Check for updates

DATA NOTE

The genome sequence of the starved wood-sedge, *Carex depauperata* Curtis ex Woodw. (Poales: Cyperaceae)

[version 1; peer review: awaiting peer review]

Maarten J. M. Christenhusz [ID][1,2], José Ignacio Márquez-Corro [ID][1],
Michael F. Fay [ID][1,3], Royal Botanic Gardens Kew Genome Acquisition Lab,
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]Curtin University, Perth, Western Australia, Australia
[3]University of Western Australia, Crawley, Western Australia, Australia

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.

Abstract

We present a genome assembly of *Carex depauperata* (starved wood-sedge; Streptophyta; Magnoliopsida; Poales; Cyperaceae). The genome sequence has a total length of 702.60 megabases. Most of the assembly (99.61%) is scaffolded into 22 chromosomal pseudomolecules. Eight unique mitochondrial sequences and a plastid genome assembly with a length of 215.32 kilobases were also assembled. 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 depauperata; starved wood-sedge; genome sequence; chromosomal; Poales

This article is included in the Tree of Life gateway.

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

Author roles: Christenhusz MJM: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Márquez-Corro JI**: Writing – Original Draft Preparation; **Fay MF**: Writing – Review & Editing;

How to cite this article: Christenhusz MJM, Márquez-Corro JI, Fay MF *et al.* **The genome sequence of the starved wood-sedge, *Carex depauperata* Curtis ex Woodw. (Poales: Cyperaceae) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2026, **11**:165 https://doi.org/10.12688/wellcomeopenres.26113.1

First published: 18 Mar 2026, **11**:165 https://doi.org/10.12688/wellcomeopenres.26113.1

Species taxonomy
Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; Liliopsida; Petrosaviidae; commelinids; Poales; Cyperaceae; Cyperoideae; Cariceae; *Carex*; *Carex* subgen. *Carex*; *Carex depauperata* Curtis ex Woodw. (NCBI:txid234461)

Background
The starved wood-sedge, *Carex depauperata* Curtis ex Woodw., is a perennial, rhizomatous plant that forms loose tufts or tussocks with stems up to around 100 cm tall. The species is found mostly in western and southern Europe, from Ireland to the Caucasus and in Central Asia (Royal Botanic Gardens, Kew, 2026). It inhabits open areas in dry deciduous woodlands, woodland margins and disturbed banks on calcareous soil (Jermy *et al.*, 2007). *Carex depauperata* may be similar in appearance to its distantly related *C. sylvatica* Huds., but its inflorescences bear fewer utricles -hence, depauperate- that are 7–9 cm long vs. 3–5 cm in the also larger and pendulous inflorescences of *C. sylvatica* (Jermy *et al.*, 2007).

This species is considered Endangered (criterion D) in Britain, as there are only two populations of around 60 individuals left in England (Stroh *et al.*, 2014), and a single population in Ireland (Jermy *et al.*, 2007). It has always been extremely rare and has only ever been recorded from 13 sites. It is now only found in Somerset, at two sites in Surrey (one a re-introduction) and a recent re-introduction site in Dorset (Rumsey & Crouch, 2023).

The decline of native woodlands, accompanied by a lack of disturbance and an increase of competition, has brought *C. depauperata* to the brink of extinction in Britain and Ireland, as the species requires partial shade, and woodland management providing repeated opening of the canopy has been recommended to benefit *C. depauperata* for all the sites (Rich & Birkinshaw, 2001). The critical status of the species led to its inclusion on Schedule 8 of the UK Wildlife and Countryside Act 1981, and on Schedule A of the Flora (Protection) Order, 2022, in Ireland.

Genetic fingerprinting of the British and Irish populations (Fay *et al.*, 2003) has shown that the British populations are genetically close to those from France and Spain, whereas the Irish sample in that study was a genetic outlier. *Carex depauperata* is part of a small lineage that presents a diverse species assemblage, having distinct morphologies and disjunct distributions (Roalson *et al.*, 2021). Therefore, this high-quality genome assembly of *C. depauperata*, together with other recent publications focused on sedges (Kim *et al.*, 2023; Planta *et al.*, 2022; Qu *et al.*, 2022) may help to enhance our understanding of the genetic diversity in this species and will allow us to better understand the importance of these species in an ecological and evolutionary context.

Methods
Sample acquisition, flow cytometry and DNA barcoding
A specimen of *Carex depauperata* (specimen ID KDTOL10087, ToLID lpCarDepa1; Figure 1) was used for genome sequencing. It was collected from Rock Garden, Royal Botanic Gardens Kew, Surrey, UK (latitude 51.4815, longitude – 0.2896) on 2020-09-03. The specimen was collected and identified by Maarten Christenhusz (Royal Botanic Gardens Kew). The herbarium voucher associated with the sequenced plant is *MC9082* and is deposited in the herbarium of RBG Kew (K) (K001400681).

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 and 0.08% (v/v) beta-mercaptoethanol was used for isolation of nuclei (Loureiro *et al.*, 2007), and the internal calibration standard was *Petroselinum crispum* 'Champion Moss Curled' with an assumed 1C-value of 2 200 Mb (Obermayer *et al.*, 2002).

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 lpCarDepa1 sample was weighed and triaged to determine the appropriate extraction protocol. Leaf tissue was homogenised by cryogenic disruption using the Covaris cryoPREP® Automated Dry Pulverizer. HMW DNA was extracted using the Automated Plant MagAttract v2

Figure 1. Photographs of the *Carex depauperata* (lpCarDepa1) plant from which samples were taken for genome sequencing.

protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation), using AMPure PB beads (Pacific Biosciences) and the Thermo Fisher KingFisher™ Apex 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.

RNA was extracted from leaf tissue of lpCarDepa1 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. Libraries were 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 (ThermoFisher 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.

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. The SMRT link software, a PacBio web-based end-to-end 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 lpCarDepa1 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 in the total RNA solution was isolated using oligo (dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq 6000 to generate 150-bp 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 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 MERQURY.FK (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 five breaks and seven joins. This reduced the scaffold count by 20.2% and reduced the scaffold N50 by 3.0%. The curation process is documented 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 Merqury.FK 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 databases ($k = 31$) 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 from NCBI Taxonomy (Schoch *et al.*, 2020). 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 package management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *Carex depauperata* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 26.23 Gb (gigabases) from 1.55 million reads, which were used to assemble the genome. Genome-Scope2.0 analysis estimated the haploid genome size at 723.71 Mb, with a heterozygosity of 0.16% and repeat content of 43.42% (Figure 2). Using flow cytometry, the genome size (1C-value) of the sample was estimated to be 0.88 pg, equivalent to 860.00 Mb. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately $33\times$ coverage. Hi-C sequencing produced 155.61 Gb from 1 030.54 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.
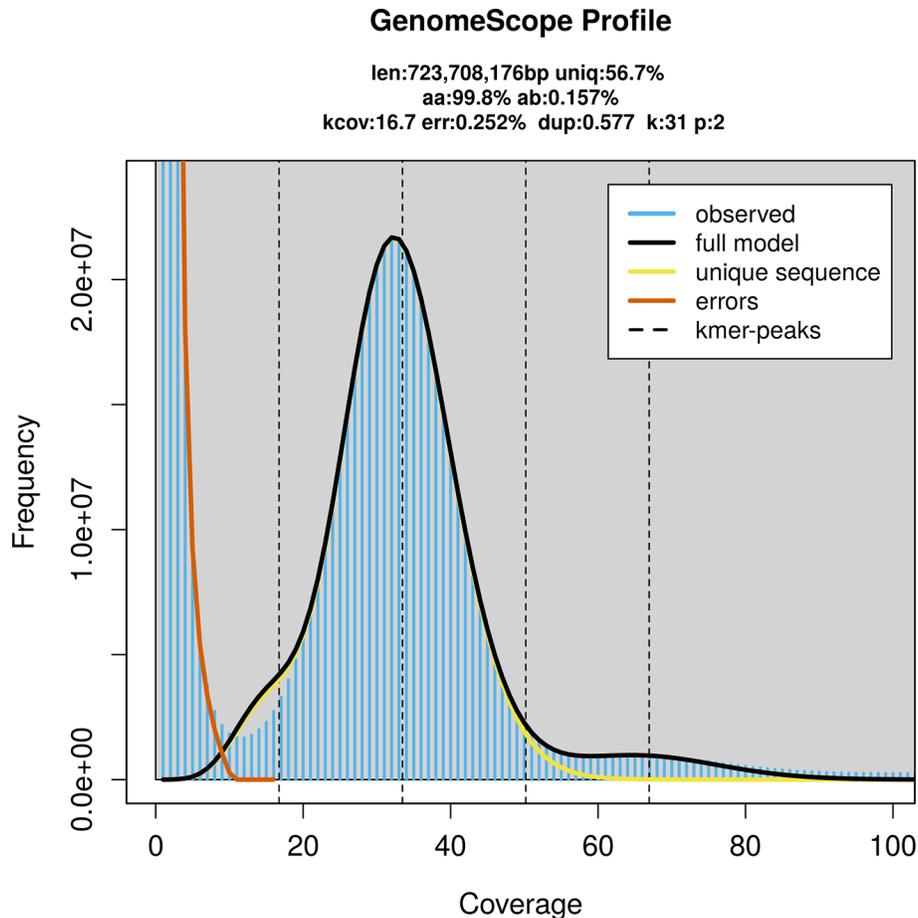
GenomeScope Profile

len:723,708,176bp uniq:56.7%
aa:99.8% ab:0.157%
kcov:16.7 err:0.252% dup:0.577 k:31 p:2

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.

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 702.60 Mb in 70 scaffolds, with 283 gaps, and a scaffold N50 of 32.58 Mb (Table 2).

Most of the assembly sequence (99.61%) was assigned to 22 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

The mitochondrial and plastid genomes 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 70.0. The k-mer completeness is 98.77% for the primary assembly, 2.35% for the alternate haplotype, and 98.84% for the combined assemblies (Figure 4).

BUSCO v.5.5.0 analysis using the poales_odb10 reference set (n = 4 896) identified 75.6% of the expected gene set (single = 68.2%, duplicated = 7.3%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric calculated for the primary assembly is **6.C.Q70**, meeting the recommended reference standard.

Table 1. Specimen and sequencing data for BioProject PRJEB50879.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	lpCarDepa1	lpCarDepa1	lpCarDepa1
Specimen ID	KDTOL10087	KDTOL10087	KDTOL10087
BioSample (source individual)	SAMEA7522407	SAMEA7522407	SAMEA7522407
BioSample (tissue)	SAMEA7522491	SAMEA7522488	SAMEA7522492
Tissue	leaf	leaf	leaf
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR8705861; ERR8705862	ERR8702787	ERR11606285
Read count total	1.55 million	1 030.54 million	59.93 million
Base count total	26.23 Gb	155.61 Gb	9.05 Gb

Table 2. Genome assembly statistics.

|	
Assembly name	lpCarDepa1.1
Assembly accession	GCA_964197715.1
Alternate haplotype accession	GCA_964197725.1
Assembly level	chromosome
Span (Mb)	702.60
Number of chromosomes	22
Number of contigs	353
Contig N50	3.88 Mb
Number of scaffolds	70
Scaffold N50	32.58 Mb
Organelles	Mitochondrial sequences: 131.78; 24.82; 57.14; 165.49; 132.12; 1122.02; 23.34 and 64.22 kb; Plastid: 215.32 kb
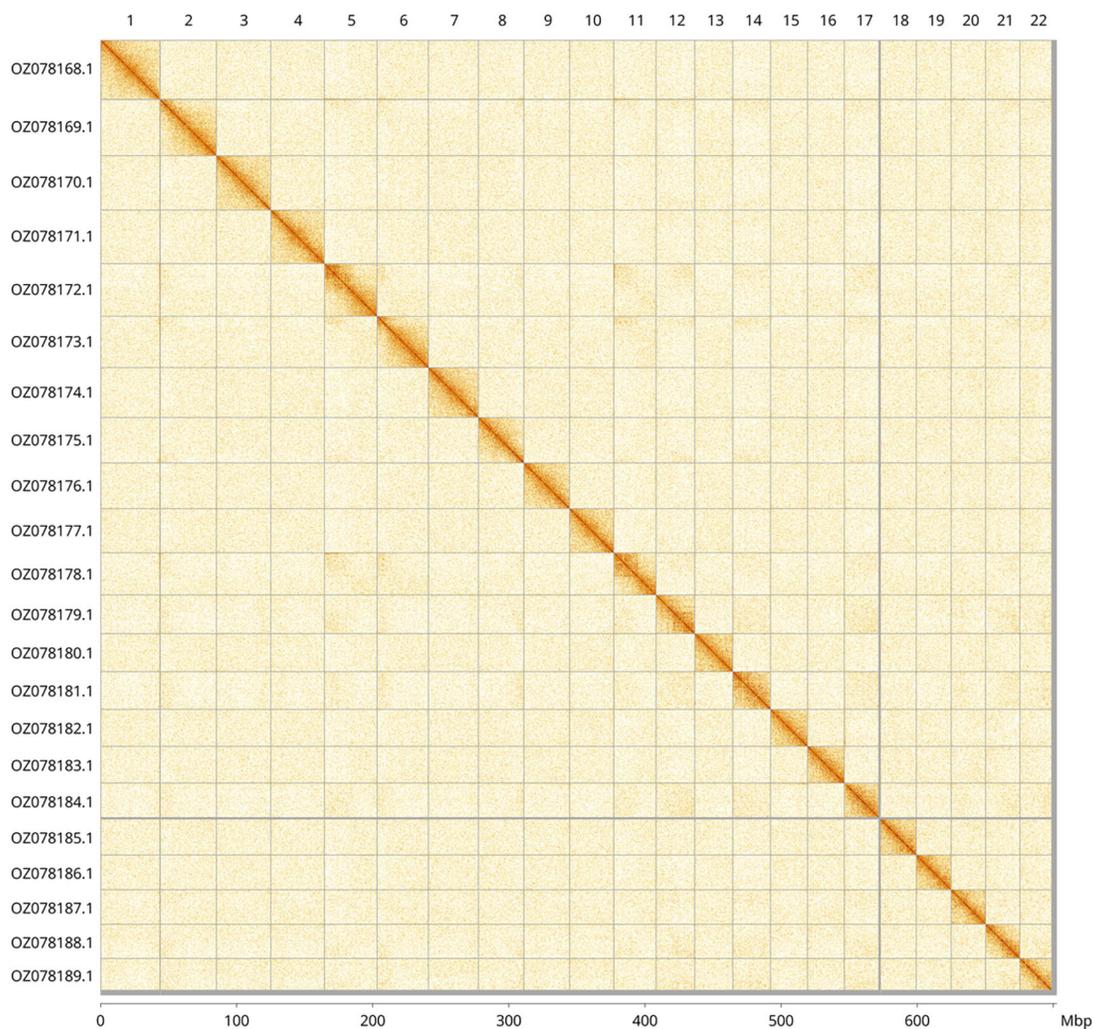
Figure 3. **Hi-C contact map of the *Carex depauperata* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

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

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Carex depauperata* lpCarDepa1.

| INSDC accession	Molecule	Length (Mb)	GC%
OZ078168.1	1	43.85	36
OZ078169.1	2	41.43	36
OZ078170.1	3	40.03	35.50
OZ078171.1	4	39.64	35.50
OZ078172.1	5	38.68	35.50
OZ078173.1	6	37.89	35.50
OZ078174.1	7	36.80	35.50
OZ078175.1	8	33.60	35.50
OZ078176.1	9	33.58	35.50
OZ078177.1	10	32.58	36
OZ078178.1	11	31.13	35.50
OZ078179.1	12	28.51	35
OZ078180.1	13	28.03	36
OZ078181.1	14	27.81	35
OZ078182.1	15	27.27	35.50
OZ078183.1	16	27.11	35.50
OZ078184.1	17	25.57	35
OZ078185.1	18	26.47	35.50
OZ078186.1	19	25.72	36
OZ078187.1	20	25.39	35.50
OZ078188.1	21	25.34	35.50
OZ078189.1	22	23.40	35.50
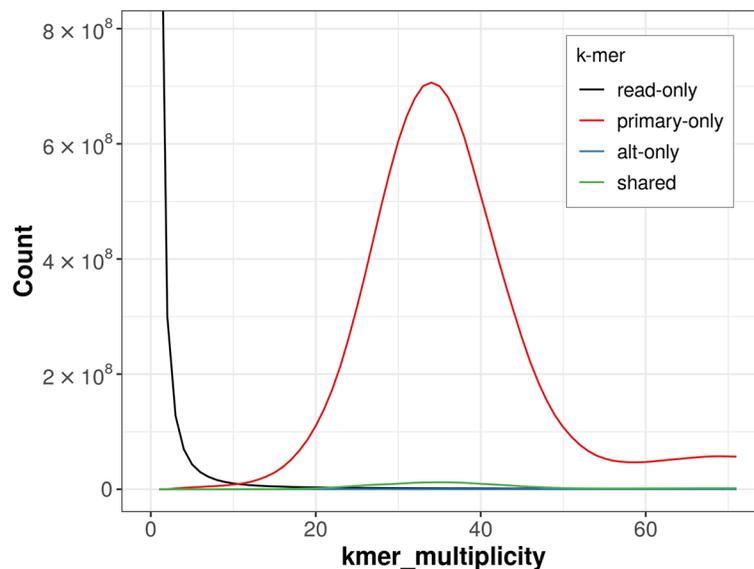
Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.
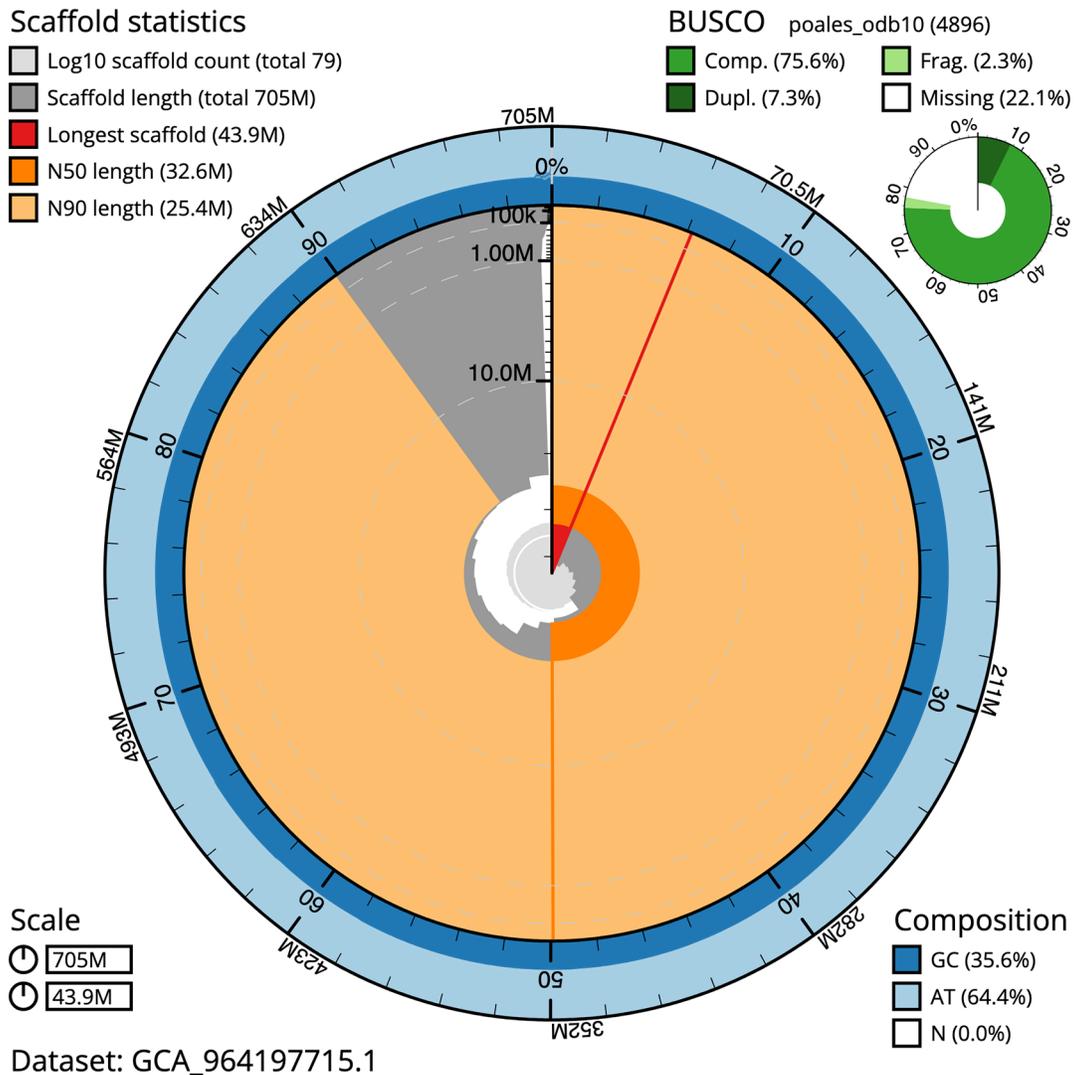
Figure 5. Assembly metrics for lpCarDepa1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue 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.

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:
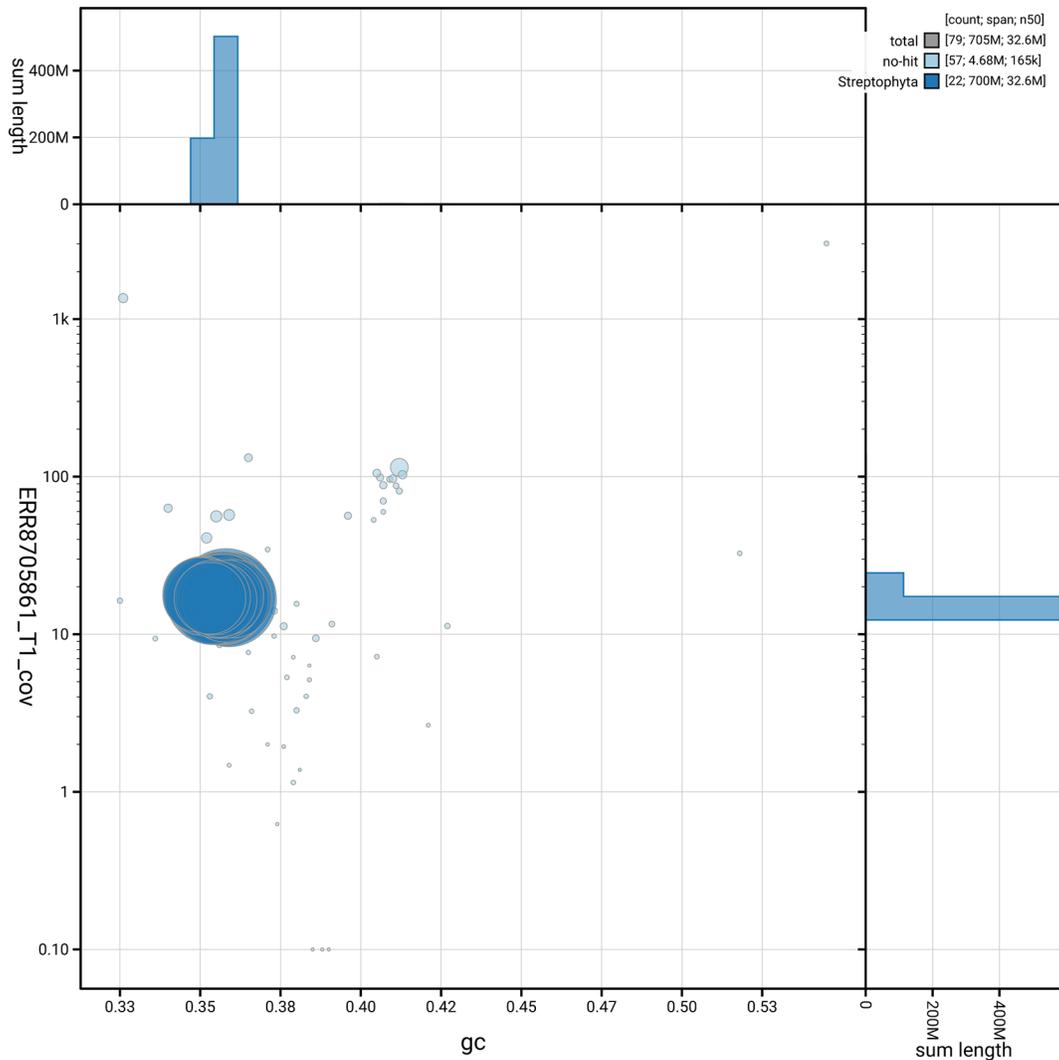
Figure 6. BlobToolKit GC-coverage plot for lpCarDepa1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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.

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

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q70	6.C.Q40
Contig N50 length	3.88 Mb	≥ 1 Mb
Scaffold N50 length	32.58 Mb	= chromosome N50
Consensus quality (QV)	Primary: 70.9; alternate: 59.6; combined: 70.0	≥ 40
k-mer completeness	Primary: 98.77%; alternate: 2.35%; combined: 98.84%	≥ 95%
BUSCO	C:99.2% [S:89.4%, D:9.8%], F:0.4%, M:0.4%, n:255	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.77%	≥ 90%

Notes: 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 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://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
Cooler	0.8.11	https://github.com/open2c/cooler
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
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	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	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
samtools	1.19.2	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
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

- 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 depauperata. Accession number PRJEB50879. The genome sequence is released openly for reuse. The *Carex depauperata* 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.

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**

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

Fay MF, O'Rourke A, Rich TCG: **A preliminary investigation of genetic variation in Western European *Carex depauperata* Curtis ex With. (Cyperaceae), Starved Wood-sedge.** *Watsonia.* 2003; **24**: 507–511.
Reference Source

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**

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: Sustainable and comprehensive software distribution for the life sciences.** *Nat. Methods.* 2018; **15**(7): 475–476.
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).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Jermy AC, Simpson DA, Foley MJY, *et al.*: *Sedges of the British Isles: BSBI Handbook No. 1.* BSBI; 2007.

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**

Kim K-R, Yu J-N, Hong JM, *et al.*: **Genome assembly and microsatellite marker development using Illumina and PacBio sequencing in *Carex pumila* (Cyperaceae) from Korea.** *Genes.* 2023; **14**(11): 2063.
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**

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).
Publisher Full Text

Obermayer R, Leitch IJ, Hanson L, *et al.*: **Nuclear DNA C-values in 30 species double the familial representation in pteridophytes.** *Ann. Bot.* 2002; **90**(2): 209–217.
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

Planta J, Liang Y-Y, Xin H, *et al.*: **Chromosome-scale genome assemblies and annotations for Poales species *Carex cristatella*, *Carex scoparia*, *Juncus effusus*, and *Juncus inflexus*.** *G3: Genes|Genomes|Genetics.* 2022; **12**(10): jkac211.
Publisher Full Text

Qu G, Bao Y, Liao Y, *et al.*: **Draft genome assembly and annotation of *Carex parvula* and *Carex kokanica* reveals stress-specific genes.** *Sci. Rep.* 2022; **12**(1): 4970.
Publisher Full Text

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.** *Mol. Ecol. Notes.* 2007; **7**(3): 355–364.
Reference Source . | **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).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rich TCG, Birkinshaw CR: **Conservation of Britain's biodiversity: *Carex depauperata* With. (Cyperaceae), Starved Wood-sedge.** *Watsonia.* 2001; **23**: 401–411.
Reference Source

Roalson EH, Jiménez-Mejias P, Hipp AL, *et al.*: **A framework infrageneric classification of *Carex* (Cyperaceae) and its organizing principles.** *J. Syst. Evol.* 2021; **59**(4): 726–762.
Publisher Full Text

Royal Botanic Gardens, Kew: **Plants of the World Online.** 2026.
Reference Source

Rumsey FJ, Crouch HJ: **Somerset Rare Plant Register account: *Carex depauperata*.** 2023.
Reference Source

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**

Stroh PA, Leach SJ, August TA, *et al.*: *A Vascular Plant Red List for England.* Bristol: Botanical Society of Britain; Ireland; 2014.

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project.** *Wellcome Open Res.* 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 Biol.* 2025; **26**(235).
Publisher Full Text

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