Check for updates

GENOME NOTE

The genome sequence of spiked sedge, *Carex spicata* Huds. (Poales: Cyperaceae)

[version 1; peer review: 1 approved]

Maarten J. M. Christenhusz[1,2], Paul J. Kersey[3], José Ignacio Márquez-Corro[1], Ilia J. Leitch[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, Surrey, England, UK
[2]Botanical Society of Britain and Ireland, St Albans, Hertfordshire, England, UK
[3]Royal Botanic Garden Edinburgh, Edinburgh, Scotland, UK

v1 **First published:** 27 Nov 2026, **11**:1126
https://doi.org/10.12688/wellcomeopenres.28543.1
Latest published: 27 Nov 2026, **11**:1126
https://doi.org/10.12688/wellcomeopenres.28543.1

Abstract

We present a genome assembly of *Carex spicata* (spiked sedge; Streptophyta; Magnoliopsida; Poales; Cyperaceae). The primary assembly has a total length of 385.54 megabases. Most of the primary assembly (99.84%) is scaffolded into 29 chromosomal pseudomolecules. The mitochondrial genome was also assembled. The plastid genome assembly has a length of 183.59 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 spicata, spiked sedge, genome sequence, chromosomal, Poales

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓

	1
version 1 27 Nov 2026	✓ view

1. **Automated Benchmarking**, Wellcome Sanger Institute, Hinxton, UK

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; **Kersey PJ**: Investigation, Resources; **Márquez-Corro JI**: Investigation, Resources; **Leitch IJ**: 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.

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

How to cite this article: Christenhusz MJM, Kersey PJ, Márquez-Corro JI *et al.* **The genome sequence of spiked sedge, *Carex spicata* Huds. (Poales: Cyperaceae) [version 1; peer review: 1 approved]** Wellcome Open Research 2026, **11**:1126 https://doi.org/10.12688/wellcomeopenres.28543.1

First published: 27 Nov 2026, **11**:1126 https://doi.org/10.12688/wellcomeopenres.28543.1

Species taxonomy

Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; Liliopsida; Petrosaviidae; commelinids; Poales; Cyperaceae; Cyperoideae; Cariceae; *Carex*; *Carex subgen. Vignea*; *Carex spicata* Huds. (NCBI:txid240682)

Background

Carex spicata Huds. (spiked sedge) grows in dense tufts with narrow green leaves and purple-tinged roots. The flowering stems carry three to eight spikes, usually close together. Each spike bears male flowers at the tip and female flowers below (Jermy *et al.*, 2007). In Britain and Ireland, it grows in rough grassland, on road verges and along hedgebanks, typically on moist, heavy soils that are neutral or slightly base-rich (Porter & Foley, 2023). Its native range includes Madeira and Europe and extends to south-western Siberia and north-western Iran. It has been introduced to parts of North America, the Russian Far East and New Zealand's South Island (POWO, 2026).

We sequenced the genome of *Carex spicata* as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity. The sequenced specimen was collected at Radbrook Common, Wytham Woods, Oxfordshire, UK (Figure 1).

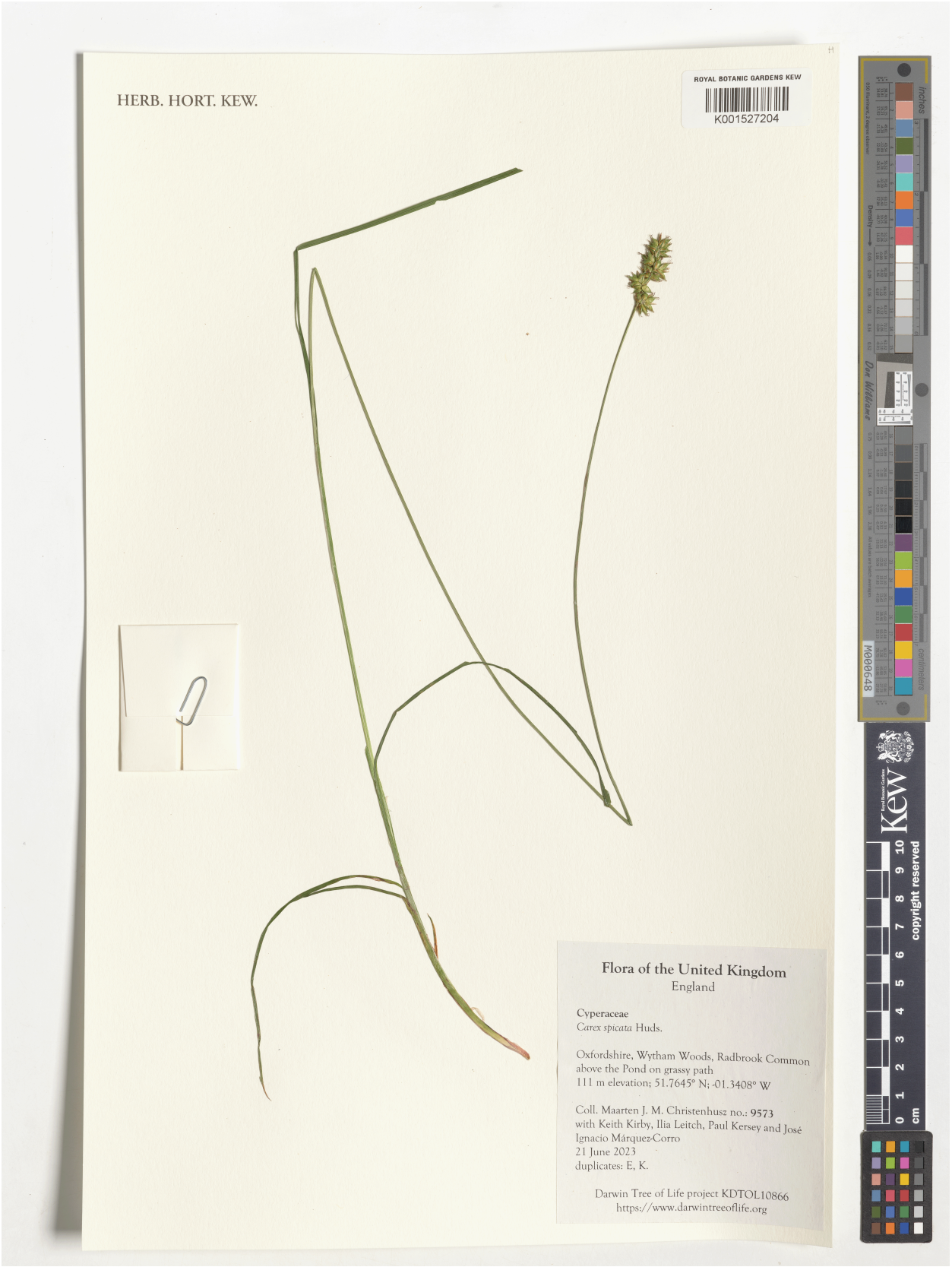
HERB. HORT. KEW.
ROYAL BOTANIC GARDENS KEW
K001527204
Flora of the United Kingdom
England
Cyperaceae
Carex spicata Huds.
Oxfordshire, Wytham Woods, Radbrook Common
above the Pond on grassy path
111 m elevation; 51.7645° N; -01.3408° W
Coll. Maarten J. M. Christenhusz no.: 9573
with Keith Kirby, Ilia Leitch, Paul Kersey and José
Ignacio Márquez-Corro
21 June 2023
duplicates: E, K.
Darwin Tree of Life project KDTOL10866
https://www.darwintreeoflife.org
inches
centimeters
M000648
Kew
copyright reserved

Figure 1. ***Carex spicata*** **(lpCarSpic1) from which samples were taken for genome sequencing.**

Methods

Sample acquisition, flow cytometry and DNA barcoding

A specimen of *C. spicata* (specimen ID KDTOL10866, ToLID lpCarSpic1; Figure 1) was used for genome sequencing. It was collected on a grassy path above the pond at Radbrook Common, Wytham Woods, Oxfordshire, UK (latitude 51.7645, longitude −1.3408) on 21 June 2023 by Maarten Christenhusz, Keith Kirby, Paul Kersey, José Ignacio Márquez-Corro and Ilia Leitch and identified by José Ignacio Márquez-Corro.

A second specimen from the same locality (specimen ID KDTOL10867, ToLID lpCarSpic2) was used for Hi-C and RNA sequencing. It was collected by Maarten Christenhusz, Keith Kirby, Paul Kersey, José Ignacio Márquez-Corro and Ilia Leitch and identified by José Ignacio Márquez-Corro.

We recorded specimen and sample metadata using the Darwin Tree of Life Sample Manifest and the guidance developed by the project's Samples Working Group (Lawniczak *et al.*, 2022). The herbarium voucher for the sequenced plant (collector number 9573) is deposited at Kew (K001527204).

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). 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 sequencing, 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.

Sample preparation

We followed the Tree of Life triage protocol described by Howard *et al.* (2025). During triage, we weighed the material supplied and, where necessary, divided it into portions according to the tissue type, the amount available and the intended use. Leaf tissue from specimen lpCarSpic1 was used for genomic DNA extraction. Leaf tissue from specimen lpCarSpic2 was used for Hi-C. A separate leaf preparation supplied material for RNA sequencing.

DNA extraction

Tissue (106 mg) was prepared for DNA extraction by cryogenic bead beating. We washed the homogenate by sorbitol wash before extraction. We extracted high molecular weight DNA from leaf tissue using the automated plant organic extraction protocol. The DNA was sheared on a Megaruptor 3 and then purified using automated SPRI with AMPure PB beads on a KingFisher Apex. We measured the mean fragment length on an Agilent Femto Pulse system and the DNA concentration with a Qubit fluorometer and the dsDNA High Sensitivity Assay Kit. The mean fragment length was 16.4 kb, and the DNA concentration after SPRI clean-up was 36.4 ng/μL.

PacBio HiFi library preparation and sequencing

We prepared a low-input library with the SMRTbell Prep Kit 3.0. The library was sequenced in a four-plex pool on a PacBio Revio using the Revio sequencing plate at a loading concentration of 295 pM with 24 h of data collection per SMRT Cell.

Hi-C tissue preparation, library preparation and sequencing

We isolated nuclei from lpCarSpic2 leaf tissue after homogenisation using cryogenic bead beating and fixed them for 10 min in phosphate-buffered saline containing 2% formaldehyde, following the Tree of Life plant protocol. After proximity ligation with the Arima-HiC v2 kit, we enriched fragments containing biotin-labelled ligation junctions and prepared a library with the NEBNext Ultra II DNA Library Prep Kit. The library was sequenced in paired-end mode on an Illumina NovaSeq X.

RNA tissue preparation, library preparation and sequencing

We ground leaf tissue from specimen lpCarSpic2 to a powder by cryogenic bead beating. We extracted 2,673 ng of total RNA using the automated MagMAX mirVana protocol. Poly(A) RNA was selected with oligo(dT) beads and a stranded cDNA library was prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. The library was sequenced in paired-end mode on an Illumina NovaSeq X.

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. The *k*-mer frequency distribution was analysed with GenomeScope 2.0 (Ranallo-Benavidez *et al.*, 2020), providing fitted point estimates of haploid genome size, heterozygosity and repeat content.

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option.

The Hi-C reads (Rao *et al.*, 2014) were mapped to 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 (Tegenfeldt *et al.*, 2025) 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 *Carex pseudochinensis* (NC_087862.1). The comparator for the plastid assembly was the chloroplast genome of *Carex siderosticta* (PP790566.1).

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. The flat files and maps for curation were generated with TreeVal. Manual curation was conducted primarily in PretextView. Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). The primary and alternate assemblies were combined for curation. Manual corrections included one join. This increased the scaffold count by 2.6%. The curation process is described at sanger-tol/curation-resources. A Hi-C contact map of the primary assembly was generated with PretextSnapshot.

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.

Gene-set completeness was assessed with BUSCO (Tegenfeldt *et al.*, 2025), using the poales_odb12 lineage dataset ($n = 6{,}282$) in eukaryotic genome mode, with miniprot (Li, 2023). This assessment was produced by sanger-tol/busco.

The primary assembly was assessed using BlobToolKit (Challis *et al.*, 2020). PacBio reads were mapped to the assembly to derive sequence coverage. Within this workflow, BUSCO (Tegenfeldt *et al.*, 2025) was used to assess gene-set completeness. DIAMOND BLASTp searches of extracted BUSCO proteins and DIAMOND BLASTx searches of assembly sequences against UniProt Reference Proteomes (Buchfink *et al.*, 2021), followed by BLASTn searches of sequences without DIAMOND BLASTx hits against the NCBI nt database (Altschul *et al.*, 1990), supported taxonomic assignment. Coverage, sequence composition, taxonomic assignment and the workflow's BUSCO results were integrated into a BlobDir for interactive inspection and visualisation. The snail and blob plots were rendered locally from this hosted BlobDir using BlobTk. The scaffold-length radial scale function for the snail plot was set to linear (Challis & Blaxter, 2026). The blob plot used circular markers with area proportional to scaffold length (Challis *et al.*, 2020).

Genome sequence report

Sequence data

The genome of a specimen of *C. spicata* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 11.89 Gb (gigabases) from 0.82 million reads. These reads were used to assemble the genome. GenomeScope 2.0 analysis using a $p = 2$ model gave a *k*-mer-based haploid genome size estimate of 421.19 Mb, with heterozygosity of 0.13% and repeat content of 33.99%; the full-model fit was 98.41% (Figure 2). Using flow cytometry, the genome size (1C-value) of the sample was estimated to be 0.46 pg, equivalent to 450.00 Mb. These estimates guided expectations for the assembly. The NCBI assembly record reports 27× genome coverage. Hi-C sequencing produced 133.04 Gb from 440.53 million read pairs, which were used to scaffold the primary

assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

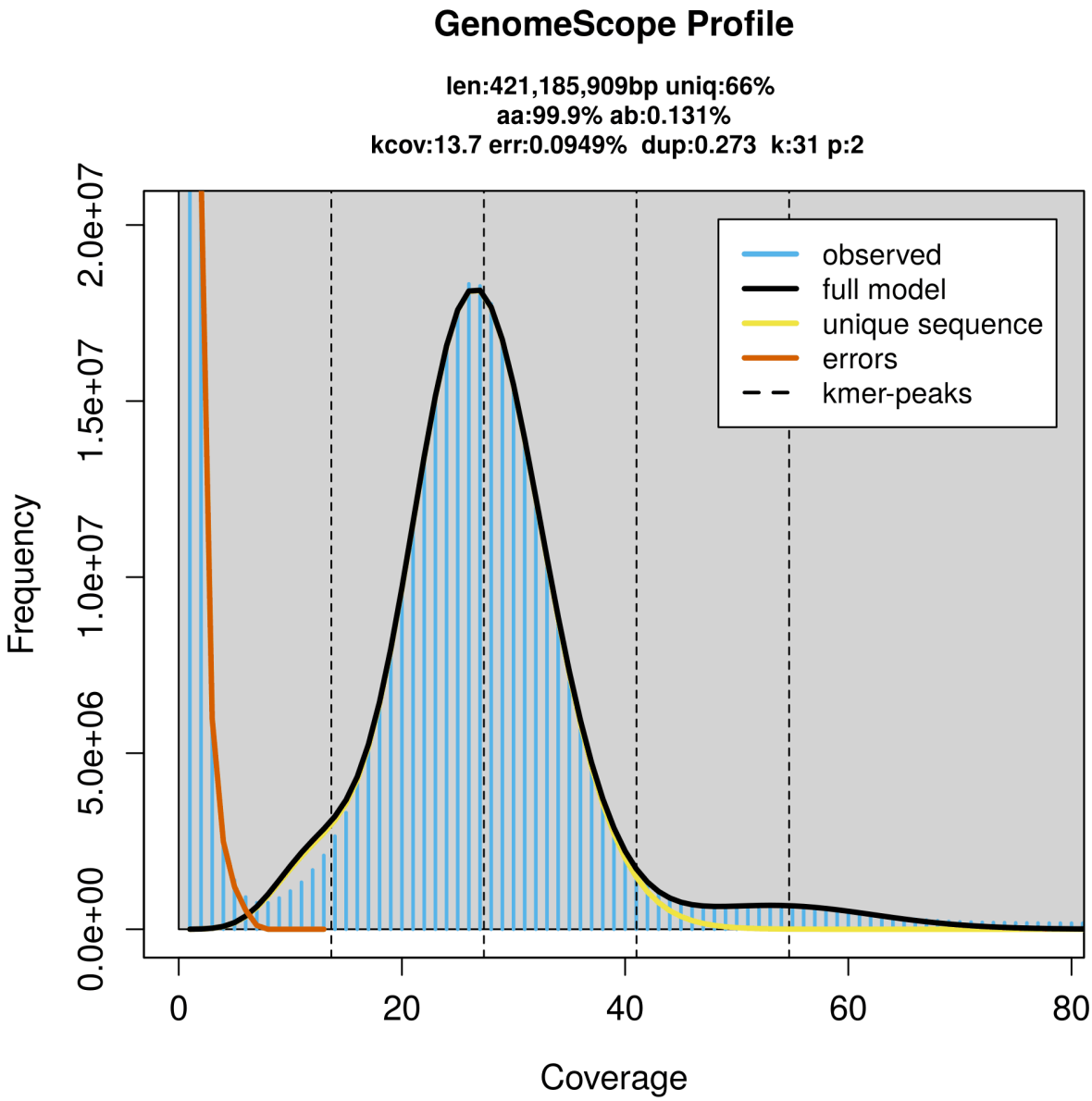

Figure 2. GenomeScope 2.0 model fitted to the *k*-mer frequency distribution. The plot shows the observed and modelled *k*-mer spectra used to estimate haploid genome size, heterozygosity and repeat content from unassembled sequencing reads.

Table 1. Specimen and sequencing data for *C. spicata* (BioProject PRJEB97530)

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	lpCarSpic1	lpCarSpic2	lpCarSpic2
Specimen ID	KDTOL10866	KDTOL10867	KDTOL10867

Platform	PacBio HiFi	Hi-C	RNA-seq
BioSample (source individual)	SAMEA114563450	SAMEA114563451	SAMEA114563451
BioSample (tissue)	SAMEA114563476	SAMEA114563477	SAMEA114563477
Tissue	leaf	leaf	leaf
Instrument	Revio	Illumina NovaSeq X	Illumina NovaSeq X
Library ID & run accession(s)	DTOL15020190: ERR15606357	SQPP-59463-T:E6: ERR15610345	ERR16910282
Read count total	0.82 million reads	440.53 million read pairs	44.31 million read pairs
Base count total	11.89 Gb	133.04 Gb	13.38 Gb

Assembly statistics

The primary assembly was generated and an alternate haplotype assembly was also deposited at scaffold level in INSDC databases. The primary assembly is proposed as the reference assembly for this species. It has a total length of 385.54 Mb in 38 scaffolds (excluding organelle sequences), with 10 gaps, and its scaffold N50 is 13.14 Mb (Table 2). The scaffold auN for the primary assembly is 14.3 Mb.

Table 2. Genome assembly statistics for *C. spicata*

Genome assembly	Primary assembly	Alternate haplotype assembly
Assembly name	lpCarSpic1.1	lpCarSpic1.1 alternate haplotype
Assembly accession	GCA_977035945.1	GCA_977035845.1
Assembly level	chromosome	scaffold
Span (Mb)	385.54	32.12
Number of chromosomes	29	-
Number of contigs	48	378
Contig N50 (Mb)	13.13	0.31
Number of scaffolds (excluding organelle sequences)	38	291
Scaffold N50 (Mb)	13.14	0.67
Scaffold auN (Mb)	14.3	-
Longest scaffold length (Mb)	24.66	2.28
Organelles	Mitochondrial genome: 1,362.88 kb; Plastid genome: 183.59 kb	-

Most of the primary assembly sequence (99.84%) was assigned to 29 chromosome-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

The mitochondrial genome (length 1362.88 kb, OZ365402.1) and plastid genome (length 183.59 kb, OZ365403.1) were also assembled. These sequences are included as contigs in the multifasta file of the primary genome submission and as standalone records.

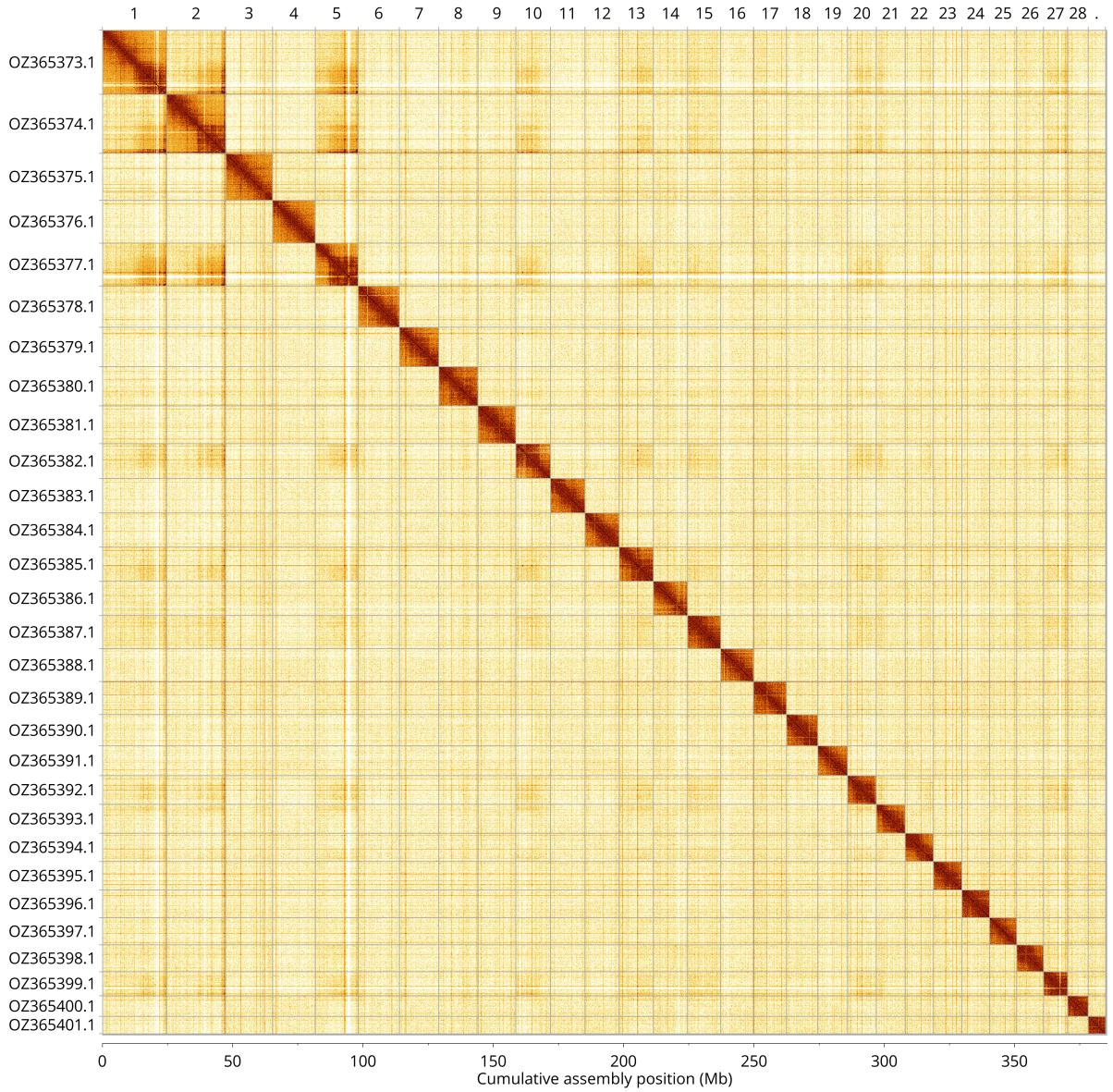

Figure 3. Hi-C contact map of the *C. spicata* primary assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *C. spicata* lpCarSpic1

INSDC accession	Molecule	Length (Mb)	GC%
OZ365373.1	1	24.66	34.50
OZ365374.1	2	22.70	34
OZ365375.1	3	17.93	34.50
OZ365376.1	4	16.43	34
OZ365377.1	5	16.41	34.50
OZ365378.1	6	15.87	34
OZ365379.1	7	15.11	34

INSDC accession	Molecule	Length (Mb)	GC%
OZ365380.1	8	15.04	34.50
OZ365381.1	9	14.51	34.50
OZ365382.1	10	13.29	34.50
OZ365383.1	11	13.24	33.50
OZ365384.1	12	13.14	34.50
OZ365385.1	13	13.13	34.50
OZ365386.1	14	13.05	34.50
OZ365387.1	15	12.75	34
OZ365388.1	16	12.66	34
OZ365389.1	17	12.60	34.50
OZ365390.1	18	11.95	34
OZ365391.1	19	11.41	34
OZ365392.1	20	11.09	34
OZ365393.1	21	11.01	34
OZ365394.1	22	10.87	34
OZ365395.1	23	10.87	34
OZ365396.1	24	10.62	34
OZ365397.1	25	10.42	34
OZ365398.1	26	10.31	34
OZ365399.1	27	9.28	34
OZ365400.1	28	7.87	34.50
OZ365401.1	29	6.70	34

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 70.8. The *k*-mer completeness is 98.80% for the primary assembly, 7.56% for the alternate haplotype, and 99.37% for the combined assemblies (Figure 4).

BUSCO analysis of the primary assembly using the poales_odb12 reference set (n = 6,282) identified 97.9% of the expected gene set (single = 94.4%, duplicated = 3.5%).

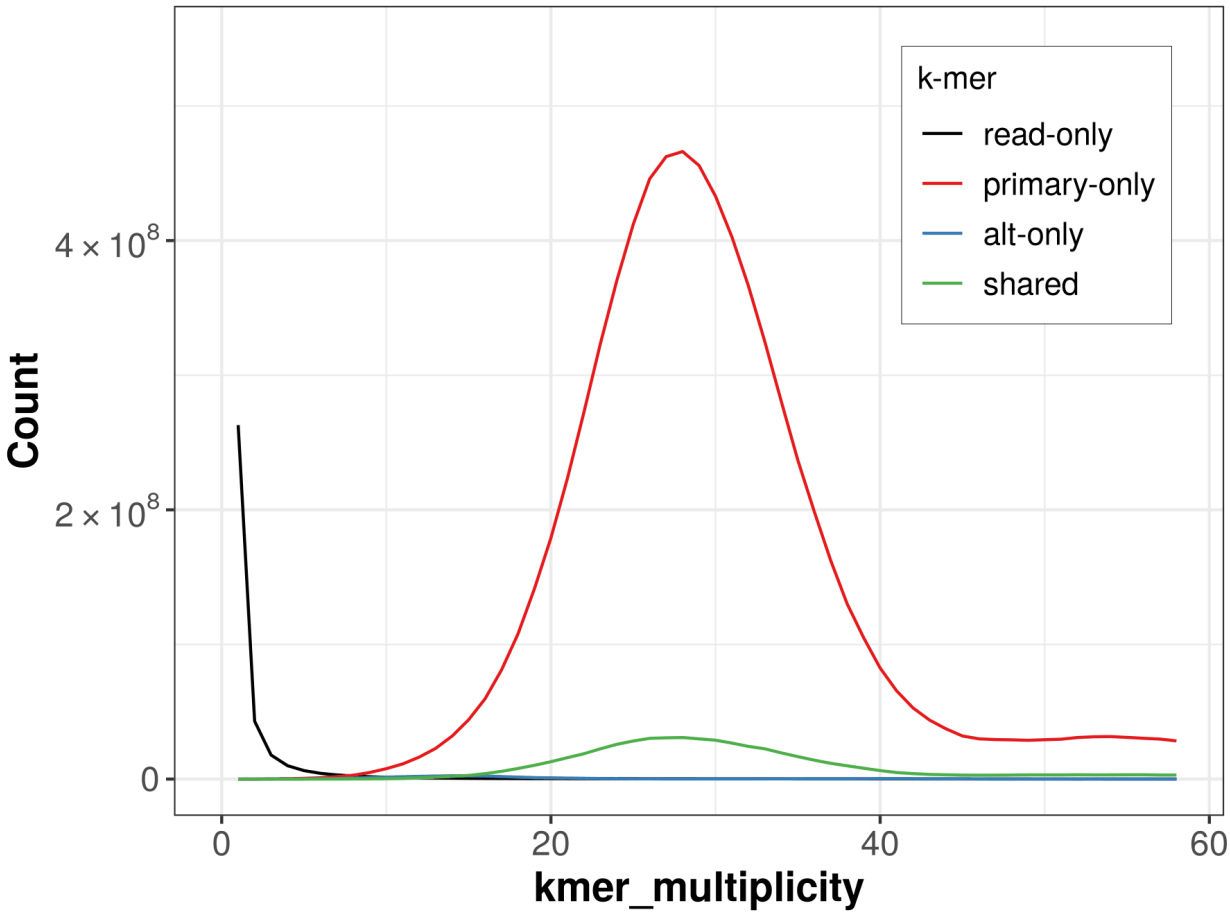

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the primary assembly, the alternate haplotype assembly and the combined assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows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

The snail plot in Figure 5 provides a summary of assembly contiguity, cumulative scaffold count and sequence composition for the primary assembly (Challis & Blaxter, 2026). The relative sizes of the longest scaffold, scaffold N50 and N90, together with the dark-grey scaffold-length distribution, show how much of the assembly is contained in long scaffolds. The outer composition track shows variation in GC, AT and N content, while the central track shows the cumulative scaffold count.

The blob plot in Figure 6 displays scaffolds from the primary assembly by GC proportion and base coverage. Marker area is proportional to scaffold length, and colour indicates taxonomic assignment (Challis *et al.*, 2020).

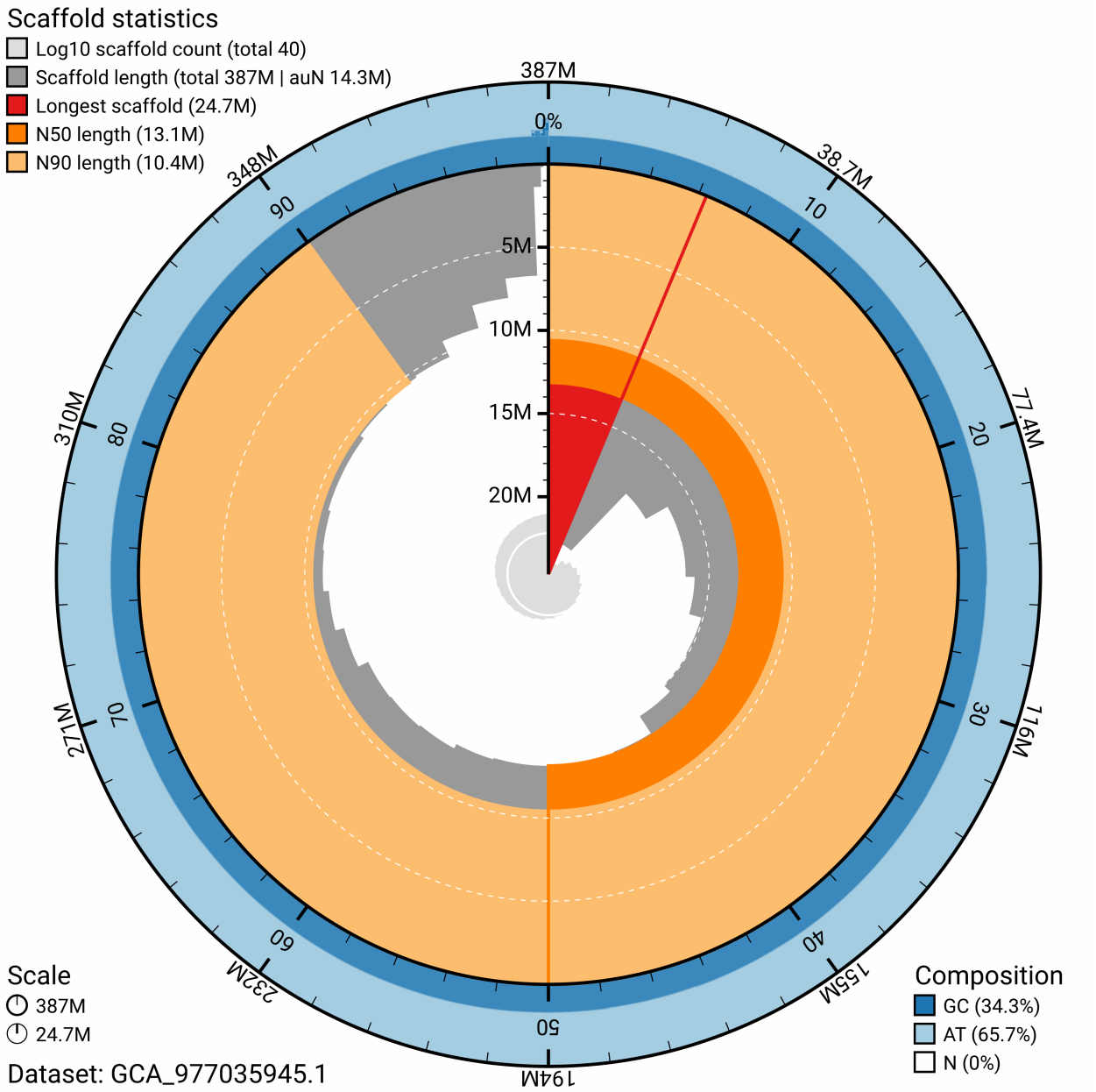

Figure 5. Assembly metrics for lpCarSpic1.1, the primary assembly. The BlobToolKit snail plot provides an overview of assembly contiguity and sequence composition. The circumference represents the length of the primary assembly, and the main plot is divided into 1,000 bins around the circumference. The 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. Scaffold length is shown on a linear radial scale, scaled to the longest scaffold. Red shading indicates the longest scaffold, and the deeper orange and pale orange shading represent the N50 and N90 lengths. The summary statistics include scaffold auN, the area under the scaffold Nx curve. The light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. The assembly can be explored interactively in the BlobToolKit viewer.

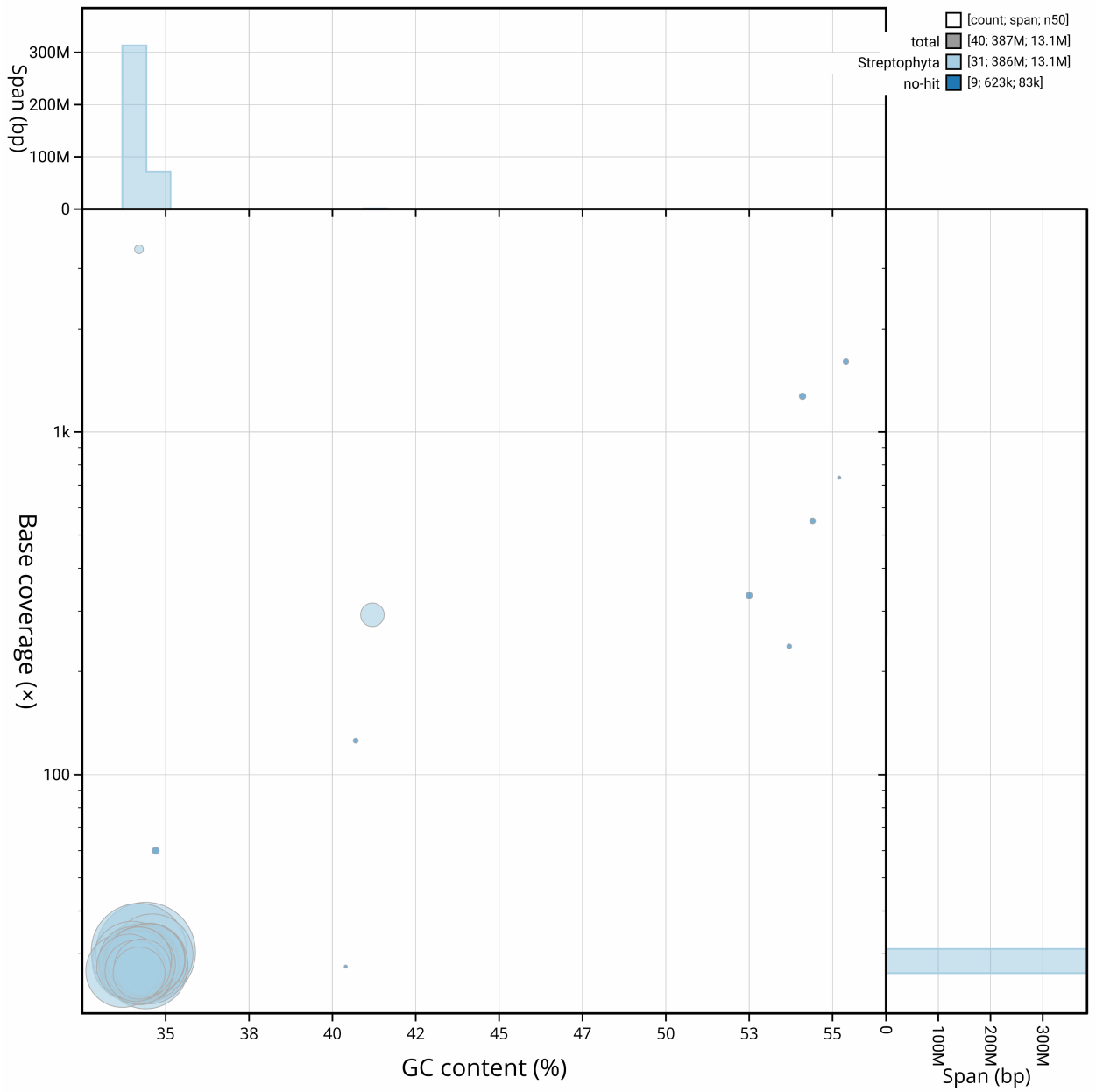

Figure 6. BlobToolKit blob plot for lpCarSpic1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). Circles represent scaffolds; marker area is proportional to scaffold length, and colour indicates taxonomic assignment. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. The assembly can be explored interactively in the BlobToolKit viewer.

Table 4 presents assembly quality metrics alongside the Earth BioGenome Project (EBP) benchmarks described by Howe *et al.* (2026). The three-part EBP metric calculated for the primary assembly is **7.C.Q77**.

Table 4. Earth BioGenome Project summary metrics for the *C. spicata* assembly

Measure	Value	*Benchmark*
EBP summary (primary)	7.C.Q77	*6.C.Q40*

Measure	Value	*Benchmark*
Contig N50 length	13.13 Mb	*> 1 Mb*
Scaffold N50 length	13.14 Mb	*= chromosome N50*
Consensus quality (QV)	Primary: 77.1; alternate: 63.3; combined: 70.8	*# 40*
k-mer completeness	Primary: 98.80%; alternate: 7.56%; combined: 99.37%	*> 90% (combined assembly)*
BUSCO	C:97.9% [S:94.4%, D:3.5%], F:0.8%, M:1.3%, n:6,282	*C > 90%*
Percentage of assembly assigned to chromosomes	99.84%	*# 90%*

Notes: The EBP summary reports log10(contig N50), rounded down; C for chromosome-scale scaffolding or log10(scaffold N50), rounded down; and Q for consensus quality (Merqury QV). BUSCO v6.1.0, poales_odb12: C=complete; S=single-copy; D=duplicated; F=fragmented; M=missing; n=orthologues. BUSCO short summary: primary assembly. For plants, total BUSCO completeness (C) replaces the EBP single-copy benchmark to accommodate biological duplication.

Data availability

European Nucleotide Archive: Carex spicata. Accession number PRJEB97530; https://identifiers.org/ena.embl/PRJEB97530. The genome sequence is released openly for reuse. The *C. spicata* 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 primary and alternate assemblies 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 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

Software	Version	Source
bedtools	2.31.0; 2.31.1	https://github.com/arq5x/bedtools2
BLAST	2.16.0+	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobTk (hosted BlobDir)	0.8.1	https://github.com/genomehubs/blobtk
BlobTk (snail and blob plot renderer)	0.8.3	https://github.com/genomehubs/blobtk
BlobToolKit	4.4.6	https://github.com/genomehubs/blobtoolkit
BUSCO	6.0.0; 6.1.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.16	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.2	https://github.com/thegenemyers/FASTK
GenomeScope 2.0	2.1.0	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.25.0-r726	https://github.com/chhylp123/hifiasm

Software	Version	Source
MerquryFK	1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209; 2.29-r1283	https://github.com/lh3/minimap2
miniprot	0.18-r281	https://github.com/lh3/miniprot
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4; 25.04.6; 25.10.0; 25.10.4	https://github.com/nextflow-io/nextflow
Oatk	0.9	https://github.com/c-zhou/oatk
PretextSnapshot	0.0.9	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
samtools	1.17; 1.2; 1.21; 1.23.1	https://github.com/samtools/samtools
sanger-tol/ascc	v0.5.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.11.1	https://github.com/sanger-tol/blobtoolkit
sanger-tol/busco	v0.3.1; 0.3.1	https://github.com/sanger-tol/busco
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	v1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2.2	https://github.com/c-zhou/yahs

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.

References

Altschul S. F., Gish W., Miller W.: **Basic Local Alignment Search Tool**. *Journal of Molecular Biology*; 1990; **215**(3): 403–410. PubMed Abstract | 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 |

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

Challis Richard, Blaxter Mark: **Snail plots are badges of genome assembly quality**. *G3 Genes|Genomes|Genetics*; 2026; **16**(6): jkag074. PubMed Abstract | Publisher Full Text |

Chase M. W., Hills H. H.: **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 G. T., Feng X.: **Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm**. *Nature Methods*; 2021; **18**(2): 170–175. PubMed Abstract | Publisher 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 |

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

Howard Caroline, Denton Amy, Jackson Benjamin W.: **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 |

Howe K., Agius D. R., Alioto T.: **Generating eukaryotic reference genome assemblies: Earth BioGenome Project quality standards and recommendations**. *arXiv*; 2026 2610.03075. Reference Source Publisher Full Text |

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

Jermy A. C., Simpson D. A., Foley M. J. Y.: *Sedges of the British Isles* Oswald P. H.: London: Botanical Society of the British Isles; 2007. 3 ed. 9780901158666

Kurtzer G. M., Sochat V., Bauer M. W.: **Singularity: Scientific containers for mobility of compute**. *PLoS One*; 2017; **12**(5): e0177459. PubMed Abstract | Publisher Full Text |

Lawniczak M. K. N., Davey R. P., Rajan J.: **Specimen and sample metadata standards for biodiversity genomics: A proposal from the Darwin Tree of Life project**. *Wellcome Open Research*; 2022; **7**: 187. Publisher Full Text |

Li H.: **Protein-to-genome alignment with miniprot**. *Bioinformatics*; 2023; **39**(1): btad014. PubMed Abstract | Publisher Full Text |

Loureiro J., Rodriguez E., Doležel J.: **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 |

Pellicer J., Powell R. F., Leitch I. J.: **The application of flow cytometry for estimating genome size, ploidy level endopolyploidy, and reproductive modes in plants**. *Molecular plant taxonomy: Methods and protocols* Besse P.: New York, NY: Humana; 2021. **2222**: 325–361. PubMed Abstract | Publisher Full Text |

Porter M. S., Foley M. J. Y.: ***Carex spicata* Huds.**. *BSBI Online Plant Atlas 2020* Stroh P. A., Humphrey T. A., Burkmar R. J., Pescott O. L., Roy D. B., Walker K. J.: Botanical Society of Britain and Ireland; 2023. Accessed: 2026-11-24. Reference Source

: **Plants of the World Online: *Carex spicata* Huds.**. Royal Botanic Gardens, Kew. 2026. Accessed: 2026-11-24. Reference Source

Ranallo-Benavidez T. R., Jaron K. S., Schatz M. C.: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes**. *Nature Communications*; 2020; **11**(1): 1432. PubMed Abstract | Publisher Full Text |

Rao S. S. P., Huntley M. H., Durand N. C.: **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 |

Ratnasingham S., Hebert P. D. N.: **BOLD: The Barcode of Life Data System (http://www.barcodinglife.org)**. *Molecular Ecology Notes*; 2007; **7**(3): 355–364. PubMed Abstract | Publisher Full Text |

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

Tegenfeldt Fredrik, Kuznetsov Dmitry, Manni Mosè: **OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes**. *Nucleic Acids Research*; 2025; **53**(D1): D516–D522. PubMed Abstract | Publisher Full Text |

Twyford A. D., Beasley J., Barnes I.: **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 |

Vasimuddin Md., Misra S., Li H.: **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 Chenxi, Brown Max, Blaxter Mark: **Oatk: A de novo assembly tool for complex plant organelle genomes**. *Genome Biology*; 2025; **26**(1): 235. Publisher Full Text |

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

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 27 November 2026

https://doi.org/

© **2026 Benchmarking A.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Automated Benchmarking**

Wellcome Sanger Institute, Hinxton, England, UK

As this article is a Genome Note, Automated Benchmarking has been used to determine whether the reported assembly passes four defined checks established by the Earth BioGenome Project: the raw and assembled data are publicly available, the Earth BioGenome Project three-part metric is met, the k-mer completeness exceeds 90%, and the BUSCO score passes the stated threshold.

For Carex spicata (lpCarSpic1), all four checks passed.

Data available: pass. ENA study PRJEB97530, raw-read accessions ERR15606357, ERR15610345, and ERR16910282, and assembly accessions GCA_977035945.1 and GCA_977035845.1 are public.

Earth BioGenome Project three-part metric: pass. The primary assembly is reported as 7.C.Q77, meeting the 6.C.Q40 reference standard.

k-mer completeness: pass. The combined haplotype assemblies have 99.37% k-mer completeness, exceeding the Earth BioGenome Project target of 90%.

BUSCO: pass. The primary assembly, which is proposed as the reference assembly for this species, has 97.9% complete BUSCOs (94.4% single-copy and 3.5% duplicated). BUSCO software version: 6.1.0; lineage dataset: poales_odb12. It meets the plant-specific target of greater than 90% complete BUSCOs.

This decision uses values reported in the Genome Note and public accession records.

Competing Interests: This Automated Benchmarking peer-review report was generated automatically and relies on analyses performed by the Tree of Life programme, which also produced the Genome Note under review.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.