24$ ) is a genus of broadleaved trees and shrubs of the Fagaceae family that includes seven species (although the taxonomic identity of some entities is still debated [1, 2]) that are native to temperate deciduous forests of the Northern Hemisphere. Among these, three species are cultivated for fruit: Japanese chestnut (*C. crenata* Sieb. et Zucc), Chinese chestnut (*C. mollissima* Bl.), and European chestnut (*C. sativa* Mill.). European chestnut, also known as sweet chestnut, is the only European species of the genus and is native to central-southern Europe (northern Iberian Peninsula, southern France, central-northern Italy, southern Balkan Peninsula) and Asia Minor (western and northern Turkey, Caucasus); however, it has been widely planted and cultivated outside its natural range in temperate regions worldwide (e.g., South and North America and Australia). Sweet chestnut is a species of remarkable ecological and economic importance. In addition to being the dominant tree in its native range, mesophilous, broad-leaved forests have been cultivated for millennia for timber production (coppice and high forest) and fruit production (traditional orchards), providing a broad range of secondary products and ecosystem services [3].

The sweet chestnut Marrone genotype was selected for its exceptional qualities, such as its above-average fruit weight (maximum 70 fruits/kg), mono-embryonic nuts, and thin and easy-to-remove epispem (cuticle), which is not deep in the cotyledons and their floury paste or sugary and is consistent [4, 5]. Reference genomes have been published for three *Castanea* species so far. In 2020, a scaffold-level genome assembly of *C. mollissima* cultivar N11-1, generated using PacBio sequencing technology, was published (GenBank assembly accession: GCA\_014183005.1), and in the same year the scaffold-level genome for *C. mollissima* cultivar “Vanuxem” was assembled [6]. In 2021, contig- and scaffold-level genomes were generated from PacBio sequencing also for *C. crenata* (GenBank: GCA\_019972055.1 and GCA\_020976635.1). In 2022, a chromosome-level genome assembly was published for *C. crenata* by combining Nanopore long reads and Hi-C sequencing [7]. Two chromosome-scale and haplotype-resolved reference genome assemblies were then recently generated for *C. mollissima*, “Mahogany” and “Nanking” cultivar (HudsonAlpha Institute for Biotechnology; [http://phytozome.jgi.doe.gov/info/CmollissimaMahoganyHAP2\\_v1\\_1](http://phytozome.jgi.doe.gov/info/CmollissimaMahoganyHAP2_v1_1)).

Here, we describe a chromosome-scale de novo genome assembly of *Castanea sativa* Mill., cv. “Marrone di Chiusa Pesio”, which, to our knowledge, is the first reported *C. sativa* genome assembly. We believe this study will provide important resources for better investigating the evolutionary history and domestication

process of this species and elucidating the genetic basis of resistance to diseases and environmental stressors.

## Data description

Fresh, young leaves were collected from a single true-to-type plant, ‘Marrone di Chiusa Pesio’, which was provided by the Chestnut R&D Center Piemonte (<https://centro-castanicoltura.org/en/>). DNA extraction was performed using Macherey Nagel’s NucleoSpin Plant II Midi following the manufacturer’s protocol.

Illumina libraries were constructed from the genomic DNA following the Illumina TruSeq kit protocol and sequenced (PE150) by Novogene, yielding 89 Gb of data. Genomic DNA was also sequenced using the ONT Minion device with Flowcell version R9.4.1. Additionally, a Hi-C library was prepared with the Omni-C Kit from Dovetail Genomics following the manufacturer’s protocol with minor adjustments (Supplementary Appendix and Table 1).

K-mer analysis of the Illumina reads with a kmer size of 23 using GenomeScope [8] indicated an estimated genome size of 654 Mbps.

The ONT reads were assembled using the NextDenovo assembler v2.5.0 [9], and the assembled sequence was polished with NextPolish v1.4.0 [10], resulting in 238 scaffolds with an N50 value of 21.8 Mbps and an N80 value of 7.1 Mbps, for a total of 750 Mbps. The scaffold-to-chromosome ratio was 19.83. Chromosome-level scaffolding was performed with Omni-C data with standard parameters (<https://omni-c.readthedocs.io/en/latest>). A manual inspection of the contact maps was conducted without highlighting any issues.

The genome was anchored using two published genetic maps for *Castanea sativa*: ‘Bouche de Betizac’ and ‘Madonna’ [11]. ALLMAPS [12] was used with standard parameters after the SNPs were mapped to the assembled genome and the alignments were filtered (98% identity and coverage of the probe sequence and uniqueness of each haplotype). The mapey plots are reported in the Supplementary materials.

The two haplotypes were reconstructed by phasing (with ONT reads) the structural variants identified by Illumina reads with WhatsHap software v.1.0 [13] and the BCFtools v.1.7 consensus command [14].

The final anchored sequence contained 715 Mbps (Haplotype 1) and 713 Mbps (Haplotype 2) (Sup. Table 2).

BUSCO (v. 5.2.2) [15] analysis indicated high genome assembly completeness (98.6%). Furthermore, 97% of the filtered Illumina reads were aligned to the genome assembly using BWA v.0.7.17 [16], with 93% being properly paired (see Supplementary materials).

The LTR assembly index (LAI) was computed with LTR\_Retrieve v2.9.7 [17] for both haplotypes, with

**Table 1** Overview of the data files and datasets

Label	Name of data file/dataset	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Genome Haplotype1	Fasta file (.fasta)	<a href="https://treegenesdb.org/FTP/Genomes/Cast/v1.0/genome/Cast.1_0.hap1.fa">https://treegenesdb.org/FTP/Genomes/Cast/v1.0/genome/Cast.1_0.hap1.fa</a> [33]
Data file 2	Genome Haplotype2	Fasta file (.fasta)	<a href="https://treegenesdb.org/FTP/Genomes/Cast/v1.0/genome/Cast.1_0.hap2.fa">https://treegenesdb.org/FTP/Genomes/Cast/v1.0/genome/Cast.1_0.hap2.fa</a> [34]
Data file 3	Gene prediction on Haplotype1	GFF3 file (.gff)	<a href="https://treegenesdb.org/FTP/Genomes/Cast/v1.0/annotation/Cast.1_0.hap1.gff">https://treegenesdb.org/FTP/Genomes/Cast/v1.0/annotation/Cast.1_0.hap1.gff</a> [35]
Data file 4	Gene prediction on Haplotype2	GFF3 file (.gff)	<a href="https://treegenesdb.org/FTP/Genomes/Cast/v1.0/annotation/Cast.1_0.hap2.gff">https://treegenesdb.org/FTP/Genomes/Cast/v1.0/annotation/Cast.1_0.hap2.gff</a> [36]
Data file 5	K-mer spectrum of the Illumina reads	Image file (.png)	<a href="https://doi.org/10.6084/m9.figshare.25568154">https://doi.org/10.6084/m9.figshare.25568154</a> [37]
Data file 6	Omni-C contact maps	Image file (.svg)	<a href="https://doi.org/10.6084/m9.figshare.25568163">https://doi.org/10.6084/m9.figshare.25568163</a> [38]
Data file 7	Busco scores of the assembled sequence	Image file (.png)	<a href="https://doi.org/10.6084/m9.figshare.25568139">https://doi.org/10.6084/m9.figshare.25568139</a> [39]
Data file 8	miRNA analysis	Spreadsheet (.xlsx)	<a href="https://doi.org/10.6084/m9.figshare.25568064">https://doi.org/10.6084/m9.figshare.25568064</a> [40]
Accession identifier	Bioproject identifier	Website (Html)	<a href="http://identifiers.org/ncbi/bioproject:PRJNA1096137">http://identifiers.org/ncbi/bioproject:PRJNA1096137</a> [41]
DataSet 1	ONT reads of <i>C. sativa</i>	SRA file (.sra)	<a href="http://identifiers.org/ncbi/insdc.sra:SRR28552917">http://identifiers.org/ncbi/insdc.sra:SRR28552917</a> [42]
DataSet 2	Illumina PE-150 reads of <i>C. sativa</i>	SRA file (.sra)	<a href="http://identifiers.org/ncbi/insdc.sra:SRR28552918">http://identifiers.org/ncbi/insdc.sra:SRR28552918</a> [43]
DataSet3	Dovetail Omni-C of <i>C. sativa</i>	SRA file (.sra)	<a href="http://identifiers.org/ncbi/insdc.sra:SRR28552916">http://identifiers.org/ncbi/insdc.sra:SRR28552916</a> [44]

scores of 17.87 for Haplotype 1 and 16.85 for Haplotype 2, indicating good completeness. Repetitive elements in the genome were first identified using EDTA v2.0 [18]; then, RepeatMasker v4.1.2 [19] was used to identify and annotate the repetitive sequences.

Gene prediction was carried out separately for each haplotype. The dataset of mature miRNAs and the corresponding hairpin sequences of *Quercus robur*, which is the genetically nearest species, were retrieved from the PmiREN data repository [20]. These sequences were independently aligned to the two haplotypes using bowtie [21] (version 1.3.1) for the short miRNA reads and Blastn [22] for the hairpin sequences. The tRNAs were predicted by tRNAscan-SE v2.0.6 [23]. The coding genes prediction was carried out using Augustus v3.4 [24] and Maker v3.01 [25] (GeneMark-ES [26], Augustus and EVidenceModeler v1.1.1 [27]), trained with RNA-Seq data (downloaded from SRA: SRR8305473 and SRR15058346) and proteins from NCBI RefSeq (belonging to the Fagaceae family), respectively. The two predictions were merged according to the results of the GeneValidator v2.1.12 [28] tool, which retained only the best predictions for every gene site.

The genes masked by RepeatMasker were searched for domains associated with resistance genes using hmmsearch [29] and the PFAM domains reported in [30]. The two predictions were merged according to the results of the GeneValidator v2.1.12 tool, which retained only the best predictions for every gene site. The predicted proteins were functionally annotated using EggNOG-mapper [31] and the results were filtered using FunTaxIS-lite [32].

## Limitations

Although the gene space is quite complete, this assembly lacks resolution of the two haplotypes. The sequences of the two haplotypes were reconstructed based on the phasing of short-read sequencing enhanced with the use of long reads. A better resolution of the two haplotypes might be reached in the future by producing PacBio HiFi reads and integrating them with the data produced in this work.

## Abbreviations

ONT	Oxford Nanopore Technologies
SNP	Single Nucleotide Polymorphism
LTR	Long-Terminal Repeat
BUSCO	Benchmarking Universal Single-Copy Orthologs
NCBI	National Center for Biotechnology Information
SRA	Sequence Read Archive

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12863-024-01245-7>.

Supplementary Material 1

## Acknowledgements

Not applicable.

## Author contributions

C.V., M.C.M., D.T.M., S.M., L.D., C.M., R.B., F.S., D.M., L.P. conceived and designed the experiments, reviewed the initial draft of the manuscript, and approved the final draft submitted. L.B., P.F., A.M., S.T., M.R.M., S.P., S.A., V.P., P.P., D.M., designed and performed the experiments, analyzed the data, prepared the materials, drafted and revised the manuscript, and approved the final draft submitted. C.V., M.C.M., S.M., L.D., C.M., R.B., F.S., L.P. acquired the funding. All the authors approved the final manuscript. L.B., P.F., A.M., contributed equally to this study.

## Funding

Crowdfunding "Accordo di Ricerca e Cooperazione tecnologica: European Chestnut Genome: Chest-Gen. (2022). AM, PP and CM acknowledge funding support from the European Union–NextGenerationEU.

## Data availability

The data described in this Data note can be freely and openly accessed from the NCBI database under BioProject accession PRJNA1095814 and PRJNA1095812. The sequencing reads are available at the Sequence Read Archive (SRA) under BioProject accession PRJNA1096137. The assembled sequence and gene predictions are available for download at the following address: <https://treegenesdb.org/FTP/Genomes/.Cast/v1.0/>.

## Declarations

### Ethics approval and consent to participate

Permission to collect samples from Marrone di Chiusa Pesio was acquired from the Chestnut R&D Center Piemonte (Regione Gambarello 23, 12013 Chiusa di Pesio, Cuneo Province (Italy)); <https://centrocastanicoultura.org/en/>; Ref: Prof Gabriele Beccaro [gabriele.beccaro@unito.it](mailto:gabriele.beccaro@unito.it).

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 12 April 2024 / Accepted: 17 June 2024

Published online: 22 June 2024

## References

- Dane F, Lang P, Huang H, Fu Y. Intercontinental genetic divergence of *Castanea* species in eastern Asia and eastern North America. *Heredity*. 2003;9:314–21.
- Perkins MT, Zhebentyayeva T, SiscoPH, Craddock JH. Genome-wide sequence-based genotyping supports a nonhybrid origin of *Castanea alabamensis*. *Syst Bot*. 2021;46:973–84.
- Conedera M, Tinner W, Krebs P, de Rigo D, Caudullo G. *Castanea sativa* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo, Houston Durrant T, Mauri A, editors, *European Atlas of Forest Tree Species*. Publ. Off. EU, Luxembourg, p. e0125e0+. 2016. pp. 78–9.
- Breviglieri N. Indagini ed osservazioni sulle migliori varietà italiane di castagno (*Castanea sativa* Miller). *Suppl. La Ricerca Scientifica Anno 25. Centro Studi Castagno*. 1995;Pubbl.2.
- Alessandri S, Krznar M, Ajolfi D, Cabrer AMR, Pereira-Lorenzo S, Dondini L. Genetic diversity of *castanea sativa* mill. Accessions from the tuscan-emilian apennines and emilia romagna region (Italy). *Agronomy*. 2020. <https://doi.org/10.3390/agronomy10091319>.
- Staton M, Addo-Quaye C, Cannon N, Yu J, Zhebentyayeva T, Huff M, Islam-Faridi N, Fan S, Georgi LLI, Nelson CD, Bellis E, Fitzsimmons S, Henry N, Drautz-Moses D, Noorai RE, Flicklin S, Sasaki K, Manda MI, Wagner TK, Zembower N, Bodénès C, Holliday J, Westbrook J, Lasky J, Hebard FV, Schuster SC, Abbott AG, Carlson JE. A reference genome assembly and adaptive trait analysis of *Castanea mollissima* 'Vanuxem', a source of resistance to chestnut blight in restoration breeding. *Tree Genet Genomes*. 2020;16.
- Jiawei W, Po H, Qian Q, Dongzi Z, Lisi Z, Ke L, Shan S, Shuna J, Bingxue S, Shizhong Z, Qingzhong L. Chromosome-level genome assembly provides new insights into Japanese chestnut (*Castanea crenata*) genomes. *Front Plant Sci*. 2022;13. <https://doi.org/10.3389/fpls.2022.1049253>.
- Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*. 2017;33:2202–4.
- Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, et al. An efficient error correction and accurate assembly tool for noisy long reads. *bioRxiv*. 2023. <https://doi.org/10.1101/2023.03.09.531669>
- Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*. 2020;36:2253–5. <https://doi.org/10.1093/bioinformatics/btz891>.
- Torello Marinoni D, Nishio S, Valentini N, Shirasawa K, Acquadro A, Portis E et al. Development of high-density genetic linkage maps and identification of loci for Chestnut Gall Wasp Resistance in *Castanea* spp. *Plants Basel Switz*. 2020;9.
- Tang H, Zhang X, Miao C, Zhang J, Ming R, Schnable JC, et al. ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol*. 2015;16:3.
- Patterson M, Marschall T, Pisanti N, van Iersel L, Stougie L, Klau GW, et al. WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *J Comput Biol J Comput Mol Cell Biol*. 2015;22:498–509.
- Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics*. 2017;33(13):2037–39. [28205675].
- Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. *Curr Protocols*. 2021;1:e323. <https://doi.org/10.1002/cpz1.323>.
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics*. 2010;5:589 – 95 Epub. [PMID: 20080505].
- Ou S, Jiang N, LTR\_retriever: A highly accurate and sensitive program for identification of long terminal repeat Retrotransposons. *Plant Physiol*. 2018;176(2):1410–22. <https://doi.org/10.1104/pp.17.01310>.
- 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(1):275.
- Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2009;Chap. 4:4.10.1–4.10.14. <https://doi.org/10.1002/0471250953.bi0410s25>. PMID: 19274634.
- Guo Z, Kuang Z, Zhao Y, Deng Y, He H, Wan M, Tao Y, Wang D, Wei J, Li L, Yang X. PmiREN2.0: from data annotation to functional exploration of plant microRNAs. *Nucleic Acids Res*. 2021;50:1475–82.
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:25.
- Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36 Web Server issue: W5–9.
- Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol Clifton NJ*. 2019;1962:1–14.
- Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*. 2011;27(6):757–63. <https://doi.org/10.1093/bioinformatics/btr010>.
- Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*. 2008;18:188–96. <https://doi.org/10.1101/gr.6743907>.
- Lomsadze A, Ter-Hovhannisyanyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res*. 2005;33:6494–506.
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol*. 2008;9:R7.
- Drăgan M-A, Moghul I, Priyam A, Bustos C, Wurm Y. GeneValidator: identify problems with protein-coding gene predictions. *Bioinformatics*. 2016;32:1559–61.
- Eddy SR. Accelerated Profile HMM searches. *PLoS Comput Biol*. 2011;7:e1002195.
- Bayer PE, Edwards D, Batley J. Bias in resistance gene prediction due to repeat masking. *Nat Plants*. 2018;4:762–5.
- Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. Mol Biol Evol. 2021;38:5825–29. <https://doi.org/10.1093/molbev/msab293>.
- Falda M, Lavezzo E, Fontana P, Bianco L, Berselli M, Formentin E, Toppo S. Eliciting the Functional Taxonomy from protein annotations and taxa. *Sci Rep*. 2016;6:31971. <https://doi.org/10.1038/srep31971>.
- TreeGenes Database. *Hap1*. (2024). [https://treegenesdb.org/FTP/Genomes/.Cast/v1.0/genome/Cast.1\\_0.hap1.fa](https://treegenesdb.org/FTP/Genomes/.Cast/v1.0/genome/Cast.1_0.hap1.fa).
- TreeGenes Database. *Hap2*. (2024). [https://treegenesdb.org/FTP/Genomes/.Cast/v1.0/genome/Cast.1\\_0.hap2.fa](https://treegenesdb.org/FTP/Genomes/.Cast/v1.0/genome/Cast.1_0.hap2.fa).
- TreeGenes Database. *Gene prediction. Hap1*. (2024). [https://treegenesdb.org/FTP/Genomes/.Cast/v1.0/annotation/Cast.1\\_0.hap1.gff](https://treegenesdb.org/FTP/Genomes/.Cast/v1.0/annotation/Cast.1_0.hap1.gff).
- TreeGenes Database. *Gene prediction. Hap2*. (2024). [https://treegenesdb.org/FTP/Genomes/.Cast/v1.0/annotation/Cast.1\\_0.hap2.gff](https://treegenesdb.org/FTP/Genomes/.Cast/v1.0/annotation/Cast.1_0.hap2.gff).

37. Bianco L, Fontana P, Marchesini A, Moser M, Piazza S, Alessandri S, Pavese V, Pollegioni, Torre S, Vernesi C, Malnoy M, Sebastiani F, Torello Marinoni D, Murolo S, Dondini L, Mattioni C, Botta R, Micheletti D, Palmieri L. Data files for the genome of the plant *Castanea sativa*. figshare. (2024) <https://doi.org/10.6084/m9.figshare.25568154>.
38. Bianco L, Fontana P, Marchesini A, Moser M, Piazza S, Alessandri S, Pavese V, Pollegioni, Torre S, Vernesi C, Malnoy M, Sebastiani F, Torello Marinoni D, Murolo S, Dondini L, Mattioni C, Botta R, Micheletti D, Palmieri L. Data files for the genome of the plant *Castanea sativa*. figshare. (2024) <https://doi.org/10.6084/m9.figshare.25568163>.
39. Bianco L, Fontana P, Marchesini A, Moser M, Piazza S, Alessandri S, Pavese V, Pollegioni, Torre S, Vernesi C, Malnoy M, Sebastiani F, Torello Marinoni D, Murolo S, Dondini L, Mattioni C, Botta R, Micheletti D, Palmieri L. Data files for the genome of the plant *Castanea sativa*. figshare. (2024) <https://doi.org/10.6084/m9.figshare.25568139>.
40. Bianco L, Fontana P, Marchesini A, Moser M, Piazza S, Alessandri S, Pavese V, Pollegioni, Torre S, Vernesi C, Malnoy M, Sebastiani F, Torello Marinoni D, Murolo S, Dondini L, Mattioni C, Botta R, Micheletti D, Palmieri L. Data files for the genome of the plant *Castanea sativa*. figshare. (2024) <https://doi.org/10.6084/m9.figshare.25568064>.
41. Bioproject identifier. (2024). <http://identifiers.org/ncbi/bioproject:PRJNA1096137>.
42. ONT reads of *C. sativa*. (2024). <http://identifiers.org/ncbi/insdc.sra:SRR28552917>.
43. Illumina PE-150 reads of *C. sativa*. (2024). <http://identifiers.org/ncbi/insdc.sra:SRR28552918>.
44. Dovetail Omni-C of *C. sativa*. (2024). <http://identifiers.org/ncbi/insdc.sra:SRR28552916>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.