DATA NOTE

The genome sequence of fen bedstraw, *Galium uliginosum* L., 1753 (Gentianales: Rubiaceae)

[version 1; peer review: awaiting peer review]

Maarten J. M. Christenhusz[1,2], José Ignacio Márquez-Corro[1], Isabel Larridon[1], Royal Botanic Gardens Kew Genome Acquisition Lab, Plant Genome Sizing Collective, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics team, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Royal Botanic Gardens Kew, Richmond, England, UK
[2]Botanical Society of Britain and Ireland, St Albans, Hertfordshire, England, UK

v1 **First published:** 21 Sep 2026, **11**:710
https://doi.org/10.12688/wellcomeopenres.27588.1
Latest published: 21 Sep 2026, **11**:710
https://doi.org/10.12688/wellcomeopenres.27588.1

Abstract

We present a genome assembly of *Galium uliginosum* (Fen Bedstraw; Streptophyta; Magnoliopsida; Gentianales; Rubiaceae). The assembly consists of two haplotypes with total lengths of 528.28 megabases and 529.59 megabases. Most of haplotype 1 (97.84%) is scaffolded into 11 chromosomal pseudomolecules. Haplotype 2 was assembled to scaffold level. The mitochondrial sequence has a length of 189.05 kilobases and the plastid genome assembly has a length of 153.42 kilobases. Gene annotation of this assembly by Ensembl identified 25,045 protein-coding genes. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Galium uliginosum, Fen Bedstraw, genome sequence, chromosomal, Gentianales

This article is included in the Tree of Life gateway.

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.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Christenhusz MJM: Investigation, Resources; **Márquez-Corro JI**: Investigation, Resources; **Larridon I**: Investigation, Resources;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [220540, https://doi.org/10.35802/220540] and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Christenhusz MJM, Márquez-Corro JI, Larridon I *et al.* **The genome sequence of fen bedstraw, *Galium uliginosum* L., 1753 (Gentianales: Rubiaceae) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2026, **11**:710 https://doi.org/10.12688/wellcomeopenres.27588.1

First published: 21 Sep 2026, **11**:710 https://doi.org/10.12688/wellcomeopenres.27588.1

Species taxonomy

Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; eudicotyledons; Gunneridae; Pentapetalae; asterids; lamiids; Gentianales; Rubiaceae; Rubioideae; Rubieae; *Galium*; *Galium uliginosum* L., 1753 (NCBI:txid715724).

Background

Galium uliginosum L. (fen bedstraw) (Figure 1) is a perennial herb in the family Rubiaceae. Its native range extends from Europe to Mongolia and includes Morocco, and it is native in Britain and Ireland (Lambert & Meek, 2023; POWO, 2026). In Britain and Ireland, it grows in base-rich marshes and fens, generally occupying drier and more calcareous habitats than the related *G. palustre*, although the two species sometimes occur together. It is recorded from sea level to 750 m, reaching this elevation on Cross Fell in Cumberland (Lambert & Meek, 2023).

Galium uliginosum typically grows to 10–60 cm. It has weak, decumbent or ascending, four-angled stems bearing backward-pointing prickles. Its narrow, linear-lanceolate leaves are arranged in whorls of five to eight and bear recurved prickles along their margins. Small white flowers are produced in axillary panicles during July and August, followed by glabrous, wrinkled fruits approximately 1 mm long (Kliphuis, 1974).

Both diploid ($2n = 22$) and tetraploid ($2n = 44$) cytotypes have been reported, with the diploid chromosome number recorded in British material. In plants cultivated under uniform conditions, morphological characters and stomatal size did not reliably distinguish the two cytotypes (Kliphuis, 1974). The genome sequence reported here will support future comparative studies of ploidy variation in *G. uliginosum* and related species.

Methods

Sample acquisition, flow cytometry and DNA barcoding

A specimen of *G. uliginosum* (specimen ID KDTOL10676, ToLID daGalUlig1) was used for genome sequencing. It was collected from Wicken Fen, Ely, Cambridgeshire, UK (latitude 52.3113, longitude 0.2899) on 2022-09-16. The specimen was collected by Maarten Christenhusz, Jose Ignacio Marquez-Corro and Isabel Larridon, and identified by Maarten Christenhusz. The daGalUlig1 specimen was also used for Hi-C and RNA sequencing. The herbarium voucher associated with the sequenced plant is MC9443 and is deposited in the herbarium of RBG Kew (K) K001527058.

The genome size was estimated by flow cytometry following the 'one-step' method outlined in Pellicer *et al.* (2021) and using propidium iodide as the fluorochrome. LB01 was used for isolation of nuclei (Loureiro *et al.*, 2007), and the internal calibration standard was *Solanum lycopersicum* L. 'Stupické polní rané' with an assumed 1C-value of 968 Mb (Doležel *et al.*, 2007).

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). Part of the plant specimen was preserved in silica gel desiccant (Chase & Hills, 1991). DNA extracted from the dried plant was amplified by PCR for standard barcode markers, with the amplicons sequenced and compared to public sequence databases including GenBank and the Barcode of Life Database (BOLD) (Ratnasingham & Hebert, 2007). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

Nucleic acid extraction

High molecular weight (HMW) DNA extraction protocols developed by the Tree of Life Core Laboratory at the Wellcome Sanger Institute (WSI) are available on protocols.io (Howard *et al.*, 2025). Tissue from the daGalUlig1 sample was weighed and triaged to select the appropriate extraction workflow. For library DTOL13630466, 67.0 mg of whole plant tissue was homogenised using Cryoprep. HMW DNA was extracted using the Automated Plant MagAttract v.2 protocol. DNA was sheared following the Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor3 for LI PacBio protocol, and sheared DNA was purified following the Sanger Tree of Life Fragmented DNA clean up: Manual SPRI protocol.

The concentration of the sheared and purified DNA was assessed using a NanoDrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, final post-shearing DNA measurements included a Qubit concentration of 10.58 ng/μL, a yield of 476.10 ng, and an average fragment size of 11.7 kb.

Figure 1. Photograph of *G. uliginosum* by Аимаина хикари.

RNA was extracted from whole plant tissue of daGalUlig1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol. The RNA concentration was assessed using a NanoDrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Library DTOL13630466 was prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences) according to the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

Library DTOL13630466 was sequenced using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel system was in the range 40–135 pM. SMRT Link software, a PacBio web-based workflow manager, was used to set up and monitor the run, and to perform primary and secondary analysis of the data upon completion.

Hi-C

Sample preparation and crosslinking

Hi-C data were generated from 20–50 mg of the daGalUlig1 whole plant sample using the Arima-HiC v2 kit (Arima Genomics). Tissue was finely ground using the Covaris cryoPREP Dry Pulverizer (Covaris), and then subjected to nuclei isolation. Nuclei were isolated using a modified protocol based on the Qiagen QProteome Cell Compartment Kit (Qiagen), in which only the Lysis and CE2 buffers were used, with QIAshredder spin columns. After isolation, nuclei were fixed using formaldehyde to a final concentration of 2% to crosslink the DNA. The crosslinked DNA was then digested and biotinylated according to the manufacturer's instructions. A clean-up step was performed with SPRIselect

beads before library preparation. DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and the Qubit HS Assay Kit, following the manufacturer's instructions.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq X.

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA was isolated from 100 ng of total RNA using oligo (dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were prepared with an intended average fragment size of 100–300 bp. Libraries were quantified, normalised, and pooled equimolarly to a final concentration of 2.8 nM before dilution to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq X, generating paired-end reads.

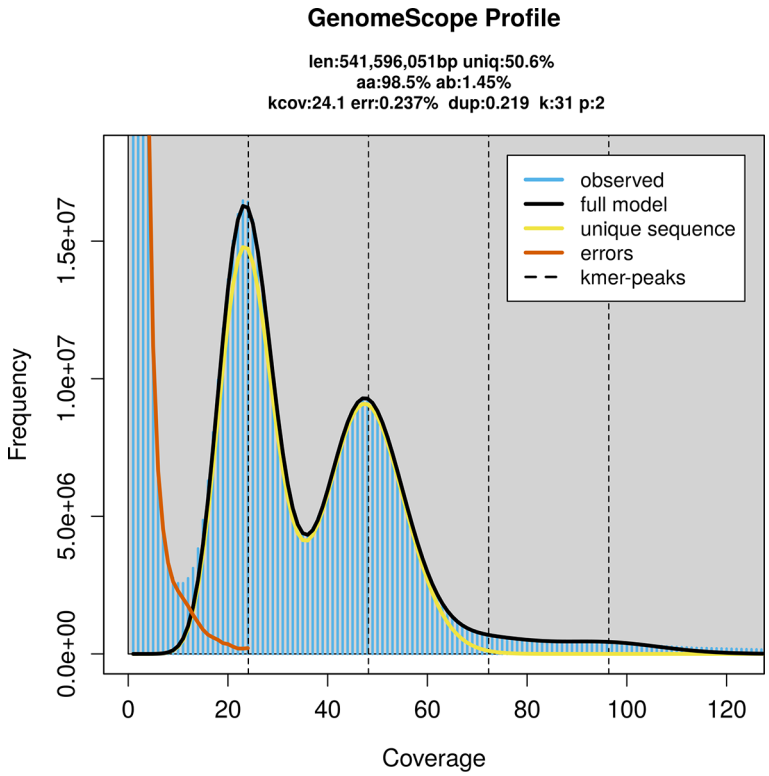

Figure 2. GenomeScope model fitted to the *k*-mer frequency distribution. The plot shows the observed and modelled *k*-mer spectra used to estimate genome size, heterozygosity and repeat content from unassembled sequencing reads.

Table 1. **Specimen and sequencing data for *G. uliginosum* (BioProject PRJEB79591).**

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	daGalUlig1	daGalUlig1	daGalUlig1
Specimen ID	KDTOL10676	KDTOL10676	KDTOL10676
BioSample (source individual)	SAMEA112287522	SAMEA112287522	SAMEA112287522
BioSample (tissue)	SAMEA112287608	SAMEA112287607	SAMEA112287607
Tissue	whole plant	whole plant	whole plant
Instrument	Sequel IIe (1 run)	Illumina NovaSeq X	Illumina NovaSeq X
Library ID & run accession(s)	DTOL13630466: ERR13649995	ERR13623232	ERR14379136
Read count total	2.72 million reads	430.19 million read pairs	46.18 million read pairs
Base count total	28.17 Gb	129.92 Gb	13.95 Gb

Table 2. **Genome assembly statistics for *G. uliginosum*.**

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	daGalUlig1.hap1.1	daGalUlig1.hap2.1
Assembly accession	GCA_965199535.1	GCA_965199655.1
Assembly level	chromosome	scaffold
Span (Mb)	528.28	529.59
Number of chromosomes	11	-
Number of contigs	432	320
Contig N50 (Mb)	7.27	6.69
Number of scaffolds	313	193
Scaffold N50 (Mb)	45.99	46.4
Longest scaffold length (Mb)	57.57	59.07
Organelles	Mitochondrial genome: 189.05 kb; Plastid genome: 153.42 kb	-

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope (version 2.1.0) (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distribution, providing fitted point estimates of genome size, heterozygosity and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MerquryFK (Rhie *et al.*, 2020). The organelle genomes were assembled using OATK (Zhou *et al.*, 2025). The comparator for the mitochondrial assembly was the mitochondrial genome of *Psychotria viridis* (NC_066984.1).

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included nine breaks and four joins. This increased the scaffold count by 1.3% and reduced

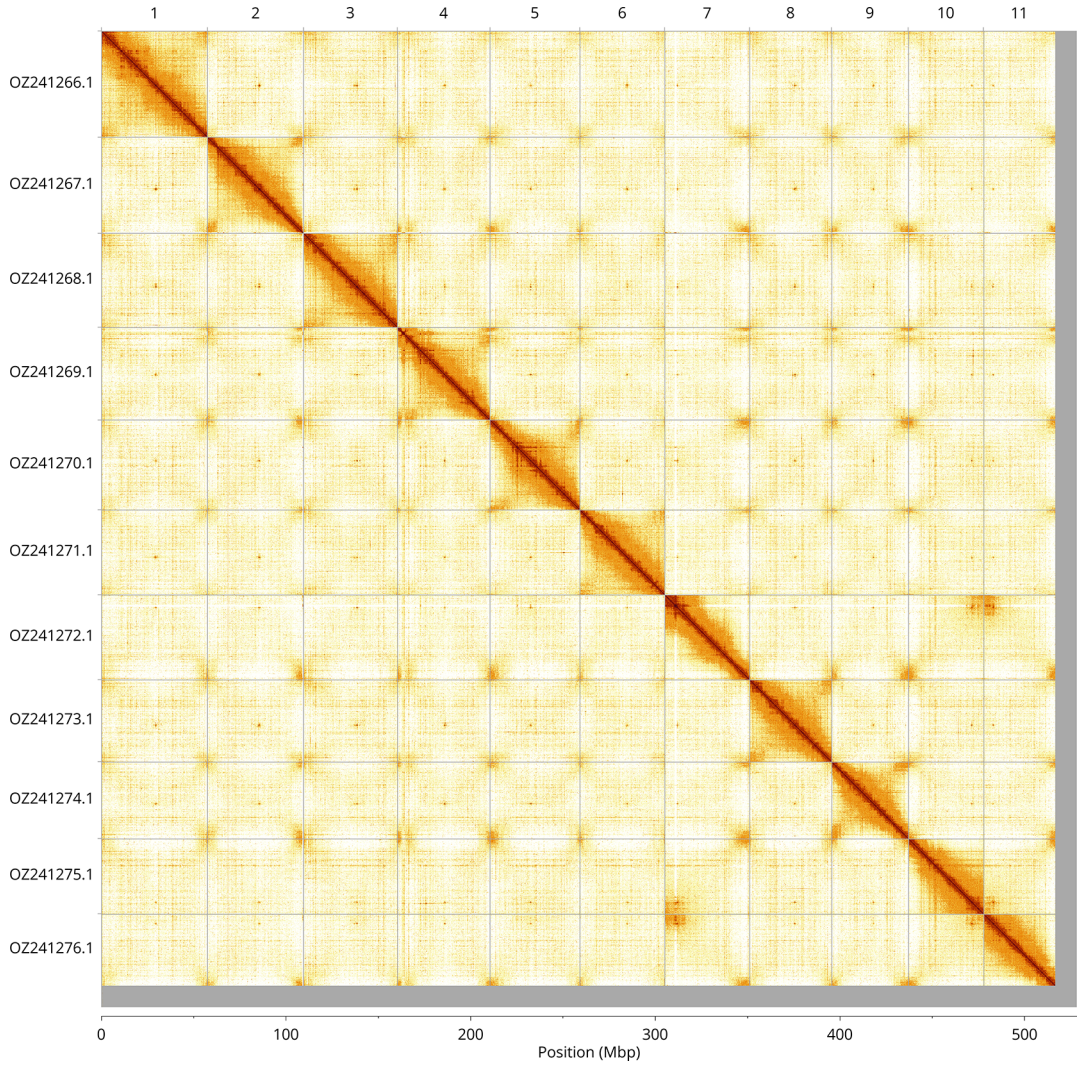

Figure 3. Hi-C contact map of the *G. uliginosum* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

the scaffold N50 by 9.6%. The curation process is described at sanger-tol/curation-resources. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The MerquryFK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer database ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). PacBio reads were aligned to the assembly using minimap2 (Li, 2018) and processed with SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. The pipeline runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three basal BUSCO lineages, eukaryota, bacteria and archaea, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND BLASTp (Buchfink *et al.*, 2021). The genome assembly is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND BLASTx. Sequences without DIAMOND BLASTx hits are extracted using seqtk and aligned to the NT database using BLASTn (Altschul *et al.*, 1990). The outputs are consolidated into a BlobDir for

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *G. uliginosum* daGalUlig1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ241266.1	1	57.57	39
OZ241267.1	2	52.01	38.50
OZ241268.1	3	50.87	38.50
OZ241269.1	4	50.15	38.50
OZ241270.1	5	48.60	39
OZ241271.1	6	45.99	39
OZ241272.1	7	45.98	39
OZ241273.1	8	44.45	38.50
OZ241274.1	9	41.58	38.50
OZ241275.1	10	40.75	38.50
OZ241276.1	11	38.93	38.50

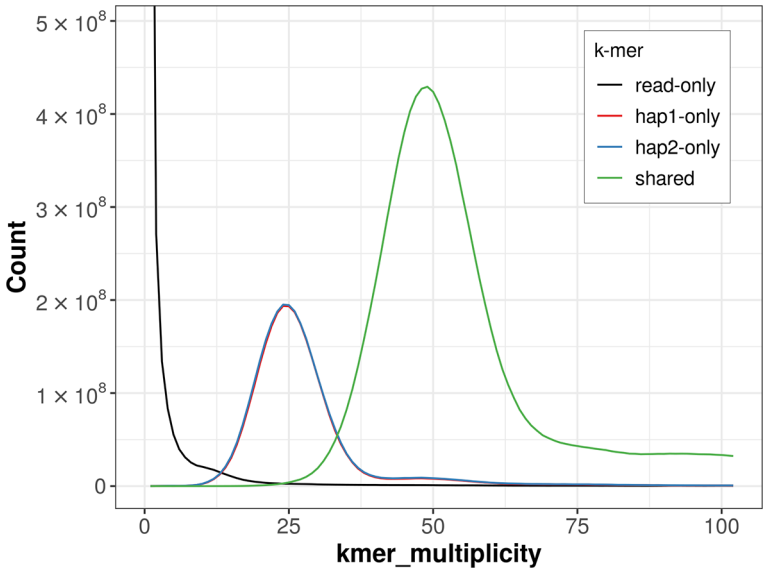

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assembly. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled, while the remaining curves correspond to *k*-mers recovered in the final assembly.

visualisation in BlobToolKit. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *G. uliginosum* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 28.17 Gb (gigabases) from 2.72 million reads, which were used to assemble the genome. GenomeScope analysis using a $p = 2$ model gave a *k*-mer-based haploid genome size estimate of 541.60 Mb, with heterozygosity of 1.45% and repeat content of 49.37%; the full-model fit was 97.68% (Figure 2). Using flow cytometry, the genome size (1C-value) of the sample was estimated to be 0.65 pg, equivalent to 640.00 Mb. These estimates guided expectations for the assembly. The sequencing data provided approximately 48× coverage. Hi-C sequencing produced 129.92 Gb from 430.19 million read pairs, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

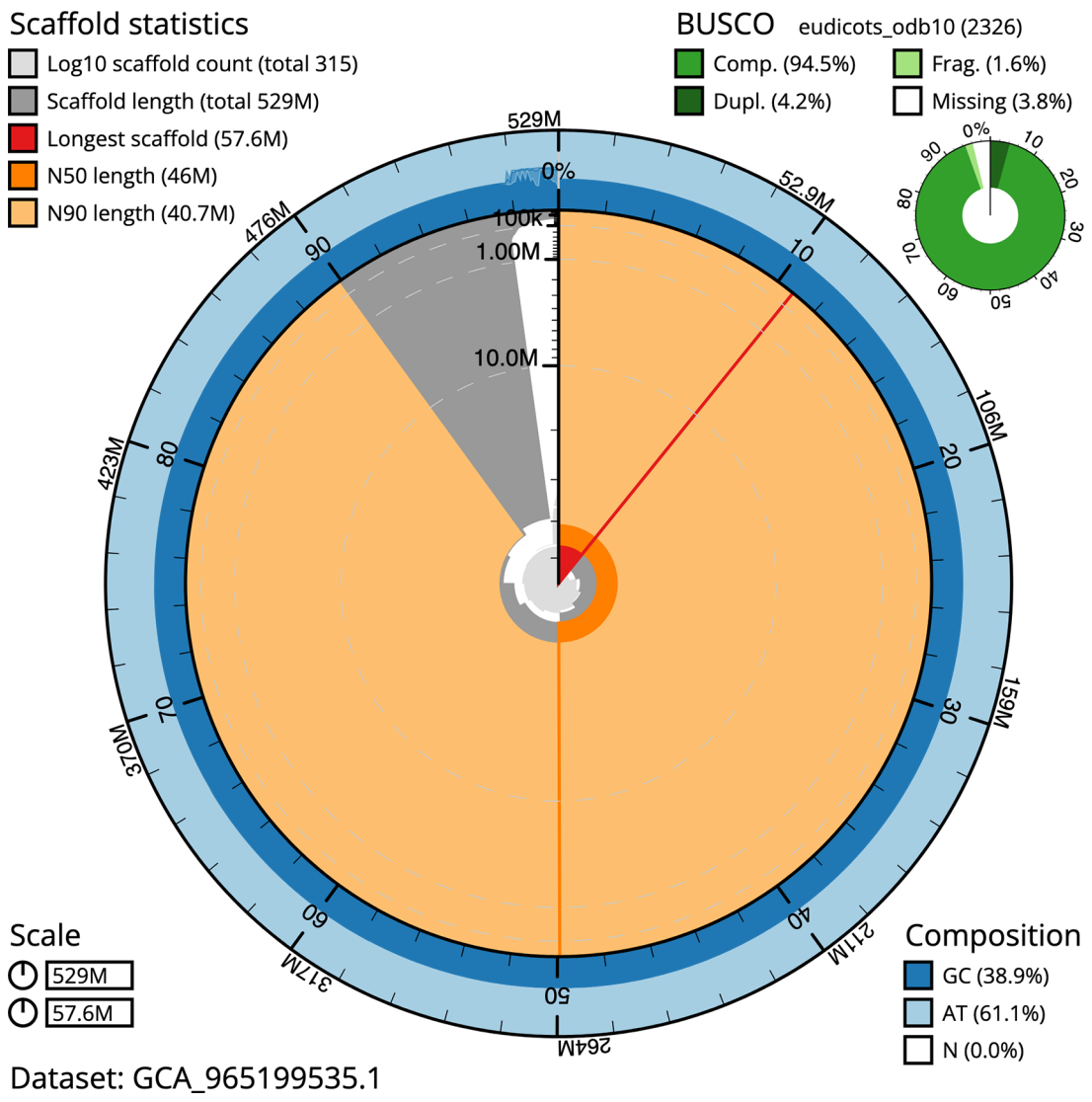

Figure 5. Assembly metrics for daGalUlig1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the eudicots_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 528.28 Mb in 313 scaffolds, with 119 gaps, and a scaffold N50 of 45.99 Mb (Table 2).

Most of the assembly sequence (97.84%) was assigned to 11 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

The mitochondrial genome (length 189.05 kb, OZ241277.1) and plastid genome (length 153.42 kb, OZ241278.1) were also assembled. These sequences are included as contigs in the multifasta file of the genome submission and as standalone records.

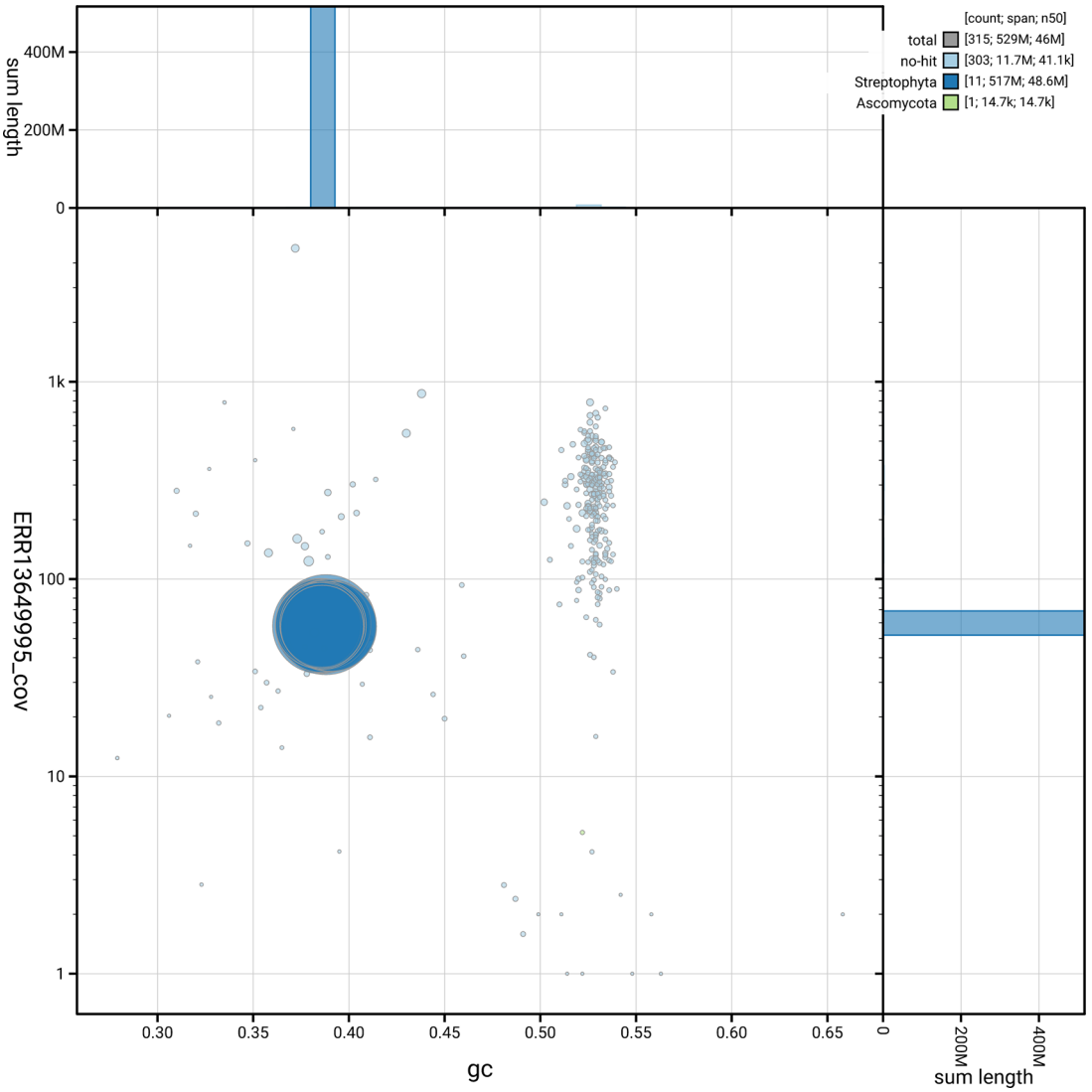

Figure 6. BlobToolKit blob plot for daGalUlig1.hap1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Assembly quality metrics

For haplotype 1, the estimated QV is 64.2, and for haplotype 2, 64.5. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.3. The *k*-mer completeness is 72.51% for haplotype 1, 73.06% for haplotype 2, and 97.78% for the combined haplotypes (Figure 4).

BUSCO analysis using the eudicots_odb10 reference set (n = 2,326) identified 94.5% of the expected gene set (single = 90.3%, duplicated = 4.2%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project (EBP) Report on Assembly Standards.

The three-part EBP metric calculated for haplotype 1 is **6.C.Q64**, meeting the 6.C.Q40 benchmark.

Table 4. Earth BioGenome Project summary metrics for the *G. uliginosum* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q64	6.C.Q40
Contig N50 length	7.27 Mb	> 1 Mb
Scaffold N50 length	45.99 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 64.2; haplotype 2: 64.5; combined: 64.3	≥ 40
k-mer completeness	Haplotype 1: 72.51%; Haplotype 2: 73.06%; combined: 97.78%	> 90%
BUSCO	Haplotype 1: C:94.5% [S:90.3%, D:4.2%], F:1.6%, M:3.8%, n:2,326; Haplotype 2: C:94.5% [S:90.2%, D:4.4%], F:1.6%, M:3.8%, n:2,326	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	97.84%	≥ 90%

Note: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO v5.7.1 using the eudicots_odb10 lineage dataset: C = complete; S = single-copy; D = duplicated; F = fragmented; M = missing; n = orthologues.

Genome annotation report

The *G. uliginosum* genome assembly (GCA_965199535.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 46,393 transcribed mRNAs from 25,045 protein-coding and 10,643 non-coding genes. The average transcript length is 2,692.35 bp, with an average of 1.30 coding transcripts per gene and 4.73 exons per transcript. For further information about the annotation, please refer to the annotation page on Ensembl.

Author information

- Members of the Royal Botanic Gardens Kew Genome Acquisition Lab
- Members of the Plant Genome Sizing Collective
- Members of the Darwin Tree of Life Barcoding collective
- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Galium uliginosum. Accession number PRJEB79591; https://identifiers.org/ena.embl/PRJEB79591. The genome sequence is released openly for reuse. The *G. uliginosum* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the Sanger Institute Tree of Life Programme (PRJEB43745). All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

Pipelines used for genome assembly at the WSI Tree of Life are available at https://pipelines.tol.sanger.ac.uk/pipelines. Table 5 lists software versions used in this study.

Table 5. Software versions and sources used for *G. uliginosum*.

Software	Version	Source
BLAST	2.14.0	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.5	https://github.com/genomehubs/blobtoolkit
blobtk	0.5.1	https://github.com/genomehubs/blobtk
BUSCO	5.7.1	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope	2.1.0	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.0-c1	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	https://github.com/lh3/minimap2
Oatk	0.9	https://github.com/c-zhou/oatk
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.6	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.1	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.4-r122	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.1.1	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J. Mol. Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin M-J, Orchard S, *et al.*: **UniProt: The Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nat. Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3: Genes, Genomes, Genetics.* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Chase MW, Hills HH: **Silica gel: An ideal material for field preservation of leaf samples for DNA studies.** *Taxon.* 1991; **40**(2): 215–220.
Publisher Full Text

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

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat. Biotechnol.* 2022; **40**(9): 1332–1335.
PubMed Abstract | Publisher Full Text | Free Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

Doležel J, Greilhuber J, Suda J: **Estimation of nuclear DNA content in plants using flow cytometry.** *Nat. Protoc.* 2007; **2**(9): 2233–2244.
PubMed Abstract | Publisher Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: Summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat. Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

Howard C, Denton A, Jackson BW, *et al.*: **On the path to reference genomes for all biodiversity: Laboratory protocols and lessons learned from processing over 2,000 species in the Sanger Tree of Life.** *GigaScience.* 2025; **14**: giaf119.
Publisher Full Text | Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Kliphuis E: **Cytotaxonomic notes on some *Galium* species. A.** *Proc. K. Ned. Akad. Wet. C.* 1974; **77**(4): 345–357.
Reference Source

Kurtzer GM, Sochat V, Bauer MW: **Singularity: Scientific containers for mobility of compute.** *PLOS ONE.* 2017; **12**(5): e0177459.
PubMed Abstract | Publisher Full Text | Free Full Text

Lambert S, Meek WR: ***Galium uliginosum* L.** Stroh PA, Humphrey TA, Burkmar RJ, *et al.*, editors. *BSBI Online Plant Atlas 2020.* Ireland: Botanical Society of Britain; 2023. UK Centre for Ecology & Hydrology.
Reference Source

Li H: **Minimap2: Pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text | Free Full Text

Loureiro J, Rodriguez E, Doležel J, *et al.*: **Two new nuclear isolation buffers for plant DNA flow cytometry: A test with 37 species.** *Ann. Bot.* 2007; **100**(4): 875–888.
PubMed Abstract | Publisher Full Text | Free Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol. Biol. Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

Merkel D: **Docker: Lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Publisher Full Text

Pellicer J, Powell RF, Leitch IJ: **The application of flow cytometry for estimating genome size, ploidy level endopolyploidy, and reproductive modes in plants.** In Besse P, editor. *Methods in Molecular Biology.* New York, NY: 2021; **2222**: 325–61.
PubMed Abstract | Publisher Full Text

POWO: ***Galium uliginosum* L.** 2026.
Reference Source

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat. Commun.* 2020; **11**(1): 1432.
PubMed Abstract | Publisher Full Text | Free Full Text

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | Publisher Full Text | Free Full Text

Ratnasingham S, Hebert PDN: **BOLD: The Barcode of Life Data System (http://www.barcodinglife.org).** *Mol. Ecol. Notes.* 2007; **7**(3): 355–364.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools.** *Database.* 2020; **2020**: baaa062.
PubMed Abstract | Publisher Full Text | Free Full Text

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project.** *Wellcome Open Research.* 2024; **9**: 339.
PubMed Abstract | Publisher Full Text | Free Full Text

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE; 2019; 314–324.
Publisher Full Text

Zhou C, Brown M, Blaxter M, *et al.*: **Oatk: A de novo assembly tool for complex plant organelle genomes.** *Genome Biol.* 2025; **26**(1): 235.
PubMed Abstract | Publisher Full Text | Free Full Text

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