Check for updates

DATA NOTE

ERGA-BGE reference genome of the European bladdernut (*Staphylea pinnata*)

[version 1; peer review: awaiting peer review]

Christian de Guttry[1], Roman A. Volkov[2], Kai-hsiu Chen[3], Astrid Böhne[4], Rita Monteiro[4], Thomas Marcussen[5,6], Torsten H. Struck[6], Rebekah A Oomen[5-9], Genoscope Sequencing Team, Karine Ladabie[10], Lola Demirdjian[11], Adama N'dar[11], Pedro H. Oliveira[11], Patrick Wincker[11], Jean-Marc Aury[11], Fergal Martin[12], Vianey Paola Barrera Enriquez[12], Leanne Haggerty[12], Robert M. Waterhouse[1], Chiara Bortoluzzi[1]

[1]SIB Swiss Institute of Bioinformatics, Amphipôle, Quartier UNIL-Sorge, Lausanne, 1015, Switzerland
[2]Yuriy Fedkovych Chernivtsi National University, Kotsiubynsky str. 2, Chernivtsi, 58002, Ukraine
[3]Department of Ecology and Evolution, Biophore, Quartier UNIL-Sorge, Lausanne, 1015, Switzerland
[4]Leibniz Institute for the Analysis of Biodiversity Change, Adenauerallee 127, Museum Koenig Bonn, Bonn, 53113, Germany
[5]Centre for Ecological & Evolutionary Synthesis, Blindernveien 31, University of Oslo, Oslo, Oslo, 0371, Norway
[6]Natural History Museum, P.O. Box 1172, Blindern, University of Oslo, Oslo, 0318, Norway
[7]Centre for Coastal Research, Universitetsveien 25, University of Agder, Kristiansand, 4630, Norway
[8]Department of Biological Sciences, 100 Tucker Park Road, University of New Brunswick Saint John, Saint John, E2K5E2, Canada
[9]Tjärnö Marine Laboratory, Hättebäcksvägen 7, University of Gothenburg, Gothenburg, 45296, Sweden
[10]Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France
[11]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France
[12]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

v1 **First published:** 06 Oct 2026, **6**:396
https://doi.org/10.12688/openreseurope.25011.1
Latest published: 06 Oct 2026, **6**:396
https://doi.org/10.12688/openreseurope.25011.1

Abstract

We present a chromosome-level genome assembly for *Staphylea pinnata* L. (Crossosomatales: Staphyleaceae), commonly known as European bladdernut. The assembly spans 1.56 Gb and comprises 48 contigs assembled into 23 scaffolds. Thirteen chromosome-scale pseudomolecules were recovered, together with the mitochondrial and chloroplast genomes. The contig N50 is 71.16 Mb, and the scaffold N50 is 130.01 Mb. BUSCO analysis using the Eudicots odb10 dataset indicated 95.5% completeness, and the assembly had a Merqury quality value of 60.8. The genome annotation includes 25,870 protein-coding genes and 16,678 non-coding genes. This reference genome provides a genomic resource for *S. pinnata* and supports future studies of its genome biology, evolution, and biodiversity.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
Staphylea pinnata, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Staphyleaceae, European bladdernut

This article is included in the Horizon Europe gateway.

This article is included in the Genome Reports from the Biodiversity Genomics Europe Project collection.

Corresponding authors: Christian de Guttry (christian.deguttry@sib.swiss), Chiara Bortoluzzi (chiara.bortoluzzi@sib.swiss)

Author roles: de Guttry C: Investigation, Project Administration, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Volkov RA**: Investigation, Resources, Writing – Review & Editing; **Chen Kh**: Investigation, Resources, Writing – Review & Editing; **Böhne A**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Marcussen T**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Struck TH**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Oomen RA**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Ladabie K**: Investigation, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Demirdjian L**: Data Curation, Formal Analysis, Writing – Review & Editing; **N'dar A**: Data Curation, Formal Analysis, Writing – Review & Editing; **Oliveira PH**: Investigation, Supervision, Writing – Review & Editing; **Wincker P**: Investigation, Supervision, Writing – Review & Editing; **Aury JM**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Martin F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Barrera Enriquez VP**: Data Curation, Formal Analysis, Writing – Review & Editing; **Haggerty L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Waterhouse RM**: Project Administration, Writing – Review & Editing; **Bortoluzzi C**: Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Biodiversity Genomics Europe (Grant no. 101059492) is funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. This work was supported by the Genoscope, the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), France Génomique (ANR-10-INBS-09-08), and the exploratory research programme 'ATLASea: Atlas of marine genomes and its targeted project SEQ-Sea (ANR-22-EXAT-0003-SEQ-Sea).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 de Guttry C *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: de Guttry C, Volkov RA, Chen Kh *et al.* **ERGA-BGE reference genome of the European bladdernut (*Staphylea pinnata*) [version 1; peer review: awaiting peer review]** Open Research Europe 2026, **6**:396 https://doi.org/10.12688/openreseurope.25011.1

First published: 06 Oct 2026, **6**:396 https://doi.org/10.12688/openreseurope.25011.1

Introduction

Staphylea pinnata L., also known by the common name of Staphylier à grappes (Figure 1), is a deciduous shrub or small tree in the family *Staphyleaceae*, commonly known as European bladdernut. It is the only species of Crossosomatales indigenous to Europe. The species native range extends from eastern Central and southeastern Europe to the Caucasus and Turkey, including Switzerland (Plants of the World Online, 2026). The species typically occurs in open deciduous woodland and shrub communities and can reach approximately 4–5 m in height. It bears pinnate leaves, pendent clusters of white to yellowish flowers, and distinctive inflated capsules that give rise to its common name (InfoFlora, 2026). *Staphylea pinnata* is currently classified as Least Concern (LC) at the global level, while in Switzerland it is considered Vulnerable (VU) on the national Red List (InfoFlora, 2026, Plants of the World Online, 2026). In Switzerland, the species occurs mainly in scattered populations in the Jura, Swiss Plateau, and northern Alpine regions, where it is considered indigenous (InfoFlora, 2026). Here, we present a chromosome-level reference genome for *S. pinnata.* The generation of this genomic resource was coordinated through the Biodiversity Genomics Europe (BGE) project of the European Reference Genome Atlas (ERGA) initiative. Through coordinated sampling, sequencing, and genome assembly across European biodiversity, ERGA seeks to expand the availability of high-quality reference genomes and facilitate their use in biodiversity research and conservation (Mazzoni et al., 2023).

Materials & methods

Sample and sampling information

On 06 June 2024, Christian de Guttry sampled one specimen of *Staphylea pinnata* (hermaphrodite monoecious). The specimen was sampled in Lausanne, Canton of Vaud, Switzerland and was identified by Kai-Hsiu Chen from the University of Lausanne by visual inspection and support from iNaturalist. After removing cuttings of the individual from its natural habitat, leaves and buds were dissected with a scalpel/scissors and immediately placed in liquid nitrogen. Until DNA extraction, samples were preserved at −80 °C. No permits were required for sampling.

Vouchering information

Physical reference materials have been deposited in the Musée et jardin botaniques cantonaux de Lausanne (https://natureum.ch/fr/). Frozen reference tissue material is not available. A total of 6 electronic voucher images of the sequenced individual is available from ERGA's EBI BioImageArchive dataset https://www.ebi.ac.uk/biostudies/bioimages/studies/S-BIAD1012?query=ERGA under accession ID SAMEA115720302 (Figure 1).

Data availability

Staphylea pinnata and the related genomic study were assigned to Tree of Life ID (ToLID) 'drStaPinn1' and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB84171. The sample information is available at the following BioSample accessions: SAMEA117814047 and SAMEA115720313. The genome assembly is accessible from ENA under accession number GCA_965153445.2, whereas the annotated genome is available for accession number GCA_965153445.1 through the Ensembl website (https://projects.ensembl.org/erga-bge/). Sequencing data produced as part of this project are available from ENA at the following accessions: ERX14170727, ERX14170728, ERX14170729, ERX14170730, and ERX14096375. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) document available at https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Staphylea_pinnata/drStaPinn1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/213841.

DNA/RNA processing

DNA was extracted from 550 mg of leaves using a conventional CTAB extraction followed by a commercial purification using Qiagen Genomic tips (Qiagen). A detailed protocol is available on protocols.io (Vacherie et al., 2022). DNA fragment size selection was performed using Short Read Eliminator (PacBio). Quantification was performed using a Qubit dsDNA HS Assay kit (Thermo Fisher Scientific) and integrity was assessed in a FemtoPulse system (Agilent). DNA was stored at 4oC until usage. RNA was extracted from leaves or buds (30–50 mg) using RNeasy Powerplant Kit (Qiagen) following manufacturer instructions. Residual genomic DNA was removed with 6 U of TURBO DNase (2 U/µl) (Thermo Fisher Scientific). Quantification was performed using a Qubit RNA HS Assay and integrity was assessed in a Bioanalyzer system (Agilent). RNA was stored at −80 °C. The RNA was highly degraded and did not allow RNA sequencing to be performed.

Library preparation and sequencing

Long-read DNA libraries were prepared with SMRTbell prep kit 3.0 following manufacturers' instructions and sequenced on a Revio system (PacBio). A supplementary purification step using the DNeasy PowerClean Pro Cleanup Kit was performed after DNA fragmentation to optimize sequencing yield. Omni-C libraries were prepared using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, CA, USA) after plant nuclei isolation. Briefly, flash-frozen young leaves (1 g) were cryoground in liquid nitrogen, and pure nuclei were first isolated following the "High Molecular

Weight DNA Extraction from Recalcitrant Plant Species" protocol described by Workman et al. (2018) (Workman et al., 2018). Once the nuclei had been isolated, the pellets were treated as mammalian cells and Omni-C libraries were prepared according to the Mammalian Cell Protocol for Sample Preparation v1.4. The final libraries were sequenced on an Illumina NovaSeq X Plus instrument (Illumina, San Diego, CA, USA) with 2x 150 read length. In total, 50x PacBio HiFi and 137x HiC data were sequenced to generate the assembly.

Genome assembly methods

The genome of *Staphylea pinnata* was assembled using the Genoscope GALOP pipeline (https://workflowhub.eu/workflows/1200). Briefly, raw PacBio HiFi reads were assembled using Hifiasm v0.19.8-r603 (Cheng et al., 2021). Remaining allelic duplications were removed using purge_dups v1.2.5 (Guan et al., 2020) with default parameters and the proposed cutoffs. The purged assembly was scaffolded using YaHS v1.2.2 (Zhou et al., 2023), and the assembled scaffolds were then curated through manual inspection using PretextView v0.2.5 (Harry, 2022) to remove false joins and



Figure 1. A specimen of the species *Staphylea pinnata*. The individual in the picture is from Lausanne, Canton of Vaud, Switzerland, and is the one sampled and sequenced for this study. Photo credit: Christian de Guttry.

Table 1. Genome assembly statistics.

Genome assembly	
Assembly name	drStaPinn1.2
Assembly accession	GCA_965153445.2
Assembly level	Chromosome
Span (Gb)	1.56
Number of chromosomes	13
Number of contigs	48
Contig N50	71,159,830
Number of scaffolds	23
Scaffold N50	130,005,016
Organelles	Chloroplast

incorporate sequences not automatically scaffolded into their respective locations within the chromosomal pseudomolecules. Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (10.48546/workflowhub.workflow.1104.1).

Genome annotation methods

A gene set was generated using the Ensembl Gene Annotation system (Aken et al., 2016), primarily by aligning publicly available short-read RNA-seq data from BioSample: SAMEA104170352 to the genome. Gaps in the annotation were filled via protein-to-genome alignments of a select set of vertebrate proteins from UniProt (Consortium, 2019), which had experimental evidence at the protein or transcript level. At each locus, data were aggregated and consolidated, prioritising models derived from RNA-seq data, resulting in a final set of gene models and associated non-redundant transcript sets. To distinguish true isoforms from fragments, the likelihood of each open reading frame (ORF) was evaluated against known vertebrate proteins. Low-quality transcript models, such as those showing evidence of fragmented ORFs, were

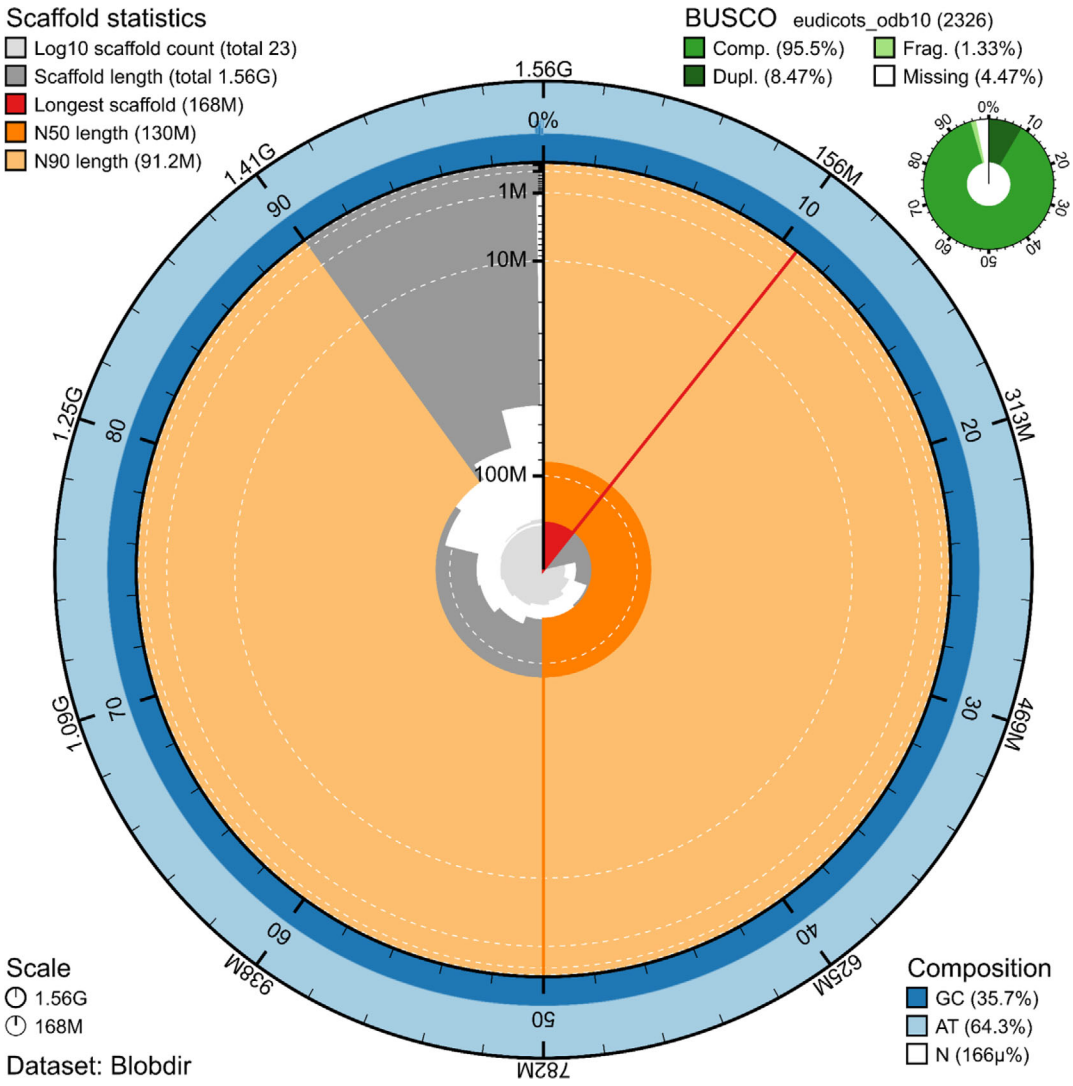


Figure 2. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 1,563,064,433 bp assembly including the mitochondrial genome and chloroplast. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (168 Mb, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (130,005,016 and 91,176,575 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log-scale, with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT, and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated, and missing BUSCO genes found in the assembled genome from the Eudicots database (odb10) is shown in the top right.

removed. In cases where RNA-seq data were fragmented or absent, homology data were prioritised, favouring longer transcripts with strong intron support from short-read data. The resulting gene models were classified into three categories: protein-coding, pseudogene, and long non-coding. Models with hits to known proteins and few structural abnormalities were classified as protein-coding. Models with hits to known proteins but displaying abnormalities, such as the absence of a start codon, non-canonical splicing, unusually small intron structures (<75 bp), or excessive repeat coverage, were reclassified as pseudogenes. Single-exon models with a corresponding multi-exon copy elsewhere in the genome were classified as processed (retrotransposed) pseudogenes. Models that did not fit any of the previously described categories, did not overlap protein-coding genes and were constructed from transcriptomic data were considered potential lncRNAs. Potential lncRNAs were further filtered to remove single-exon loci due to their unreliability. Putative miRNAs were predicted by performing a BLAST search of miRBase (Griffiths-Jones et al., 2006) against the genome, followed by RNAfold analysis. Other small non-coding loci were identified by scanning the genome with Rfam (Nawrocki et al., 2015) and passing the results through Infernal (Nawrocki & Eddy, 2013). Summary analysis of the released annotation was carried out using the ERGA-BGE Genome Report ANNOT Galaxy workflow (10.48546/workflowhub.workflow.1096.1).

Results

Genome assembly

The genome was assembled into one haplotype using Hi-C phasing and curated to chromosome level (Table 1).

The genome assembly has a total length of 1,563,064,433 bp in 23 scaffolds including the mitogenome and chloroplast (Figure 2), with a GC content of 35.7%. The assembly has a contig N50 of 71,159,830 bp and L50 of 8, and a scaffold N50 of 130,005,016 bp and L50 of 6. The assembly has a total of 25 gaps, totalling 2.6 kb in cumulative size. The single-copy gene content analysis using the Eudicots database (odb10) with BUSCO v5.8.0 (Manni et al., 2021) resulted in 95.5% completeness (87.1% single-copy and 8.5% duplicated). 89.7% of read k-mers were present in the assembly (Figure 3) and the assembly has a base accuracy Quality Value (QV) of 60.8 as calculated by Merqury (Rhie et al., 2020).

Chromosome-level scaffolds were named according to size, as confirmed by Hi-C data (Figure 4).

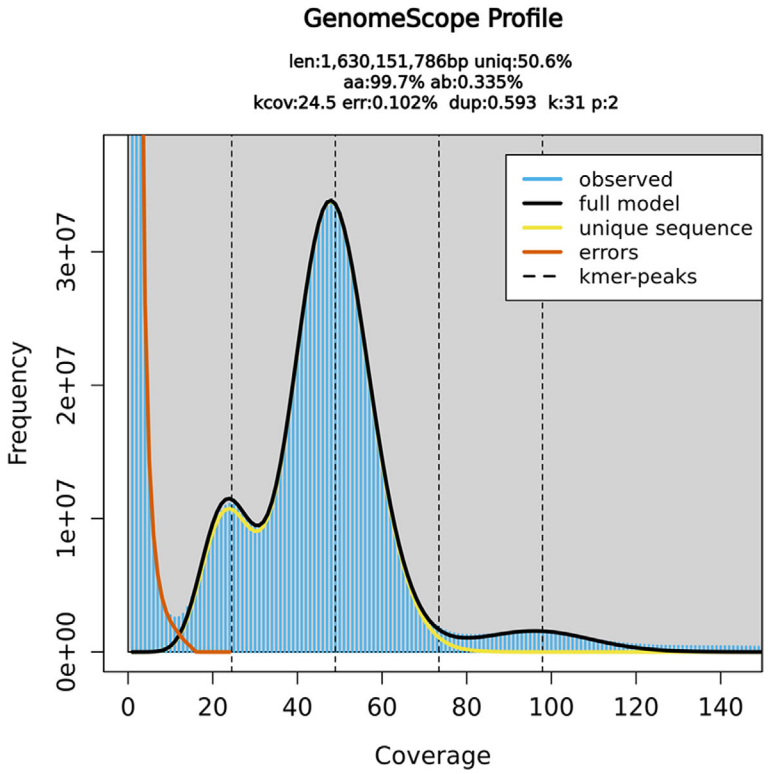


Figure 3. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Genome annotation

The genome annotation consists of 25,870 protein-coding genes with associated 33,681 transcripts, in addition to 16,678 non-coding genes (Table 2). Using the longest isoform per transcript, the single-copy gene content analysis using the Eudicots database (odb10) with BUSCO resulted in 90.7% completeness. Using the OMAmer Viridiplantae-v2.0.0.h5 database for OMArk (Nevers et al., 2025) resulted in 91.3% completeness and 84.6% consistency (Table 3).

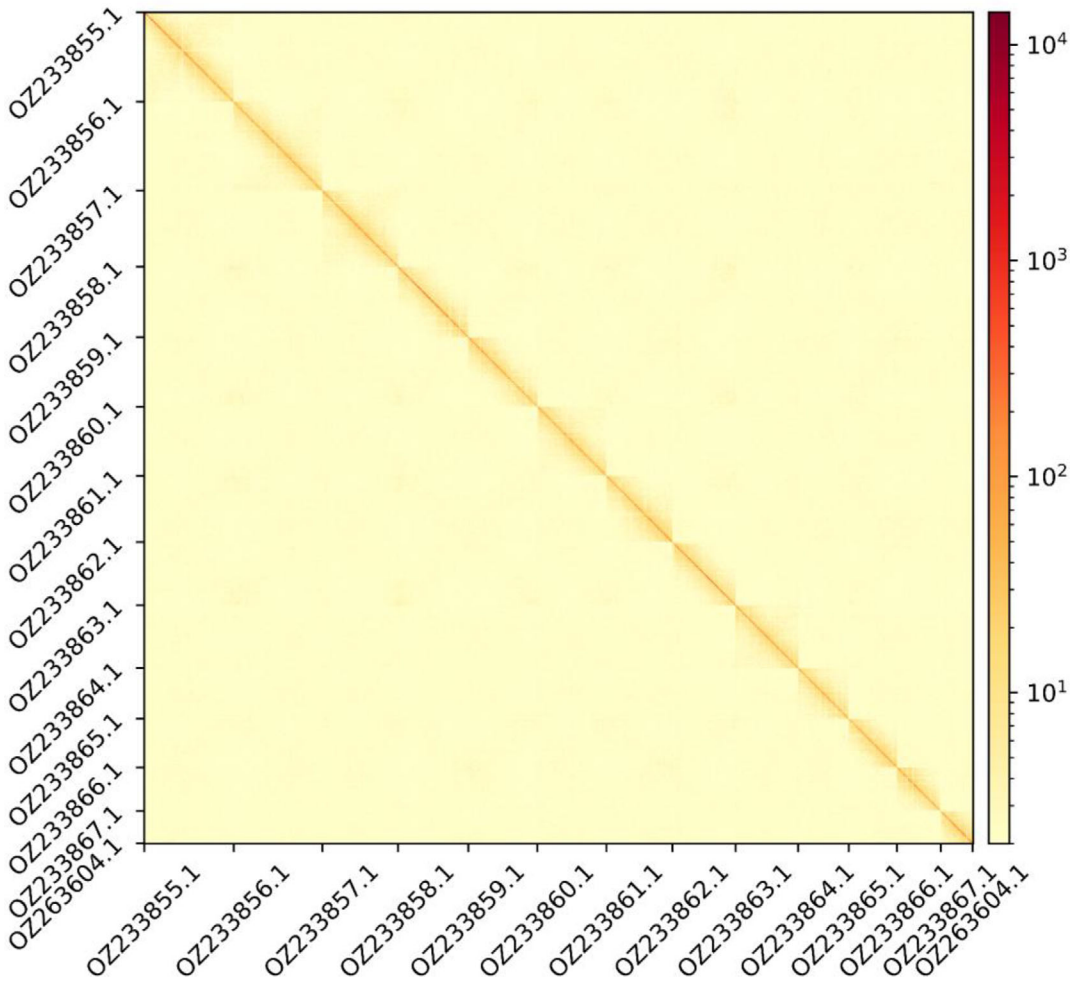


Figure 4. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 100 kb bin size for plotting. The x-axis and y-axis show the 13 autosomes and the mitochondrial genome (GenBank: OZ263604.1).

Table 2. Statistics from assembled gene models.

	No. genes	No. transcripts	Mean gene length (bp)	No. single-exon genes	Mean exons per transcript
mRNA	25,870	33,681	4,377	1,387	5.6
snoRNA	13,394	13,394	106	13,394	1.0
lncRNA	1,846	1,921	1,611	10	2.3
tRNA	463	463	75	463	1.0
snRNA	158	158	137	158	1.0
rRNA	817	817	292	817	1.0
Other ncRNA	945	945	74–122	945	1.0–1.0

Table 3. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (eudicots_odb10) and OMArk (Viridiplantae-v2.0.0.h5).

	Complete	Single copy	Duplicated	Fragmented	Missing
BUSCO	2,111 (90.7%)	1,971 (84.7%)	140 (6.0%)	65 (2.8%)	150 (6.5%)
OMArk	10,707 (91.3%)	7,625 (65.0%)	3,082 (26.3%)	-	1,013 (8.6%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	21,892 (84.6%)	2,667 (10.3%)	0.0 (0.0%)	1,311 (5.1%)	

Conflict of interest
The authors declare no conflict of interest related to this study. The funding sources had no involvement in the study design, collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. All authors have participated sufficiently in the work to take public responsibility for the content and agree to the submission of this manuscript.

Funder information
Biodiversity Genomics Europe (Grant no. 101059492) is funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. This work was supported by the Genoscope, the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), France Génomique (ANR-10-INBS-09-08), and the exploratory research programme 'ATLASea: Atlas of marine genomes and its targeted project SEQ-Sea (ANR-22-EXAT-0003-SEQ-Sea).

Author contributions
RAV proposed *Staphylea pinnata* as a target for genome sequencing and provided the rationale for its selection; CdG coordinated the project; RMW supervised the project; CdG collected the species; KHC identified the species; CdG sampled and preserved biological material and provided metadata; AsB, RM, TM, THS, and RAO provided support in sampling, shipping of biological material, metadata collection, and management; the GST extracted DNA, prepared libraries, and performed sequencing under the supervision of KL, PHO, and PW; LD and AN performed genome assembly and curation under the supervision of JMA; FM, VPBE, and LH performed genome annotation; CB generated the analysis and report. All authors contributed to the writing, review, and editing of this genome note and read and approved the final version. This work is part of the species assigned to Genoscope, which was instrumental in the wet lab, sequencing, and assembly processes, and represents a key contribution to BGE's outputs.

Author information
Members of the Genoscope Sequencing Team are listed here: https://doi.org/10.5281/zenodo.14611490.

Acknowledgements
We acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de. NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de. NBI Freiburg. We would like to acknowledge the assembly reviewer, Tom Mathers, from the Wellcome Sanger Institute.

Literature Cited

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database.* 2016; **2016**: baw093.

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm.** *Nat. Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Consortium, U: **UniProt: A worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
PubMed Abstract | Publisher Full Text | Free Full Text

Griffiths-Jones S, Grocock RJ, Van Dongen S, *et al.*: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res.* 2006;

34(suppl_1): D140–D144.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022.

InfoFlora: 2026. Retrieved 08 September 2026.
Reference Source

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO: Assessing Genomic Data Quality and Beyond.** *Current Protocols.* 2021; **1**(12): e323.
PubMed Abstract | **Publisher Full Text**

Mazzoni CJ, Ciofi C, Waterhouse RM: **Biodiversity: An atlas of European reference genomes.** *Nature.* 2023; **619**: 252–252.
Publisher Full Text

Nawrocki EP, Burge SW, Bateman A, *et al.*: **Rfam 12.0: Updates to the RNA families database.** *Nucleic Acids Res.* 2015; **43**(D1): D130–D137.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Nawrocki EP, Eddy SR: **Infernal 1.1: 100-fold faster RNA homology searches.** *Bioinformatics.* 2013; **29**(22): 2933–2935.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

POWO: *Plants of the World Online.* Kew: Facilitated by the Royal Botanic Gardens; 2026. Retrieved 08 September 2026.
Reference Source

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vacherie B, Labadie K, Falentin C: **HMW DNA extraction for Long Read Sequencing using CTAB.** 2022.

Workman R, Timp W, Fedak R, *et al.*: **High molecular weight DNA extraction from recalcitrant plant species for third generation sequencing.** 2018.
Reference Source

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**