DATA NOTE

The genome sequence of *Carex laevigata* Sm., 1800 (Poales: Cyperaceae)

[version 1; peer review: awaiting peer review]

Markus Ruhsam [1], José Ignacio Márquez-Corro [2],
Royal Botanic Garden Edinburgh 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 Garden Edinburgh, Edinburgh, Scotland, UK
[2]Royal Botanic Gardens Kew, Richmond, England, UK

v1 **First published:** 09 Sep 2026, **11**:616
https://doi.org/10.12688/wellcomeopenres.26864.1
Latest published: 09 Sep 2026, **11**:616
https://doi.org/10.12688/wellcomeopenres.26864.1

Abstract

We present a genome assembly of *Carex laevigata* (smooth-stalked sedge; Streptophyta; Magnoliopsida; Poales; Cyperaceae). The genome sequence has a total length of 376.54 megabases. Most of the assembly (99.74%) is scaffolded into 36 chromosomal pseudomolecules. The mitochondrial sequences have lengths of 1753.61, 43.94, 29.99, 720.76 and 185.64 kilobases and the plastid genome assembly has a length of 202.9 kilobases. 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

Carex laevigata, smooth-stalked sedge, genome sequence, chromosomal, Poales

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: Ruhsam M: Investigation, Resources, Writing – Review & Editing; **Márquez-Corro JI**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (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: Ruhsam M, Márquez-Corro JI, Royal Botanic Garden Edinburgh Genome Acquisition Lab *et al.* **The genome sequence of *Carex laevigata* Sm., 1800 (Poales: Cyperaceae) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2026, **11**:616 https://doi.org/10.12688/wellcomeopenres.26864.1

First published: 09 Sep 2026, **11**:616 https://doi.org/10.12688/wellcomeopenres.26864.1

Species taxonomy

Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; Liliopsida; Petrosaviidae; commelinids; Poales; Cyperaceae; Cyperoideae; Cariceae; *Carex*; Carex subgen. Carex; *Carex laevigata* Sm., 1800 (NCBI:txid372271).

Background

The smooth-stalked sedge, *Carex laevigata* Sm., is a perennial sedge with short rhizomes forming dense tussocks and is widely distributed across the western Mediterranean and Atlantic regions of Europe, with additional populations reported from Morocco (POWO, 2026). The species typically inhabits shady, moist woodlands, stream margins, and alder–ash forests on humid, often acidic soils, although it may occasionally occur in wet meadows (Jermy *et al.*, 2007; Stace, 2019). *Carex laevigata* produces robust culms reaching up to 120 cm in height (Figure 1) (Jermy *et al.*, 2007), contributing to the vegetation structure of humid forest understories and riparian habitats.

Carex laevigata belongs to a taxonomically complex group and has frequently been misidentified because of the lack of discontinuous characters separating it from related taxa, particularly the closely related *C. binervis* (Jermy *et al.*, 2007; Tutin *et al.*, 1980). Nevertheless, the species can generally be distinguished by its longer ligules (7–15 mm), larger utricles (4–6 mm), and characteristic gingery-brown staminate glumes. In contrast, *C. binervis* typically possesses shorter ligules, usually up to 3 mm and only rarely reaching 5 mm, smaller utricles up to 4.5 mm long, and purplish staminate glumes (Jermy *et al.*, 2007; Tutin *et al.*, 1980). The species flowers mainly from late spring to early summer and, like most members of the genus *Carex*, is primarily wind-pollinated (Reznicek, 1990).

Reproduction in *C. laevigata* occurs through both seed production and vegetative spread via short rhizomes, enabling local clonal expansion and the persistence of dense tussocks in stable woodland environments. Clonal growth is likely to contribute substantially to population persistence in woodland *Carex* species under shaded conditions and fluctuating moisture regimes (Bernard, 1990).

Among the chromosome diversity reported within *Carex* (Márquez-Corro *et al.*, 2024), *C. laevigata* exhibits substantial intraspecific chromosome-number variation, with diploid chromosome numbers ranging from 69 to 84 (Escudero &

Figure 1. Photographs of the *Carex laevigata* (lpCarLaev1) specimen from which samples were taken for genome sequencing. The panels show the habitat and habit of the plant, with details of the inflorescences and leaves.

Luceño, 2009; Márquez-Corro *et al.*, 2024). This chromosomal variation has been associated with the species' ecological preferences and latitudinal distribution, with populations occurring at higher latitudes and under colder climatic conditions generally exhibiting lower chromosome numbers (Escudero *et al.*, 2013). Subsequent phylogeographic analyses suggested that lineages characterized by lower chromosome numbers were preferentially associated with postglacial expansion into northern Europe, while also revealing correlations between chromosome variation and multiple environmental variables (Márquez-Corro *et al.*, 2024). These findings emphasize the potential role of chromosome evolution in ecological adaptation and range expansion within the genus.

We present a chromosome-level genome for *Carex laevigata,* which might provide a valuable resource for future investigations into genome variation, chromosomal evolution, and ecological adaptation across the species' distribution range. The assembly was produced using the Tree of Life pipeline from a specimen collected in Roslin Glen, Scotland, UK (Figure 1).

Methods

Sample acquisition, flow cytometry and DNA barcoding

A specimen of *Carex laevigata* (specimen ID EDTOL03990, ToLID lpCarLaev1; Figure 1) was used for genome sequencing. It was collected from Roslin Glen, Scotland, UK (latitude 55.8553, longitude −3.158) on 2022-05-27. The specimen was collected and identified by Markus Ruhsam (Royal Botanic Garden Edinburgh). The lpCarLaev1 specimen was also used for Hi-C and RNA sequencing. The herbarium voucher associated with the sequenced plant is E01152511 and is deposited in the herbarium of RBG Edinburgh (E) [https://data.rbge.org.uk/herb/E01152511].

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. The General Purpose Buffer (GPB) supplemented with 3% PVP 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

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The lpCarLaev1 sample was weighed and triaged to determine the appropriate extraction protocol. For HMW DNA extraction, 49.90 mg of leaf tissue was used. The tissue was homogenised by cryogenic disruption using the Covaris cryoPREP® Automated Dry Pulverizer. HMW DNA was extracted using the Automated Plant MagAttract v2 protocol. We used centrifuge-mediated fragmentation to produce DNA fragments in the 8–10 kb range, following the Covaris g-TUBE protocol for ultra-low input (ULI). Sheared DNA was purified by manual SPRI (solid-phase reversible immobilisation), using (Pacific Biosciences) AMPure PB beads to eliminate shorter fragments and concentrate the DNA. 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, the final post-shearing DNA had a Qubit concentration of 2.66 ng/μL and a yield of 125.02 ng, with a fragment size of 9.2 kb. The 260/280 spectrophotometric ratio was nan, and the 260/230 ratio was nan. The Genomic Quality Number (GQN) was 3.2.

RNA was extracted from leaf tissue of lpCarLaev1 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. Prior to library preparation, the DNA was fragmented to ~10 kb. Ultra-low-input (ULI) libraries were prepared using the PacBio SMRTbell® Express Template Prep Kit 2.0 and gDNA Sample Amplification Kit. Samples were normalised to 20 ng DNA. Single-strand

overhang removal, DNA damage repair, and end-repair/A-tailing were performed according to the manufacturer's instructions, followed by adapter ligation. A 0.85× pre-PCR clean-up was carried out with Promega ProNex beads.

The DNA was evenly divided into two aliquots for dual PCR (reactions A and B), both following the manufacturer's protocol. A 0.85× post-PCR clean-up was performed with ProNex beads. DNA concentration was measured using a Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with the Qubit HS Assay Kit, and fragment size was assessed on an Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit. PCR reactions A and B were then pooled, ensuring a total mass of ≥500 ng in 47.4 μl.

The pooled sample underwent another round of DNA damage repair, end-repair/A-tailing, and hairpin adapter ligation. A 1× clean-up was performed with ProNex beads, followed by DNA quantification using the Qubit and fragment size analysis using the Agilent Femto Pulse. Size selection was performed on the Sage Sciences PippinHT system, with target fragment size determined by Femto Pulse analysis (typically 4–9 kb). Size-selected libraries were cleaned with 1.0× ProNex beads and normalised to 2 nM before sequencing.

The sample was sequenced using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe 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 the leaf tissue of lpCarLaev1 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 6000.

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 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 6000, generating paired-end reads.

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. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) with the — primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the

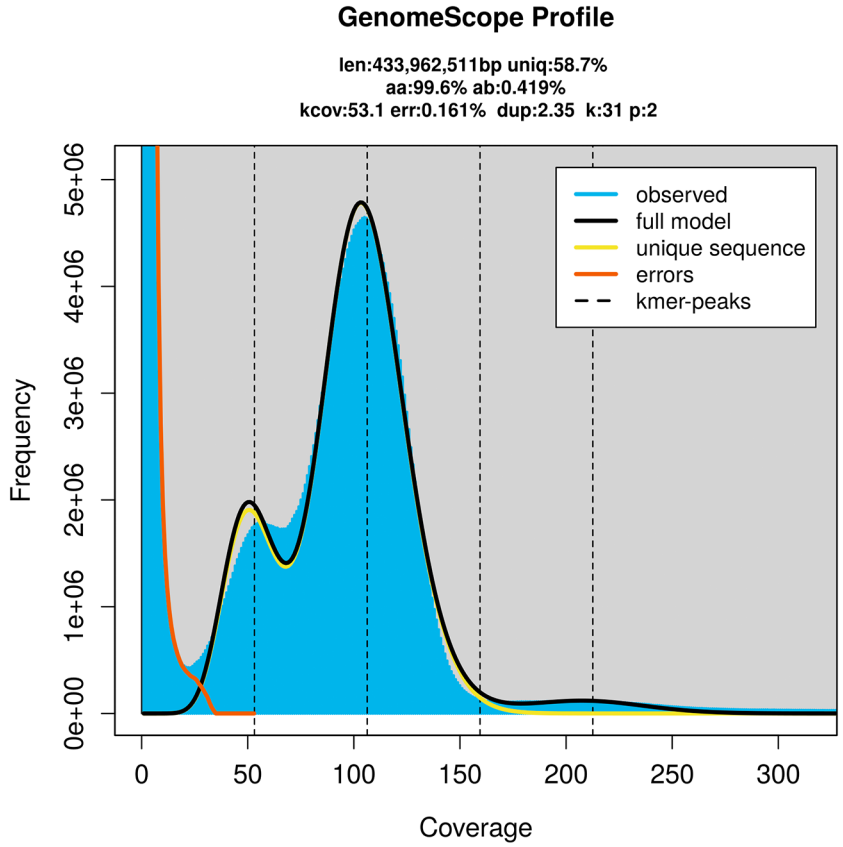

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for *Carex laevigata* (BioProject PRJEB68030).

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	lpCarLaev1	lpCarLaev1	lpCarLaev1
Specimen ID	EDTOL03990	EDTOL03990	EDTOL03990
BioSample (source individual)	SAMEA111431172	SAMEA111431172	SAMEA111431172
BioSample (tissue)	SAMEA111431215	SAMEA111431215	SAMEA111431215
Tissue	leaf	leaf	leaf
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR12205292	ERR12245623	ERR12245624
Read count total	2.24 million reads	351.10 million read pairs	37.27 million read pairs
Base count total	23.80 Gb	106.03 Gb	11.26 Gb

Table 2. **Genome assembly data for *Carex laevigata*.**

Genome assembly	Primary assembly
Assembly name	lpCarLaev1.1
Assembly accession	GCA_964214065.1
Alternate haplotype accession	GCA_964214055.1
Assembly level	chromosome
Span (Mb)	376.54
Number of chromosomes	36
Number of contigs	1 140
Contig N50	0.55 Mb
Number of scaffolds	82
Scaffold N50	10.58 Mb
Organelles	Mitochondrial genomes: 1 753.61; 43.94; 29.99; 720.76; 185.64 kb; Plastid genome: 202.90 kb

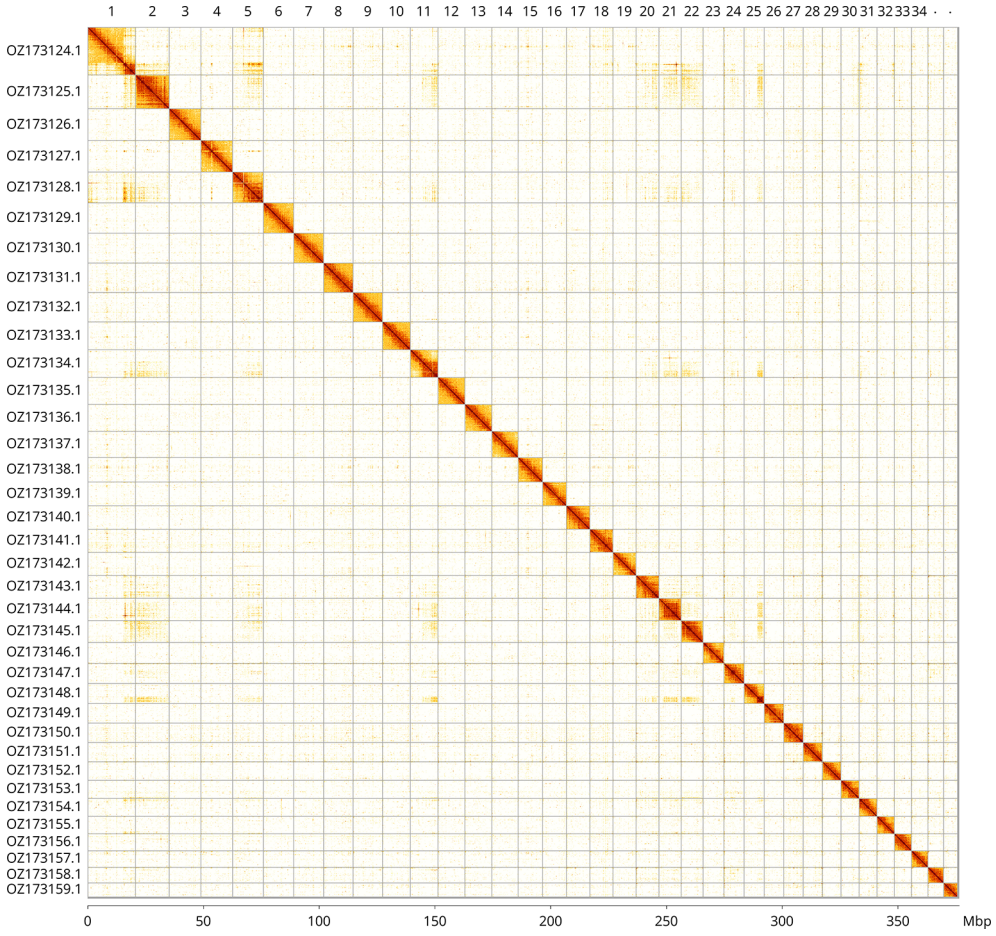

Figure 3. **Hi-C contact map of the *Carex laevigata* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Carex laevigata* lpCarLaev1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ173124.1	1	20.71	34.50
OZ173125.1	2	14.55	33.50
OZ173126.1	3	13.70	34
OZ173127.1	4	13.68	34
OZ173128.1	5	13.24	34
OZ173129.1	6	13.10	34
OZ173130.1	7	12.89	34
OZ173131.1	8	12.72	34
OZ173132.1	9	12.70	34
OZ173133.1	10	12.05	34
OZ173134.1	11	11.90	34
OZ173135.1	12	11.68	34
OZ173136.1	13	11.57	34
OZ173137.1	14	11.36	34
OZ173138.1	15	10.58	34.50
OZ173139.1	16	10.28	34
OZ173140.1	17	10.19	34
OZ173141.1	18	9.98	34.50
OZ173142.1	19	9.96	34
OZ173143.1	20	9.80	34
OZ173144.1	21	9.66	33.50
OZ173145.1	22	9.40	33.50
OZ173146.1	23	9	33.50
OZ173147.1	24	8.76	33.50
OZ173148.1	25	8.66	34
OZ173149.1	26	8.42	33.50
OZ173150.1	27	8.41	34
OZ173151.1	28	8.28	34
OZ173152.1	29	8.01	34
OZ173153.1	30	7.85	34
OZ173154.1	31	7.75	34
OZ173155.1	32	7.56	34
OZ173156.1	33	7.29	34
OZ173157.1	34	7.15	34
OZ173158.1	35	6.77	33.50
OZ173159.1	36	5.99	34

primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with 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).

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 37 breaks, 65 joins, and removal of 23 haplotypic duplications. This reduced the scaffold count by 33.8%. The curation process is described at https://gitlab.com/wtsi-grit/rapid-curation. 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 the primary and alternate 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). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified by querying NCBI datasets (O'Leary *et al.*, 2024). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome 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 with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. 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 *Carex laevigata* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 23.80 Gb (gigabases) from 2.24 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 434.63 Mb, with a heterozygosity of 0.42% and repeat content of 41.33% (Figure 2). Using flow cytometry, the genome size (1C-value) of the sample was estimated to be 0.45 pg, equivalent to 440.00 Mb. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 54× coverage. Hi-C sequencing produced 106.03 Gb from 351.10 million reads, 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.

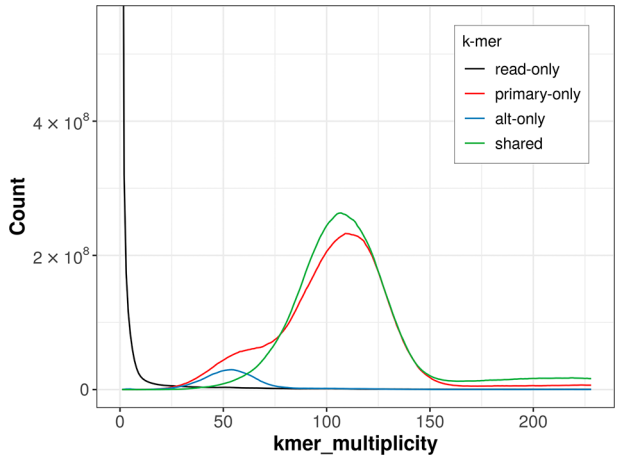

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

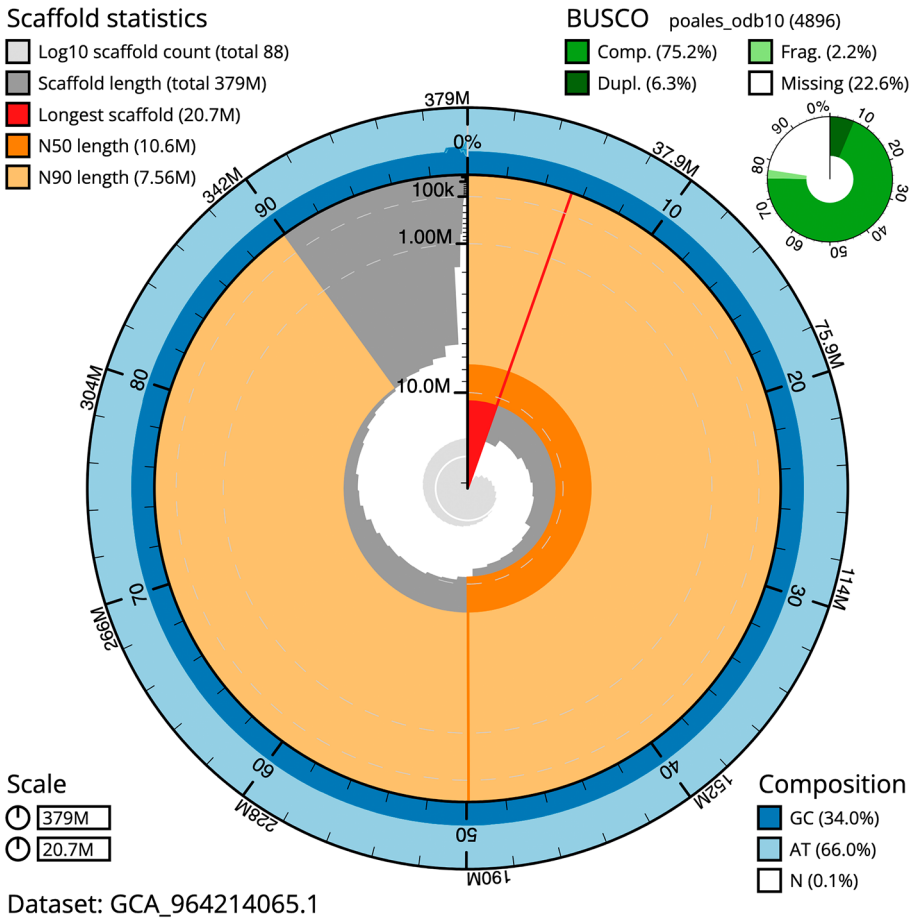

Figure 5. Assembly metrics for lpCarLaev1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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 poales_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The final assembly has a total length of 376.54 Mb in 82 scaffolds, with 1 058 gaps, and a scaffold N50 of 10.58 Mb (Table 2).

Most of the assembly sequence (99.74%) was assigned to 36 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

Multiple mitochondrial genomes (5) and the plastid genome (length 202.9 kb, OZ173165.1) were also assembled. These sequences are included as contigs in the multifasta file of the genome submission and as standalone records.

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 57.9. The *k*-mer completeness is 91.65% for the primary assembly, 51.54% for the alternate haplotype, and 97.52% for the combined assemblies (Figure 4).

BUSCO v.5.5.0 analysis using the poales_odb10 reference set (n = 4 896) identified 75.2% of the expected gene set (single = 68.9%, duplicated = 6.3%). The snail plot in Figure 5 summarises the scaffold length distribution and other

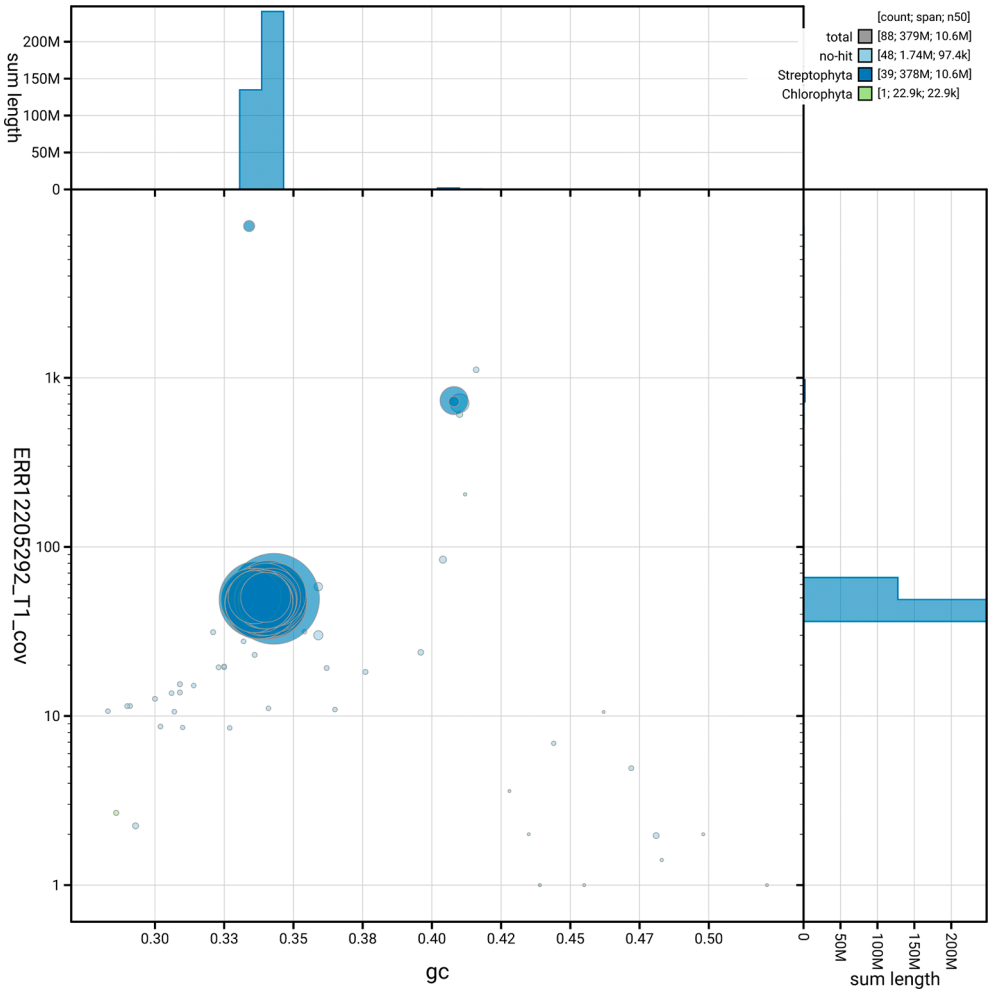

Figure 6. BlobToolKit blob plot for lpCarLaev1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total 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 statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

Table 4. Earth Biogenome Project summary metrics for the *Carex laevigata* assembly.

Measure	Value	Benchmark
EBP summary (primary)	5.C.Q58	6.C.Q40
Contig N50 length	0.55 Mb	≥ 1 Mb
Scaffold N50 length	10.58 Mb	= chromosome N50
Consensus quality (QV)	Primary: 58.5; alternate: 56.9; combined: 57.9	≥ 40
k-mer completeness	Primary: 91.65%; alternate: 51.54%; combined: 97.52%	≥ 90%
BUSCO	C:75.2% [S:68.9%, D:6.3%], F:2.2%, M:22.6%, n:4 896	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.74%	≥ 90%

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

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards January 2026. The EBP metric calculated for the primary assembly is **5.C.Q58**.

Author information

- Members of the Royal Botanic Garden Edinburgh 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: Carex laevigata. Accession number PRJEB68030; https://identifiers.org/ena.embl/PRJEB68030. The genome sequence is released openly for reuse. The *Carex laevigata* 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. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Tables 1 and 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 *Carex laevigata*.

Software	Version	Source
BLAST	2.14.0	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	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	1.2	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.5-r587	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	2.24-r1122	https://github.com/lh3/minimap2
Oatk	1.0	https://github.com/c-zhou/oatk
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.4	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
samtools	1.16.1; 1.17; 1.19.2; 1.20	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	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	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.1a.2	https://github.com/c-zhou/yahs

References

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *Journal of Molecular Biology.* 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 Research.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Bernard JM: **Life history and vegetative reproduction in *Carex*.** *Canadian Journal of Botany.* 1990; **68**(7): 1441–1448.
Publisher Full Text

Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nature 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.** *Nature Methods.* 2021; **18**(2): 170–175.
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.** *Nature Protocols.* 2007; **2**(9): 2233–2244.
PubMed Abstract | Publisher Full Text

Escudero M, Luceño M: **Systematics and evolution of *Carex* sects. Spirostachyae and Elatae (Cyperaceae).** *Plant Systematics and Evolution.* 2009; **279**(1–4): 163–89.
Publisher Full Text, 189

Escudero M, Maguilla E, Luceño M: **Selection by climatic regime and neutral evolutionary processes in holocentric chromosomes (*Carex* gr. *laevigata*: Cyperaceae): A microevolutionary approach.** *Perspectives in Plant Ecology, Evolution and Systematics.* 2013; **15**(2): 118–129.
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.** *Nature Biotechnology.* 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

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

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: Lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

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

Jermy AC, Simpson DA, Foley MJY, *et al.*: *Sedges of the British Isles.* London: Botanical Society of the British Isles; 2007.

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

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

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.** *Annals of Botany.* 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.** *Molecular Biology and Evolution.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

Márquez-Corro JI, Martín-Bravo S, Blanco-Pastor JL, *et al.*: **The holocentric chromosome microevolution: From phylogeographic patterns to genomic associations with environmental gradients.** *Molecular Ecology.* 2024; **33**(24): e17156.
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

O'Leary NA, Cox E, Holmes JB, *et al.*: **Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets.** *Scientific Data.* 2024; **11**(1): 732.
PubMed Abstract | Publisher Full Text | Free 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.** Besse P, editor. *Methods in Molecular Biology.* New York, NY: 2021; Vol. **2222**: 325–61.
PubMed Abstract | Publisher Full Text

POWO: **Plants of the World Online.** 2026.
Reference Source

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nature Communications.* 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).** *Molecular Ecology Notes.* 2007; **7**(3): 355–364.
PubMed Abstract | Publisher Full Text | Free Full Text

Reznicek AA: **Evolution in sedges (*Carex*, Cyperaceae).** *Canadian Journal of Botany.* 1990; **68**(7): 1409–1432.
Publisher 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 Biology.* 2020; **21**(1): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

Stace CA: *New Flora of the British Isles.* Middlewood Green: C & M Floristics; 2019.

Tutin TG, Heywood VH, Burges NA, *et al.*: **Flora Europaea** *Alismataceae to Orchidaceae (Monocotyledones).* Cambridge: Cambridge University Press; 1980; **5**.

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; 314–24.
Publisher Full Text

Zhou C, Brown M, Blaxter M, *et al.*: **Oatk: A de novo assembly tool for complex plant organelle genomes.** *Genome Biology.* 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