

Haplotype-Resolved Genome of the Critically Endangered, Paleo-endemic Tree, *Eidothea hardeniana*

Abhishek Soni ^{1,2}, Agnelo Furtado ^{1,2}, Maurizio Rossetto ^{3,4}, Robert J. Henry ^{1,5,*}

¹ARC Centre of Excellence for Plant Success in Nature and Agriculture, The University of Queensland, St Lucia, QLD 4072, Australia

²Centre for Crop Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD 4072 Australia

³Research Centre for Ecosystem Resilience, Royal Botanic Garden Sydney, Sydney, NSW 2000, Australia

⁴Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD 4072 Australia

⁵VinUni Big Data Research Institute, VinUniversity, Hanoi, Vietnam

*Corresponding author: E-mail: robert.henry@uq.edu.au.

Accepted: March 13, 2026

Abstract

A fully phased, chromosome-level reference genome assembly for an IUCN-declared critically endangered and paleo-endemic tree, *Eidothea hardeniana* (Nightcap Oak) was assembled de novo using long-read PacBio HiFi, Oxford Nanopore (ONT) and Hi-C sequencing technologies. Consensus assembly represented high contiguity (N50 ~ 49 Mb), (BUSCO 99.2%), assembly size in accordance with flow cytometry derived genome size estimate, and the presence of telomeric and centromeric repeats on pseudochromosomes. Haplotype 1 (598 Mb) achieved an N50 of 45.8 Mb and 99.1% BUSCO completeness, while Haplotype 2 (596 Mb) showed similar quality (N50: 47.3 Mb; BUSCO: 98.4%). Repeat content comprised approximately 54% of the genome. Gene annotation confirmed approximately 30120 protein coding genes with BUSCO completeness of 99.5%. Comparative orthology with other assembled Proteaceae genomes identified ~13,000 shared ortholog clusters and 227 *Eidothea*-specific orthologs clusters. Heterozygosity in *E. hardeniana* was estimated at 1.17 SNVs per kilobase (0.12%), which is approximately one-third of the level reported for the other sequenced critically endangered Proteaceae species, *Macadamia janseni*. This genome represents the first assembled genome from the Australian lineage of the Proteoideae subfamily within Proteaceae. As a high-quality reference for conservation genomics, it provides a foundation for SNP genotyping, population structure analyses, and the development of informed genetic rescue strategies for *E. hardeniana* and related species.

Key words: *Eidothea*, proteaceae, inbreeding, conservation, genome, IUCN.

Significance

This study presents the first chromosome-scale, haplotype-resolved reference genome for *Eidothea hardeniana*, an IUCN-listed critically endangered Australian species from the Proteoideae clade of the Proteaceae family. This genome provides an essential resource for phylogenomics, conservation genetics, genetic diversity assessments, and studies of genome evolution in *E. hardeniana* and the wider Proteaceae.

Introduction

Eidothea hardeniana P.H Weston & Kooyman, commonly Nightcap Oak, an IUCN declared critically endangered

plant species, was rediscovered in early 2000s (Weston and Kooyman 2002). The genus *Eidothea* is classified in the Proteoideae subfamily of the Proteaceae, which

consists of 27 genera and 100 species found in a wide range of habitats in the southern hemisphere (Rossetto and Kooyman 2005; Weston and Barker 2006; Sauquet et al. 2009).

Eidothea hardeniana is considered a paleo-endemic species, once more widespread but now confined to a narrow 8 to 10 km² area in New South Wales, Australia with approximately 250 individuals (Rossetto and Porter 2005). Its historical decline has been attributed to climatic shifts and habitat fragmentation over evolutionary timescales (Rossetto and Kooyman 2005; Rossetto and Porter 2005). The small population exhibits a limited capacity for seed dispersal and recruitment (Rossetto and Kooyman 2005). These characteristics, combined with low census numbers, make *E. hardeniana* especially vulnerable to inbreeding and genetic erosion. Understanding the genetic structure of the populations is critical for effective conservation. The lack of a suitable reference genome has been a major barrier to assessing key parameters such as genomic diversity, demographic bottlenecks, and functional gene retention, factors critical for informing conservation interventions.

Here we report a haplotype-resolved, chromosome-scale reference genome for *E. hardeniana*. The genome was assembled using Pacific Biosciences High Fidelity (HiFi) reads (Fig. S1) and Hi-C data, with extra-long Oxford Nanopore (ONT) sequencing reads (Fig. S2, Nurk et al. 2022; Shang et al. 2023; Espinosa et al. 2024; Lu et al. 2024). The resulting reference genome provides a crucial foundation for investigating genetic load, gene content, and genome architecture in an evolutionarily isolated lineage.

Results and Discussion

Genome Size Estimate and Ploidy Analysis

Flow cytometry-based genome size estimate for *E. hardeniana* was 1C = 605 Mbp (Fig. S3). K-mer-based size estimates varied significantly from 550 to 770 Mbp (Figs. S4 and S5). Using short k-mer (21), the ploidy inferred by Smudgeplot was tetraploid (Fig. S6), while using the best predicted k-mer (127) by Kmergenie, the Smudgeplot predicted a diploid (Fig. S7). Importantly, the flow cytometry estimate (~605 Mbp) closely matched the final assembly size, supporting interpretation of the sampled individual as diploid. This also corrected earlier overestimates (1C = 887 Mbp) that likely resulted from inaccurate reference standards (e.g. *Pisum sativum*) used in previous FCM analyses (Vesely et al. 2011; Šmarda et al. 2014).

Genome Assembly

The HiFi + Hi-C assembly produced five T2T scaffolds. Incorporating unfiltered ONT reads, this assembly consisted of one T2T contig of chromosome length

(Fig. S8, Table S1). While scaffolding achieved six T2T scaffolds. Filtering ONT reads to >20, >30 and >40 kb further reduced fragmentation at contig level assembly as the number of contigs went down from 1684 to 670 (Table S2), likewise the scaffolds number reduction from 1701 to 644 indicated less fragmentation at scaffold-level assembly. Assemblies incorporating ONT reads up to 40 kb yielded five to six chromosome-scale T2T scaffolds (Figs. S9 to S14). Whereas, the integration of extra-long ONT reads (>50 kb) produced 14 chromosome-length scaffolds, 10 of which contained telomeric repeats at both ends (Figs. S9 and S15). Hi-C contact map supported the chromosome-scale scaffolding and identified centromeric interaction patterns (Fig. S16). Four scaffolds contained telomeric repeats at only one end (Fig. S15), and were therefore, considered candidate chromosome arms. Based on (i) strong collinearity to *Protea cynaroides* genome (Fig. S17), and Hi-C interaction support for adjacency and orientation (Fig. S18 to S19), these scaffolds were joined during manual curation into two chromosome-scale pseudochromosomes. Subsequently, telomeric motifs (TTTAGGG) were identified at both ends of all manually joined scaffolds (Fig. S20), supporting their structural completeness. Hi-C contact maps (Figs. S21 and S22) and centromere predictions (Fig. S23) further supported that all 12 pseudochromosomes showed a single major centromeric signal except chromosome 1, which showed two centromere-like peaks (Fig. S22); in the absence of karyological confirmation, we interpret this cautiously as either a dicentric configuration or an artefact of centromere inference or scaffolding. Due to the critically endangered status of *E. hardeniana*, karyotyping was not feasible, necessitating reliance on the genome assembly, Hi-C contact maps, and comparative collinearity to infer chromosome structure.

The 12 pseudochromosomes in the assembly accounted for the 99.2% BUSCO with a total length of 614 Mb and N50 of 49 Mb (Table 1). Smaller unplaced scaffolds (69 Mb) were identified containing organellar DNA (mitochondrial and chloroplast DNA) (Fig. S24 to S26). The final assembly was aligned with *P. cynaroides* and *Macadamia janssenii* (Fig. S27 and S28). As expected, being in the same subfamily, Proteoideae, *E. hardeniana* showed more resemblance to the *P. cynaroides* genome.

Both haplotypes exhibited >98% BUSCO completeness against the embryophyta_odb10 dataset. Dot plot alignments confirmed high collinearity between the two haplotypes and the consensus genome (Figs. S29 to S30), indicating structurally consistent chromosome-scale assemblies.

Repeat Content Analysis

The consensus genome contained 53.9% repetitive DNA, of which a substantial proportion (~33%)

Table 1 Assembly statistics of the final consensus and haplotype genome assemblies and comparison against other Proteaceae genomes

Quality parameters	<i>Eidothea hardeniana</i>			<i>Macadamia janseni</i> (Sharma et al. 2024)	<i>Telopea speciosissima</i> (Chen et al. 2022)
	Consensus	Haplotype 1	Haplotype 2		
Number of scaffolds	12	12	12	14	11
Total size (Mb)	614.2	598.1	595.9	773	823
Min length (Mb)	40	39.8	38.4	-	-
Max length (Mb)	79.5	73.3	74.9	-	-
N90 (Mb)	40.4	39.8	39.9	-	-
N80 (Mb)	44.6	43.7	43.9	-	-
N70 (Mb)	45.3	44.4	44.1	-	-
N60 (Mb)	47.8	45.8	47.1	-	-
N50 (Mb)	48.9	45.8	47.3	54.7	69
Average length (Mb)	51.2	48.2	50.1	-	-
Number of contigs	26	35	36	-	-
Contig N50 (Mb)	24	23	23.5	-	-
Percent gaps (%)	0.00%	0.00%	0.00%	-	-
BUSCO					
Complete (C)	99.20%	99.10%	98.40%	97.70%	97.80%
Single copy (S)	86.70%	87.00%	86%	84.20%	86.70%
Duplicated (D)	12.50%	12.10%	12.40%	13.50%	11.20%
Fragmented (F)	0.20%	0.20%	0.20%	0.70%	1.70%
Missing (M)	0.60%	0.70%	1.40%	1.60%	0.50%

remained unclassified (Table S3). Both haplotypes similarly comprised just over half of their genomic sequence as repetitive DNA (53.2% in Hap1 and 53.4% in Hap2). The repeat fraction was dominated by retrotransposons, particularly LTR elements (14.38%), with Gypsy/DIRS1 (10.8%) and Ty1/Copia (2.52%) families representing the major components. LINEs accounted for approximately 3.1% of the genome, largely from L1/CIN4 subfamilies (2.28%). Overall, the repeat landscapes of the two haplotypes were highly similar (Table S4 to S5).

Structural Annotation and Gene Model Quality

Structural annotation identified 30,411 protein-coding sequences in the consensus assembly (Table S6). Comparison of the two haplotype assemblies indicated a high degree of similarity in gene content and genomic architecture, with only minor differences (Table S6). A total of 29,712 and 29,416 coding sequences were predicted in Hap1 and Hap2, respectively. Assessment of gene model completeness using BUSCO against the embryophyta_odb10 database indicated 97.7% completeness for the consensus genome assembly, with 10.3% classified as complete and duplicated BUSCOs (Table S7).

Functional Annotation of the Genome

Functional annotation of the 30,120 identified protein-coding sequences assigned at least one Gene Ontology (GO) term to 24,649 genes (Table S8). Among the 548 genes without BLAST matches, 449 were supported as

protein-coding by coding-potential analysis. Comparison of the two haplotype assemblies (Hap1 and Hap2) revealed a high degree of concordance in gene content and annotation. In Hap1, from 29,712 coding sequences, 24,304 were annotated with at least one GO term (Table S8). Whereas in Hap2, from 29,416 predicted coding sequences, 24,011 received at least one GO annotation.

Comparative Genomics and Orthologous Gene Analysis

A total of 67,982 orthologs clusters were identified across all species from which 13,086 orthologs clusters were shared across all four species (Fig. S31). Only 227 unique orthologs clusters were identified for the *E. hardeniana* (Fig. S31). These, orthologs may reflect lineage-specific gene retention or divergence since its early separation from the other Proteaceae lineages (Fig. S32). Comparative orthology analysis across Proteaceae revealed 227 species-specific orthogroups in *E. hardeniana*.

Synteny and Collinearity

Self-synteny analysis revealed limited evidence of ancient whole genome duplication followed by rearrangements within the genome (Fig. 1). Chromosomes 5, 6, 11 and 12 indicated comparatively lower chromosomal rearrangement than others (Fig. 1) with strong collinearity and structural conservation between haplotypes (Fig. S33).

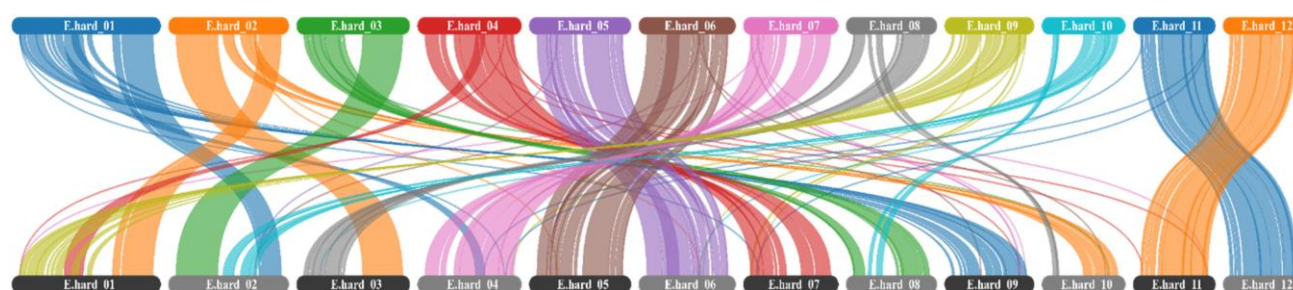


Fig. 1. Syntenic relationships within the *Eidothea hardeniana* genome based on chromosome-level scaffolds of the final consensus assembly. Ribbons connect syntenic gene blocks between chromosomes, revealing patterns of ancient segmental duplication and potential paralogy.

Structural comparison between the two haplotype-resolved assemblies of *Eidothea hardeniana* using SyRI revealed largely conserved genome architecture with only a limited number of structural rearrangements (Fig. S34). Comparative analyses between *E. hardeniana* and *P. cynaroides* revealed both deeply conserved synteny and substantial rearrangements (Fig. S35 to S36).

Antimicrobial Peptides Genes

Genes with homology to antimicrobial peptides (AMPs) in *Eidothea hardeniana* exhibited structural conservation despite a high degree of genome divergence. The g686.t1 transcript in *E. hardeniana* contained three copies of the conserved cysteine motif, in contrast to four copies in *Macadamia integrifolia* and *M. janssenii*. The *E. hardeniana* AMP was shorter, consisting of 543 amino acids compared with 636 amino acids in *M. integrifolia* (AF161884.1) (Fig. S37) (Marcus et al. 1999; Sharma et al. 2021, 2024).

Variant Calling and Heterozygosity

The number of SNVs (single nucleotide variants) was assessed across a range of allele frequency thresholds. As the minimum frequency threshold for variant detection decreased from 40% to 15%, the total number of variant positions increased from 417,292 to 720,563 (Fig. S38). This asymptotic increase indicated saturation of high-confidence variant detection at lower frequency thresholds, supporting the robustness of heterozygosity estimates. Based on the plateau observed below 20% frequency threshold, the heterozygosity rate was estimated to be approximately 1.17 SNVs per kilobase or 0.12%, considerably lower than in related Proteaceae species such as *Protea cynaroides* (1.07%) (Chang et al. 2023), *M. integrifolia* (1.55%) (Sharma et al. 2024), *M. tetraphylla* (1.33%) (Sharma et al. 2024), *M. janssenii* (0.31%) (Sharma et al. 2021), and *Telopea speciosissima* (0.76%) (Chen et al. 2022). After refining for 100% homozygous and high-confidence heterozygous calls, 719,174 SNV positions were retained. The low

heterozygosity observed in *E. hardeniana* is likely a consequence of prolonged demographic isolation and a persistently small population size. Despite this, the estimated heterozygosity indicates that the species retains a reservoir of genetic diversity that may be important for future adaptability. However, the restricted population size suggests that inbreeding and genetic drift continue to influence genomic variation. The genetic distinctiveness of each surviving individual further highlights the conservation significance of the species, as the loss of any single tree may result in the irreversible loss of unique alleles from the gene pool.

Materials and Methods

Genome Size and Ploidy Estimation

Young green leaves of *Eidothea hardeniana* (accession AA 20010147) were collected from the Royal Botanic Garden Sydney, Australia. Genome size was estimated using both flow cytometry (FCM) and k-mer approaches. For FCM, *E. hardeniana* leaf tissue was co-chopped with Nipponbare rice following established protocols (Soni and Henry 2025). Fluorescence was measured as described previously (Soni et al. 2025; Soni and Henry 2025). For k-mer analyses, Kmergenie was used to select the best k-mer and obtain an initial estimate (Chikhi and Medvedev 2014). Subsequently, GenomeScope 2.0 was used to estimate genome size and heterozygosity (Ranallo-Benavidez et al. 2020). Jellyfish and Smudgeplot were used to evaluate ploidy scenarios (Marçais and Kingsford 2011; Ranallo-Benavidez et al. 2020) (Supplementary Methods S1).

DNA/RNA Sequencing and QC

High molecular weight DNA was extracted using a modified CTAB method (GIH_SOP204-02, <https://dx.doi.org/10.17504/protocols.io.b5qqy5xw>), and long-read sequencing (PacBio HiFi, ONT), Hi-C sequencing, and short-read WGS were generated. Read QC used FastQC; adapter/quality trimming used fastp (Chen et al. 2018; Chen 2023), and HiFi reads were screened for adapter

contamination. RNA was extracted using modified CTAB-based workflows (Wang and Stegemann 2010; Rubio-Piña and Zapata-Pérez 2011) and QC-filtered with fastp (Chen et al. 2018; Chen 2023).

Assembly, Curation, Annotation, and Comparative Analyses

Haplotype-resolved assemblies were generated using HiFiasm under multiple ONT-integration strategies (Cheng et al. 2021), and Hi-C scaffolding used YaHS (Zhou et al. 2023) after mapping Hi-C reads with bwa (Li 2013). Scaffolding was validated using dot plots (Cabanettes and Klopp 2018) and curated by joining highly collinear scaffolds. Assembly quality and metrics were assessed using BUSCO and QUAST (Gurevich et al. 2013; Simão et al. 2015; Seppey et al. 2019), and Hi-C contact maps were inspected using standard Hi-C visualization workflows (Durand et al. 2016a, 2016b; Wolff et al. 2022). Repeats were identified with RepeatModeler2 and masked with RepeatMasker (Chen 2004; Flynn et al. 2020). Genes were predicted with BRAKER3 (Gabriel et al. 2024) using RNA-seq evidence and Viridiplantae peptides (OrthoDB_v11) obtained from (<https://www.orthodb.org>), and were functionally annotated with BLAST/GO/InterPro approaches (Conesa et al. 2005; Jones et al. 2014). Variant calling and heterozygosity was estimated using short reads, mapped under multiple alignment stringencies in CLC Genomics Workbench. Comparative Orthology was conducted using OrthoFinder (Emms and Kelly 2019). Synteny was inferred using DIAMOND and MCScanX (Wang et al. 2012; Buchfink et al. 2021). Chromosomal rearrangements were inferred with minimap2 and SyRI (Li 2018; Goel et al. 2019). Full protocols, software versions, and parameter settings are provided in **Supplementary Methods S1**.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Dr Subash Rai from Genome Innovation Hub, The University of Queensland, for his assistance in DNA extractions and long-read sequencing. We also thank Stephanie Chen, Richard Dimon, Samantha Yap from Royal Botanic Garden Sydney for their assistance in plant material collection and logistics.

Author Contributions

A.S.: Experimental design, sample coordination, data collection and analysis, manuscript drafting, review,

and editing. **R.J.H.:** Conceptualization, project supervision, funding acquisition, manuscript review and editing. **A.F.:** Supervision and guidance, **M.R.:** Supervision and guidance.

Funding

AS was awarded a PhD Scholarship and RH was funded by the ARC Centre of Excellence for Plant Success in Agriculture and Nature (Grant CE200100015).

Data Availability

All sequencing data for *Eidothea hardeniana* are deposited at NCBI under BioProject PRJNA1330521 with raw reads linked to BioSample SAMN51526000. Short-read WGS (SRR35515829), RNA-seq (SRR35515830), Hi-C (SRR35515831), PacBio HiFi (SRR35515833), and ONT long reads (SRR35534503) are available at the Sequence Read Archive (SRA). Genome assemblies include the consensus assembly (SUB15652535), haplotype 1 (SUB15653222), and haplotype 2 (SUB15653255). **figures** are supplied as separate files.

For peer review, data are available via NCBI reviewer access:

BioProject PRJNA1330521 (reviewer link):

<https://dataview.ncbi.nlm.nih.gov/object/PRJNA1330521?reviewer=7hq106mjmcog25t5lto0jf7pn9>

SRA accessions: WGS (SRR35515829), RNA-seq (SRR35515830), Hi-C (SRR35515831), PacBio HiFi (SRR35515833), ONT (SRR35534503). **figures** accompany the manuscript.

Literature Cited

- Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>.
- Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ*. 2018;6:e4958. <https://doi.org/10.7717/peerj.4958>.
- Chang J, et al. The genome of the king protea, *Protea cynaroides*. *Plant J*. 2023;113:262–276. <https://doi.org/10.1111/tjp.16044>.
- Chen N. Using repeat masker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2004;5:4–10. <https://doi.org/10.1002/0471250953.bi0410s05>.
- Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta*. 2023;2:e107. <https://doi.org/10.1002/imt.2.107>.
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Chen SH, et al. Chromosome-level de novo genome assembly of *Telopea speciosissima* (New South Wales waratah) using long-reads, linked-reads and Hi-C. *Mol Ecol Resour*. 2022;22:1836–1854. <https://doi.org/10.1111/1755-0998.13574>.

- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics*. 2014;30:31–37. <https://doi.org/10.1093/bioinformatics/btt310>.
- Conesa A, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>.
- Durand NC, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 2016a;3:99–101. <https://doi.org/10.1016/j.cels.2015.07.012>.
- Durand NC, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016b;3:95–98. <https://doi.org/10.1016/j.cels.2016.07.002>.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Espinosa E, Bautista R, Larrosa R, Plata O. Advancements in long-read genome sequencing technologies and algorithms. *Genomics*. 2024;116:110842. <https://doi.org/10.1016/j.ygeno.2024.110842>.
- Flynn JM, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020;117:9451–9457. <https://doi.org/10.1073/pnas.1921046117>.
- Gabriel L, et al. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res*. 2024;34:769–777. <https://doi.org/10.1101/gr.278090.123>.
- Goel M, Sun H, Jiao W-B, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol*. 2019;20:277. <https://doi.org/10.1186/s13059-019-1911-0>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Jones P, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Li H. 2013 Mar 16. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [preprint]. *arXiv preprint arXiv:1303.3997*. <https://doi.org/10.48550/arXiv.1303.3997>
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Lu D, et al. Nanopore ultra-long sequencing and adaptive sampling spur plant complete telomere-to-telomere genome assembly. *Mol Plant*. 2024;17:1773–1786. <https://doi.org/10.1016/j.molp.2024.10.008>.
- Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27:764–770. <https://doi.org/10.1093/bioinformatics/btr011>.
- Marcus JP, Green JL, Goulter KC, Manners JM. A family of antimicrobial peptides is produced by processing of a 7S globulin protein in *Macadamia integrifolia* kernels. *Plant J*. 1999;19:699–710. doi: <https://doi.org/10.1046/j.1365-3113x.1999.00569.x>.
- Nurk S, et al. The complete sequence of a human genome. *Science*. 2022;376:44–53. <https://doi.org/10.1126/science.abj6987>.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432. <https://doi.org/10.1038/s41467-020-14998-3>.
- Rossetto M, Kooyman RM. The tension between dispersal and persistence regulates the current distribution of rare palaeo-endemic rain forest flora: a case study. *J Ecol*. 2005;93:906–917. <https://doi.org/10.1111/j.1365-2745.2005.01046.x>.
- Rossetto M, Porter C. Isolation and characterization of microsatellite loci from *Eidothea hardeniana*, an early Proteaceae. *Mol Ecol Notes*. 2005;5:106–108. <https://doi.org/10.1111/j.1471-8286.2004.00852.x>.
- Rubio-Piña JA, Zapata-Pérez O. Isolation of total RNA from tissues rich in polyphenols and polysaccharides of mangrove plants. *Electron J Biotechnol*. 2011;14:11–11. <https://doi.org/10.2225/vol14-issue5-fulltext-10>.
- Sauquet H, et al. Contrasted patterns of hyperdiversification in Mediterranean hotspots. *Proc Natl Acad Sci U S A*. 2009;106:221–225. <https://doi.org/10.1073/pnas.0805607106>.
- Seppy M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. *Gene Prediction: Methods and Protocols*. 2019:1962:227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.
- Shang L, et al. A complete assembly of the rice Nipponbare reference genome. *Mol Plant*. 2023;16:1232–1236. <https://doi.org/10.1016/j.molp.2023.08.003>.
- Sharma P, et al. The genome of the endangered *Macadamia janse-nii* displays little diversity but represents an important genetic resource for plant breeding. *Plant Direct*. 2021;5:e364. <https://doi.org/10.1002/pld3.364>.
- Sharma P, et al. Genome sequences to support conservation and breeding of *Macadamia*. *Tropical Plants*. 2024;3. <https://doi.org/10.48130/tp-0024-0029>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Šmarda P, et al. Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc Natl Acad Sci U S A*. 2014;111:E4096–E4102. <https://doi.org/10.1073/pnas.1321152111>.
- Soni A, Constantin L, Furtado A, Henry RJ. A flow cytometry protocol for measurement of plant genome size using frozen material. *Applied Biosciences*. 2025;4:28. <https://doi.org/10.3390/applbiosci4020028>.
- Soni A, Henry RJ. Re-calibration of flow cytometry standards for plant genome size estimation. *Front Plant Sci*. 2025;16:1548766. <https://doi.org/10.3389/fpls.2025.1548766>.
- Veselý P, Bureš P, Šmarda P, Pavlíček T. Genome size and DNA base composition of geophytes: the mirror of phenology and ecology? *Ann Bot*. 2011;109:65–75. <https://doi.org/10.1093/aob/mcr267>.
- Wang L, Stegemann JP. Extraction of high quality RNA from polysaccharide matrices using cetyltrimethylammonium bromide. *Biomaterials*. 2010;31:1612–1618. <https://doi.org/10.1016/j.biomaterials.2009.11.024>.
- Wang Y, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40:e49–e49. <https://doi.org/10.1093/nar/gkr1293>.

- Weston PH, Barker NP. A new suprageneric classification of the Proteaceae, with an annotated checklist of genera. *Telopea*. 2006;11:314–344. <https://doi.org/10.7751/telopea20065733>.
- Weston PH, Kooyman RM. Systematics of *Eidothea* (Proteaceae), with the description of a new species, *E. hardeniana*, from the nightcap range, north-eastern New South Wales. *Telopea*. 2002;9:821–832. <https://doi.org/10.7751/telopea20024022>.
- Wolff J, Backofen R, Grüning B. Loop detection using Hi-C data with HiCExplorer. *Gigascience*. 2022;11:giac061. <https://doi.org/10.1093/gigascience/giac061>.
- Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*. 2023;39:btac808. <https://doi.org/10.1093/bioinformatics/btac808>.
- Associate editor: Vincent Castirc