Check for updates

GENOME NOTE

The genome sequence of tufted sedge, *Carex elata* All. (Poales: Cyperaceae)

[version 1; peer review: 1 approved]

Maarten J. M. Christenhusz[1,2], José Ignacio Márquez-Corro[1], Isabel Larridon[3], 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 Gardens, Kew, Richmond, Surrey, England, UK

v1 **First published:** 27 Nov 2026, **11**:1124
https://doi.org/10.12688/wellcomeopenres.28541.1
Latest published: 27 Nov 2026, **11**:1124
https://doi.org/10.12688/wellcomeopenres.28541.1

Abstract

We present a genome assembly of *Carex elata* (tufted sedge; Streptophyta; Magnoliopsida; Poales; Cyperaceae). The assembly consists of two haplotypes with total lengths of 414.86 megabases and 412.35 megabases. Most of the haplotype 1 assembly (98.33%) was scaffolded into 40 chromosomal pseudomolecules. Haplotype 2 was assembled to scaffold level. Multiple mitochondrial sequences were also assembled. The plastid genome assembly has a length of 213.59 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 elata, tufted 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; **Márquez-Corro JI**: Investigation, Resources; **Larridon I**: Investigation, Resources;

Competing interests: No competing interests were disclosed.

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

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, Márquez-Corro JI, Larridon I *et al.* **The genome sequence of tufted sedge, *Carex elata* All. (Poales: Cyperaceae) [version 1; peer review: 1 approved]** Wellcome Open Research 2026, **11**:1124 https://doi.org/10.12688/wellcomeopenres.28541.1

First published: 27 Nov 2026, **11**:1124 https://doi.org/10.12688/wellcomeopenres.28541.1

Species taxonomy

Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; Liliopsida; Petrosaviidae; commelinids; Poales; Cyperaceae; Cyperoideae; Cariceae; *Carex*; *Carex subgen. Carex*; *Carex elata* All. (NCBI:txid241206)

Background

Carex elata All. (tufted sedge) forms dense tussocks up to 40 cm high and has long, blue-green leaves. Its flowering stems are sharply triangular in cross-section and reach 1 m (Jermy *et al.*, 2007). In Britain and Ireland it grows in lowland wetlands, often in calcareous conditions. It occurs in fens, along lake and river margins, and in seasonally flooded ditches and wet alder or willow woodland (Porter & Foley, 2023). Its native range extends from Europe to Iran and includes Algeria and Morocco (POWO, 2026).

We sequenced the genome of *Carex elata* as part of the Darwin Tree of Life Project. The sequenced specimen was collected from Wicken Fen, Ely, Cambridgeshire, UK (Figure 1).

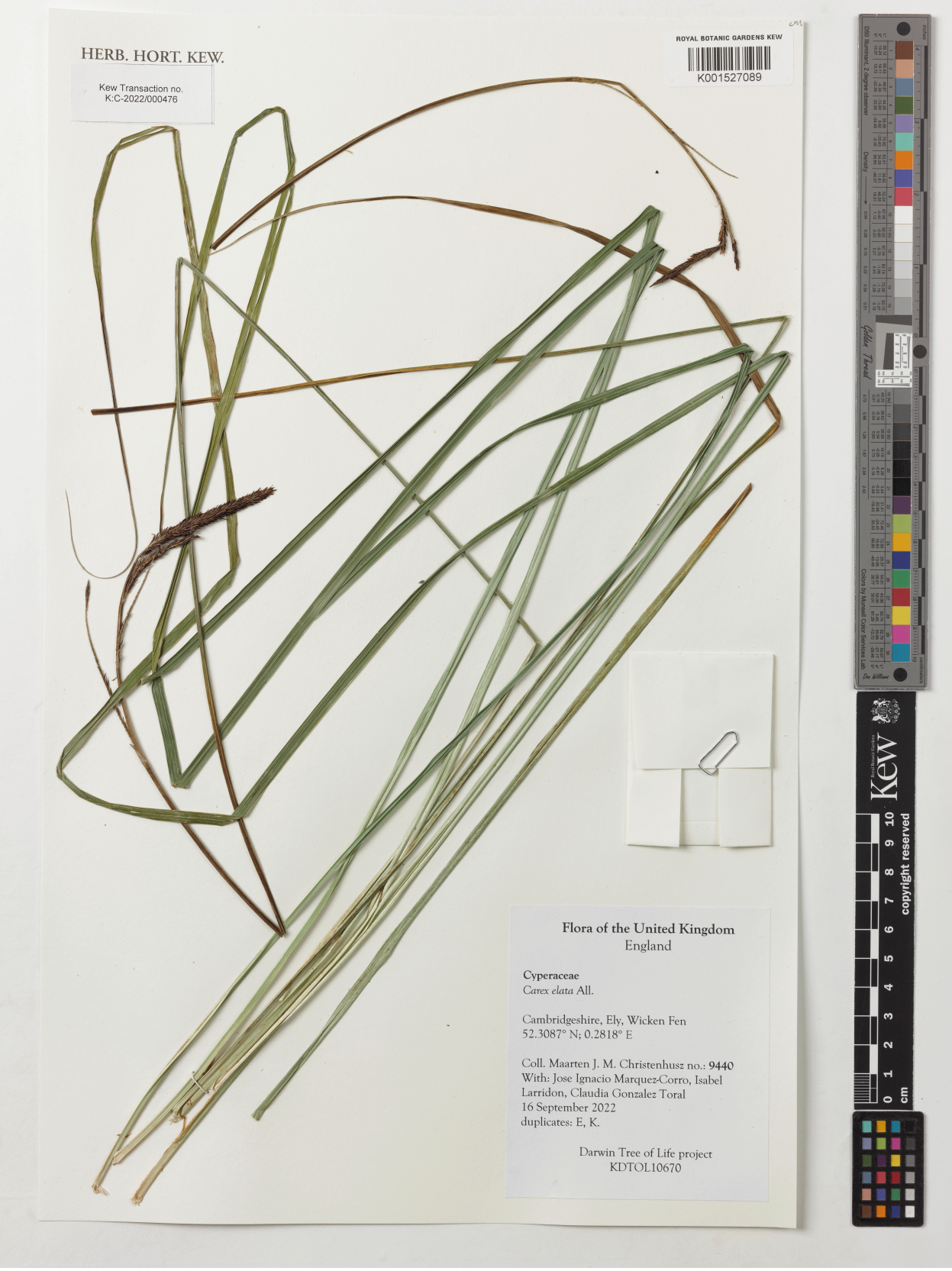
HERB. HORT. KEW.
Kew Transaction no.
K:C-2022/000476
ROYAL BOTANIC GARDENS KEW
K001527089
Flora of the United Kingdom
England
Cyperaceae
Carex elata All.
Cambridgeshire, Ely, Wicken Fen
52.3087° N; 0.2818° E
Coll. Maarten J. M. Christenhusz no.: 9440
With: Jose Ignacio Marquez-Corro, Isabel Larridon, Claudia Gonzalez Toral
16 September 2022
duplicates: E, K.
Darwin Tree of Life project
KDTOL10670
Kew
copyright reserved
cm

Figure 1. ***Carex elata*** **(lpCarElat1) from which samples were taken for genome sequencing.**

Methods

Sample acquisition, flow cytometry and DNA barcoding

A specimen of *C. elata* (specimen ID KDTOL10670, ToLID lpCarElat1; Figure 1) was used for genome sequencing. It was collected from Wicken Fen, Ely, Cambridgeshire, UK (latitude 52.3087, longitude 0.2818) on 2022-09-16. The specimen was collected by Maarten Christenhusz, Jose Ignacio Marquez-Corro and Isabel Larridon, and identified by Maarten Christenhusz. The lpCarElat1 specimen was also used for Hi-C and RNA sequencing. 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 MC9440) is deposited at Kew (K001527089).

The genome size was estimated by flow cytometry following the 'one-step' method outlined in Pellicer *et al.* (2021) and using propidium iodide as the fluorochrome. LB01 was used for isolation of nuclei (Loureiro *et al.*, 2007). 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. The February 2023 tissue preparation supplied material for genomic DNA extraction (library DTOL13456439) and Hi-C. The July 2023 tissue preparation from specimen lpCarElat1 supplied material for genomic DNA extraction (library DTOL14180134). A separate tissue preparation supplied material for RNA sequencing.

DNA extraction

We ground 116 mg of stem from specimen lpCarElat1 to a powder by cryogenic bead beating. We extracted high molecular weight DNA using the automated plant organic extraction protocol. DNA for library DTOL14180134 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 sheared fragments averaged 16.4 kb. The eluate contained 1,153.16 ng of DNA at a post-SPRI DNA concentration of 25.4 ng/μL.

The February 2023 stem preparation used for library DTOL13456439 and Hi-C was pulverised using cryoPREP. High molecular weight DNA for this library was extracted from 58 mg of the prepared stem tissue using the automated plant MagAttract v2 protocol. DNA was sheared in a Covaris g-TUBE for ultra-low input (ULI) library preparation.

PacBio HiFi library preparation and sequencing

We prepared the low-input library DTOL14180134 with the SMRTbell Prep Kit 3.0. For ULI library DTOL13456439, we amplified the DNA by PCR with the SMRTbell gDNA Sample Amplification Kit and prepared the library with the SMRTbell Express Template Prep Kit 2.0. The libraries were sequenced on a PacBio Revio using the Revio sequencing plate with 24 h of data collection per SMRT Cell. Library DTOL14180134 (run ERR14835591) was sequenced in a four-plex pool at a loading concentration of 149 pM. Library DTOL13456439 (run ERR14835901) was sequenced in a four-plex pool at a loading concentration of 229 pM.

Hi-C tissue preparation, library preparation and sequencing

We isolated nuclei from the Hi-C stem aliquot 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 6000.

RNA tissue preparation, library preparation and sequencing

We extracted 2,457 ng of total RNA from stem tissue from specimen lpCarElat1 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 in Hi-C phasing mode (Cheng *et al.*, 2021, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021; 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* (PP465050.1). The comparator for the plastid assembly was the chloroplast genome of *Carex giraldiana* (NC_081502.1).

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 two haplotype assemblies were combined for curation. Manual corrections included 34 breaks and 109 joins. This reduced the scaffold count by 20.7% and increased the scaffold N50 by 0.9%. The curation process is described at sanger-tol/curation-resources. A Hi-C contact map of the haplotype 1 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 both 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 haplotype 1 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. elata* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 32.78 Gb (gigabases) from 2.87 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 378.13 Mb, with heterozygosity of 2.13% and repeat content of 39.88%; the full-model fit was 95.18% (Figure 2). Using flow cytometry, the genome size (1C-value) of the sample was estimated to be 0.5 pg, equivalent to 490.00 Mb. These estimates guided expectations for the assembly. The NCBI assembly record reports 81× genome coverage. Hi-C sequencing produced 94.28 Gb from 312.18 million read pairs, which were used to scaffold the haplotype

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

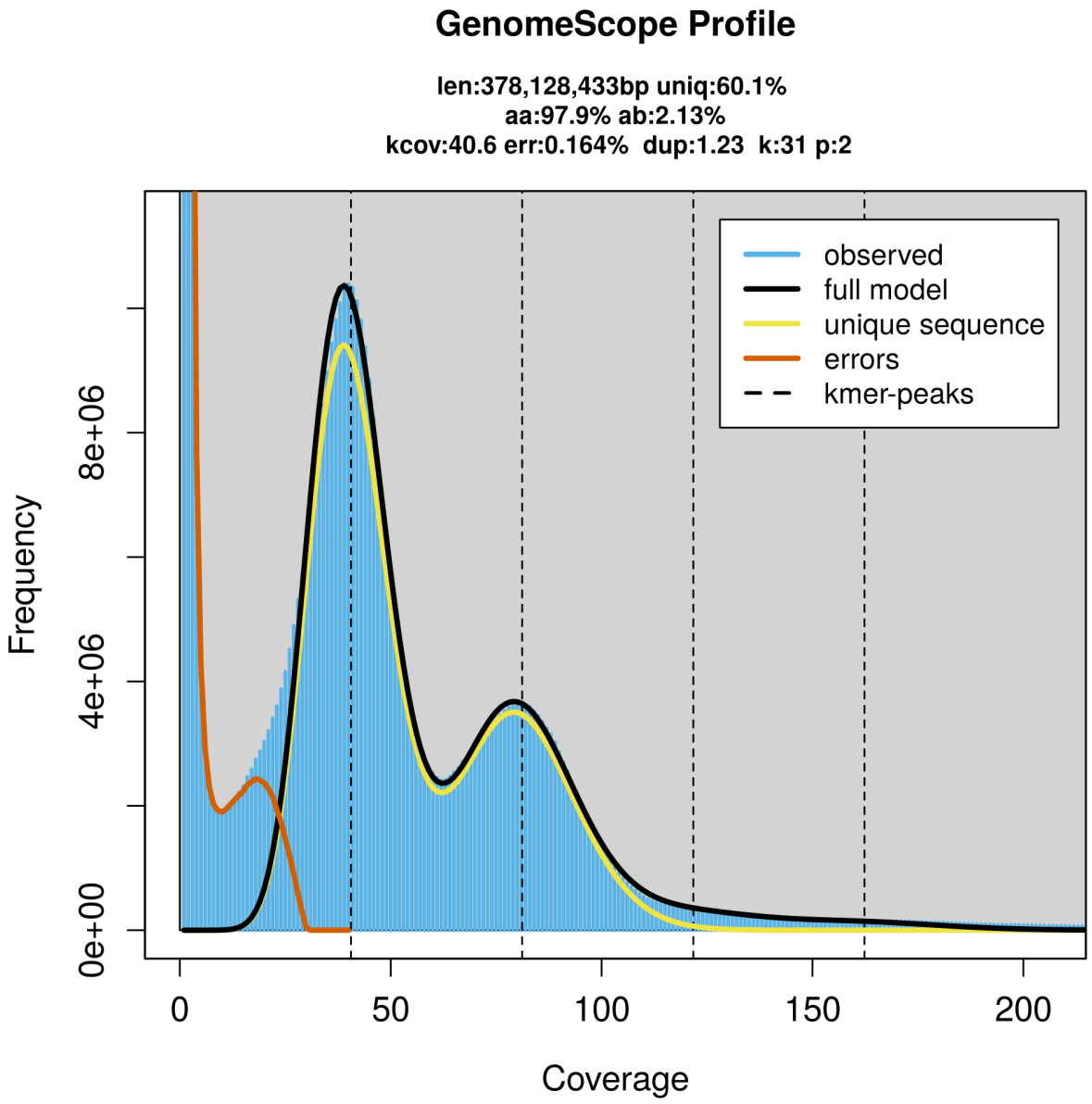

Figure 2. GenomeScope 2.0 model fitted to the *k*-mer frequency distribution. The plot shows the observed and modelled *k*-mer spectra used to estimate haploid genome size, heterozygosity and repeat content from unassembled sequencing reads.

Table 1. Specimen and sequencing data for *C. elata* (BioProject PRJEB87945)

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	lpCarElat1	lpCarElat1	lpCarElat1
Specimen ID	KDTOL10670	KDTOL10670	KDTOL10670

Platform	PacBio HiFi	Hi-C	RNA-seq
BioSample (source individual)	SAMEA112287516	SAMEA112287516	SAMEA112287516
BioSample (tissue)	SAMEA112287596	SAMEA112287596	SAMEA112287596
Tissue	stem	stem	stem
Instrument	Revio (2 runs)	Illumina NovaSeq 6000	Illumina NovaSeq X
Library ID & run accession(s)	DTOL14180134: ERR14835591; DTOL13456439: ERR14835901	SQPP-23224-N:E1: ERR14835509	ERR15551511
Read count total	2.87 million reads	312.18 million read pairs	61.57 million read pairs
Base count total	32.78 Gb	94.28 Gb	18.59 Gb

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. The haplotype 1 assembly was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The haplotype 1 assembly, which is proposed as the reference assembly for this species, has a total length of 414.86 Mb in 148 scaffolds (excluding organelle sequences), with 366 gaps, and its scaffold N50 is 10.68 Mb (Table 2). The scaffold auN for the haplotype 1 assembly is 11.4 Mb.

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

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	lpCarElat1.hap1.2	lpCarElat1.hap2.2
Assembly accession	GCA_965260875.2	GCA_965260805.2
Assembly level	chromosome	scaffold
Span (Mb)	414.86	412.35
Number of chromosomes	40	-
Number of contigs	514	493
Contig N50 (Mb)	1.77	1.69
Number of scaffolds (excluding organelle sequences)	148	126
Scaffold N50 (Mb)	10.68	10.64
Scaffold auN (Mb)	11.4	11.4
Longest scaffold length (Mb)	21.81	21.35
Organelles	Mitochondrial genomes: 1,078.57; 15.67; 100.56; 129.90 kb; Plastid genome: 213.59 kb	-

Most of the haplotype 1 assembly sequence (98.33%) was assigned to 40 chromosome-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Chromosome 39 and 40 seem to form a tetrad with their haplotype 2 pairs.

Multiple mitochondrial genomes (4) and the plastid genome (length 213.59 kb, OZ337051.1) were also assembled. These sequences are included as contigs in the multifasta file of the haplotype 1 genome submission and as standalone records.

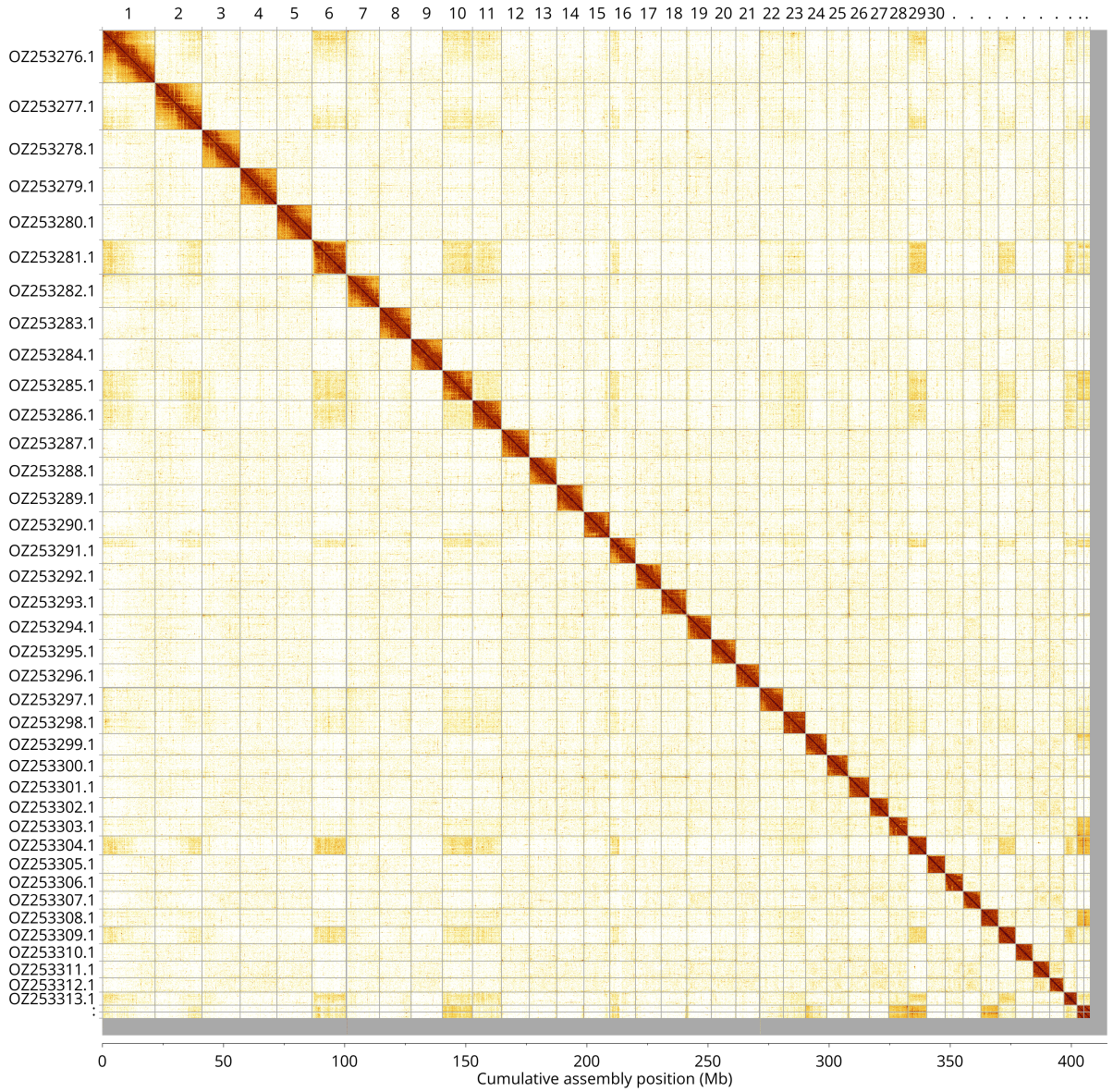

Figure 3. Hi-C contact map of the *C. elata* haplotype 1 assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *C. elata* lpCarElat1

INSDC accession	Molecule	Length (Mb)	GC%
OZ253276.1	1	21.81	35
OZ253277.1	2	19.43	35
OZ253278.1	3	15.77	34.50
OZ253279.1	4	15.12	35
OZ253280.1	5	14.46	34.50
OZ253281.1	6	14.32	35
OZ253282.1	7	13.63	35.50

INSDC accession	Molecule	Length (Mb)	GC%
OZ253283.1	8	13	34.50
OZ253284.1	9	12.97	34.50
OZ253285.1	10	12.36	35
OZ253286.1	11	11.97	35
OZ253287.1	12	11.48	35
OZ253288.1	13	11.37	35
OZ253289.1	14	11.21	35
OZ253290.1	15	10.68	35
OZ253291.1	16	10.68	35
OZ253292.1	17	10.55	35
OZ253293.1	18	10.49	34.50
OZ253294.1	19	10.17	35
OZ253295.1	20	10.12	34.50
OZ253296.1	21	9.87	34.50
OZ253297.1	22	9.73	35
OZ253298.1	23	9.17	34.50
OZ253299.1	24	8.86	34.50
OZ253300.1	25	8.80	35
OZ253301.1	26	8.69	35
OZ253302.1	27	8	34.50
OZ253303.1	28	7.90	35
OZ253304.1	29	7.77	34.50
OZ253305.1	30	7.58	34.50
OZ253306.1	31	7.47	34.50
OZ253307.1	32	7.33	34.50
OZ253308.1	33	7.19	34
OZ253309.1	34	7.16	34
OZ253310.1	35	7.05	35
OZ253311.1	36	6.94	34
OZ253312.1	37	5.89	34
OZ253313.1	38	5.43	34
OZ253314.1	39	2.76	35
OZ253315.1	40	2.76	34

Assembly quality metrics

For the haplotype 1 assembly, the estimated QV is 62.9, and for the haplotype 2 assembly, 62.8. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.8. The *k*-mer completeness is 68.86% for the haplotype 1 assembly, 68.97% for the haplotype 2 assembly, and 98.84% for the combined haplotypes (Figure 4).

BUSCO analysis using the poales_odb12 reference set (n = 6,282) identified 97.4% of the expected gene set (single = 93.1%, duplicated = 4.3%) in the haplotype 1 assembly. For the haplotype 2 assembly, BUSCO analysis identified 97.5% of the expected gene set (single = 93.1%, duplicated = 4.4%).

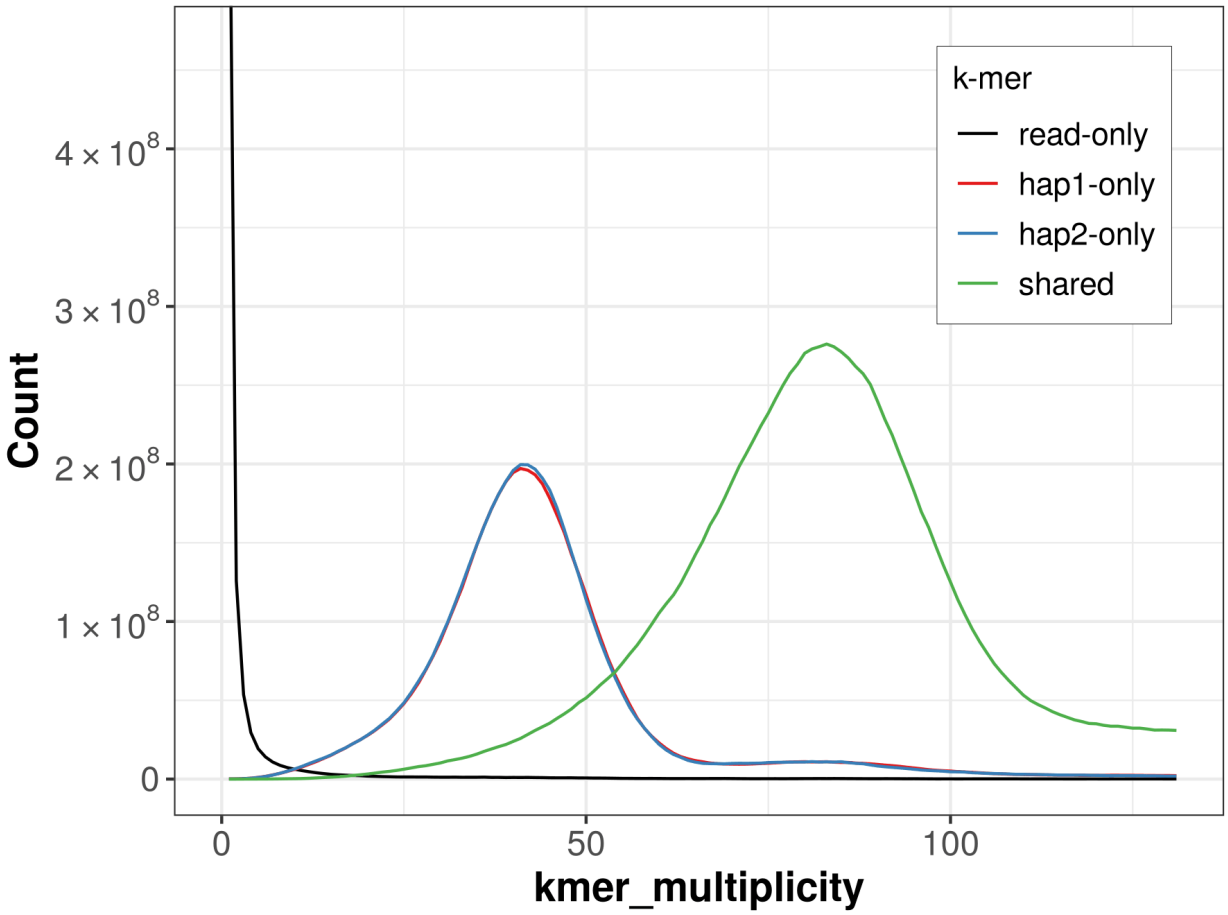

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the haplotype 1 assembly, the haplotype 2 assembly and the combined haplotypes. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows 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 haplotype 1 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 haplotype 1 assembly by GC proportion and base coverage. Marker area is proportional to scaffold length, and colour indicates taxonomic assignment (Challis *et al.*, 2020).

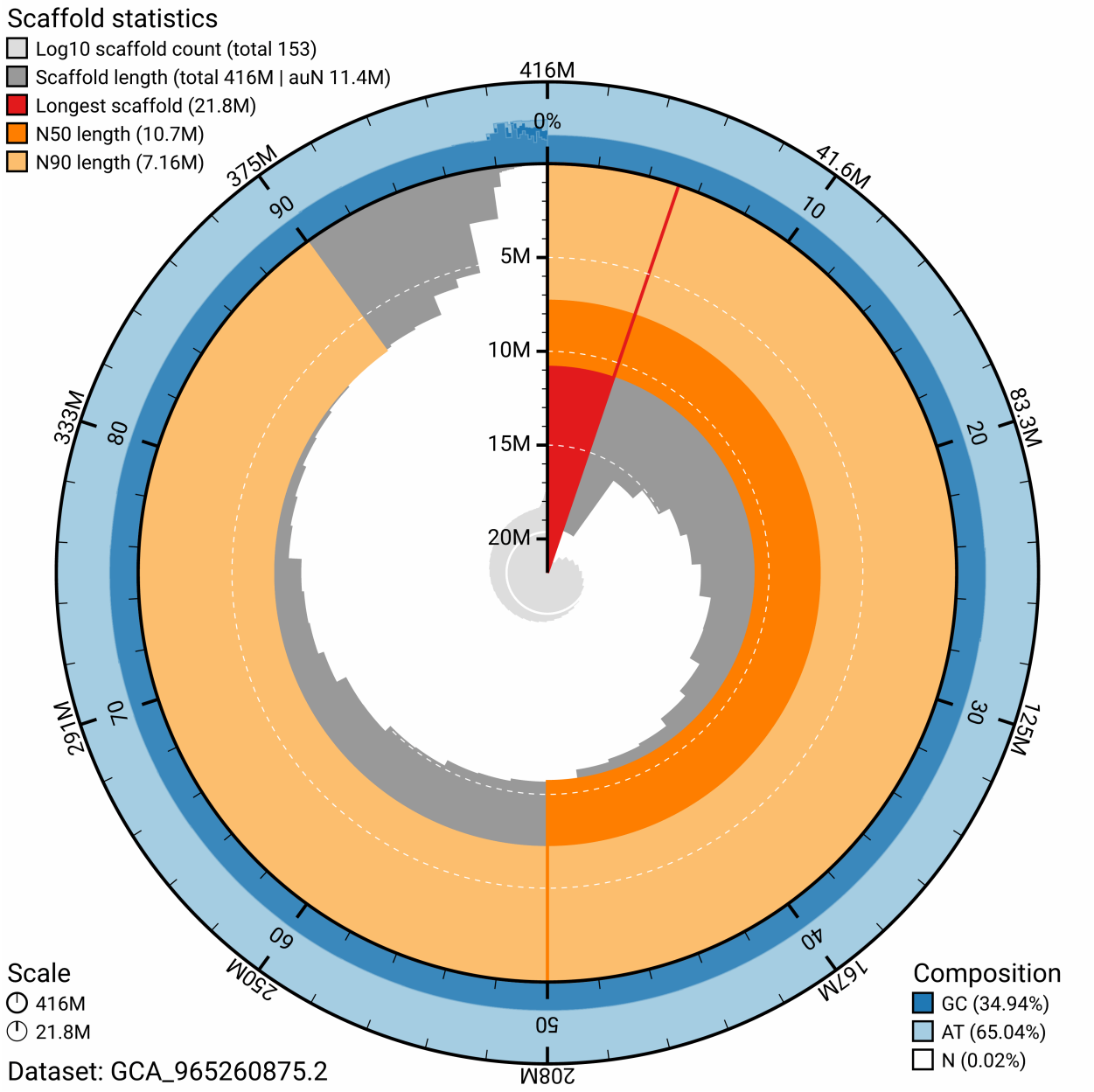

Figure 5. Assembly metrics for lpCarElat1.hap1.2, the haplotype 1 assembly. The BlobToolKit snail plot provides an overview of assembly contiguity and sequence composition. The circumference represents the length of the haplotype 1 assembly, and the main plot is divided into 1,000 bins around the circumference. 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.

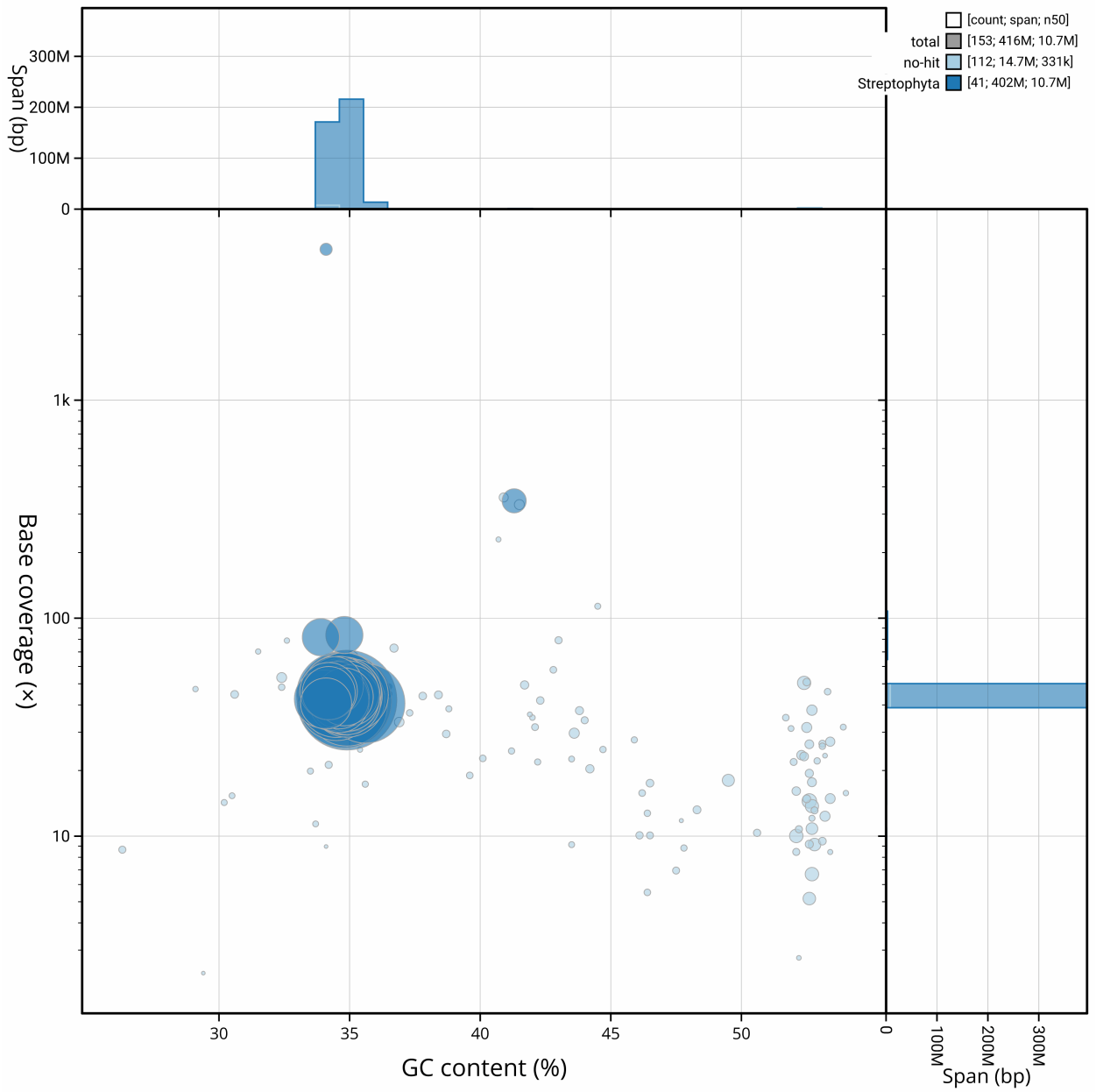

Figure 6. BlobToolKit blob plot for lpCarElat1.hap1.2, the haplotype 1 assembly. The plot shows base coverage (vertical axis) and GC content (horizontal axis). Circles represent scaffolds; marker area is proportional to scaffold length, and colour indicates taxonomic assignment. The histograms along 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 haplotype 1 assembly is **6.C.Q62**.

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

Measure	Value	*Benchmark*
EBP summary (haplotype 1)	6.C.Q62	*6.C.Q40*

Measure	Value	*Benchmark*
Contig N50 length	1.77 Mb	*> 1 Mb*
Scaffold N50 length	10.68 Mb	*= chromosome N50*
Consensus quality (QV)	Haplotype 1: 62.9; haplotype 2: 62.8; combined: 62.8	*# 40*
k-mer completeness	Haplotype 1: 68.86%; Haplotype 2: 68.97%; combined: 98.84%	*> 90% (combined assembly)*
BUSCO	Haplotype 1: C:97.4% [S:93.1%, D:4.3%], F:0.9%, M:1.7%, n:6,282; Haplotype 2: C:97.5% [S:93.1%, D:4.4%], F:0.9%, M:1.6%, n:6,282	*C > 90%*
Percentage of assembly assigned to chromosomes	98.33%	*# 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 summaries: haplotype 1 and haplotype 2. For plants, total BUSCO completeness (C) replaces the EBP single-copy benchmark to accommodate biological duplication.

Data availability

European Nucleotide Archive: Carex elata. Accession number PRJEB87945; https://identifiers.org/ena.embl/PRJEB87945. The genome sequence is released openly for reuse. The *C. elata* 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 haplotype 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	2.14.0	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobTk (hosted BlobDir)	0.5.1	https://github.com/genomehubs/blobtk
BlobTk (snail and blob plot renderer)	0.8.3	https://github.com/genomehubs/blobtk
BlobToolKit	4.4.5; 4.4.6	https://github.com/genomehubs/blobtoolkit
BUSCO	6.0.0; 5.7.1; 6.1.0	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
GenomeScope 2.0	2.1.0	https://github.com/tbenavi1/genomescope2.0

Software	Version	Source
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
MerquryFK	1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	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	23.10.0; 24.10.4; 25.10.0; 25.10.4	https://github.com/nextflow-io/nextflow
Oatk	1.0	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; 1.20; 1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.9.0; v0.7.1	https://github.com/sanger-tol/blobtoolkit
sanger-tol/busco	v0.3.0; 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	1.2.2	https://github.com/sanger-tol/treeval
YaHS	1.2a.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., Jarvis E. D., Fedrigo O.: **Haplotype-resolved assembly of diploid genomes without parental data**. *Nature Biotechnology*; 2022; **40**(9): 1332–1335. PubMed Abstract | 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 |

Manni Mosè, Berkeley Matthew R, Seppey Mathieu: **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 |

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 elata* All.**. *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 elata* All.**. 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 elata (lpCarElat1), all four checks passed.

Data available: pass. ENA study PRJEB87945, raw-read accessions ERR14835591, ERR14835901, ERR14835509, and ERR15551511, and assembly accessions GCA_965260875.2 and GCA_965260805.2 are public.

Earth BioGenome Project three-part metric: pass. Haplotype 1 is reported as 6.C.Q62, meeting the 6.C.Q40 reference standard.

k-mer completeness: pass. The combined haplotype assemblies have 98.84% k-mer completeness, exceeding the Earth BioGenome Project target of 90%.

BUSCO: pass. Haplotype 1, which is proposed as the reference assembly for this species, has 97.4% complete BUSCOs (93.1% single-copy and 4.3% 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.