DATA NOTE

REVISED The genome sequence of the Bicolour Sedge, *Triaenodes bicolor* (Curtis, 1834) (Trichoptera: Leptoceridae)

[version 2; peer review: 3 approved, 1 approved with reservations]

Derek Coleman[1], Maja Hrynowiecka[2],
Natural History Museum Genome Acquisition Lab,
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]Independent researcher, Carshalton, England, UK
[2]University of Nottingham, Nottingham, England, UK

v2 **First published:** 12 Jan 2026, **11**:26
https://doi.org/10.12688/wellcomeopenres.25183.1
Latest published: 05 Feb 2026, **11**:26
https://doi.org/10.12688/wellcomeopenres.25183.2

Abstract

We present a genome assembly from an individual male *Triaenodes bicolor* (Bicolour Sedge; Arthropoda; Insecta; Trichoptera; Leptoceridae). The assembly contains two haplotypes with total lengths of 615.75 megabases and 613.86 megabases. Most of haplotype 1 (99.48%) is scaffolded into 21 chromosomal pseudomolecules, including the Z sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 16.2 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

Triaenodes bicolor; Bicolour Sedge; genome sequence; chromosomal; Trichoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ? ✓ ✓ ✓

	1	2	3	4
version 2 (revision) 05 Feb 2026		✓ view	✓ view	✓ view
version 1 12 Jan 2026	? view			

1. **Christophe Klopp**, INRAE, Castanet-Tolosan, France
2. **Michael Hiller**, Senckenberg Research Institute, Frankfurt, Germany
3. **Ricardo Rodriguez-de-la-vega**, Université Paris-Saclay, Gif-sur-Yvette, France
4. **Panagiotis Ioannidis**, Foundation for Research & Technology - Hellas, Crete, Greece

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: Coleman D: Investigation, Resources; **Hrynowiecka M**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

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

How to cite this article: Coleman D, Hrynowiecka M, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the Bicolour Sedge, *Triaenodes bicolor* (Curtis, 1834) (Trichoptera: Leptoceridae) [version 2; peer review: 3 approved, 1 approved with reservations]** Wellcome Open Research 2026, **11**:26 https://doi.org/10.12688/wellcomeopenres.25183.2

First published: 12 Jan 2026, **11**:26 https://doi.org/10.12688/wellcomeopenres.25183.1

REVISED **Amendments from Version 1**

We have replaced the snail plot in Figure 5 with a corrected version, as the static image was different to the hosted version at https://blobtoolkit.genomehubs.org/view/Triaenodes%20bicolor/dataset/GCA_964276685.1/snail.

Any further responses from the reviewers can be found at the end of the article

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Trichoptera; Integripalpia; Brevitentoria; Leptoceroidea; Leptoceridae; Leptocerinae; Triaenodini; *Triaenodes*; *Triaenodes bicolor* (Curtis, 1834) (NCBI:txid699847)

Background

Triaenodes bicolor (Curtis, 1834) is a caddisfly species in the family Leptoceridae, commonly known as the 'Bicolour Sedge', or 'Angler's Bicolour Sedge' due to contrasting chestnut forewings and dark grey hind wings, and its use by the fishing community as bait (Barnard & Ross, 2012). The Leptoceridae are nicknamed the Long-horned Caddisflies for possessing long antennae that are often longer than twice the length of the forewings (Crofts, 2019).

This species is widespread across Britain & Ireland, and across much of mainland Europe, particularly northern Europe (GBIF Secretariat, 2025), and is one of two *Triaenodes* species present in Britain. The other, *T. ochreellus*, has rare sightings, no current UK breeding site and the adult is distinguished from *T. bicolor* by its bright yellow wings (Wallace *et al.*, 2022).

An adult *T. bicolor* has its flight period between May and September. During this time, it can be found around slow-flowing waters such as ponds, lakes, slow rivers, and permanently wet fens, where it is an important food source for fish such as trout (Barnard & Ross, 2012; Crofts, 2019). It can be identified by its long brown antennae that are annulated with white, in addition to its chestnut forewings and grey hindwings that are only noticeable from the side. The length of its wings ranges from 6–10 mm, with the wings of the female being notably larger by 1–3 mm. Furthermore, the abdomen of a female is larger and more robust, possessing two large lateral valves in contrast to the thin abdomen of a male (Barnard & Ross, 2012; McLachlan, 1867).

T. bicolor larvae occur in lakes and ponds with rich aquatic vegetation. They are easily distinguishable due to their swimming behaviour that is facilitated by metathoracic legs with fringed hairs which allow them to propel themselves with their case through the water. Cases are long and tapering, made up of consistently shaped pieces of leaf or small whole leaves in a spiral whorl, reaching 35 mm in length while the larva itself reaches 12–13 mm in length (Hickin, 1946).

This species can be easily confused with *Adicella reducta*, *Oecetis furva* and *Athripsodes aterrimus* (pale form) as a result of their similar colouring, however *A. reducta* has a pale patch on its wings, *O. furva* does not have a raised tornus and *A. aterrimus* has an obvious pale spot at the arculus (Wallace *et al.*, 2022).

There are no other known reported genome sequences of *T. bicolor,* and this one generated as part of the Darwin Tree of Life project will aid in understanding the biology, physiology, and ecology of the species. The assembly was produced using the Tree of Life pipeline from a specimen collected in Thompson Common, Norfolk, UK (Figure 1).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Triaenodes bicolor* (specimen ID NHMUK015059315, ToLID iiTriBico1; Figure 1), collected from Thompson Common, Norfolk, UK (latitude 52.53, longitude 0.85) on 2022-07-06. The specimen was collected and identified by Derek Coleman. For the Darwin Tree of Life sampling and metadata approach, refer to Lawniczak *et al.* (2022).

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). 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 iiTriBico1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the whole organism was homogenised by powermashing using a PowerMasher II tissue disruptor.

Figure 1. Photograph of the *Triaenodes bicolor* (iiTriBico1) specimen used for genome sequencing.

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol. We used centrifuge-mediated fragmentation to produce DNA fragments in the 8–10 kb range, following the Covaris g-TUBE protocol for ultra-low input (ULI). Sheared DNA was purified by manual SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 3.67 ng/μL and a yield of 172.49 ng, with a fragment size of 13.4 kb. The 260/280 spectrophotometric ratio was 2.13, and the 260/230 ratio was 6.93.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Prior to library preparation, the DNA was fragmented to ~10 kb. Ultra-low-input (ULI) libraries were prepared using the PacBio SMRTbell® Express Template Prep Kit 2.0 and gDNA Sample Amplification Kit. Samples were normalised to 20 ng DNA. Single-strand overhang removal, DNA damage repair, and end-repair/A-tailing were performed according to the manufacturer's instructions, followed by adapter ligation. A 0.85× pre-PCR clean-up was carried out with Promega ProNex beads.

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

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

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 μL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the iiTriBico1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

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 and equimolar and/or weighted 2.8 nM pools were created. 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.

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 in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

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 29 breaks, 37 joins, and removal of 19 haplotypic duplications. 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 both haplotypes using the *k*-mer database (k = 31) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). 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 the 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 containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Triaenodes bicolor* specimen generated 23.00 Gb (gigabases) from 2.10 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 581.46 Mb, with a heterozygosity of 1.14% and repeat content of 14.73% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided

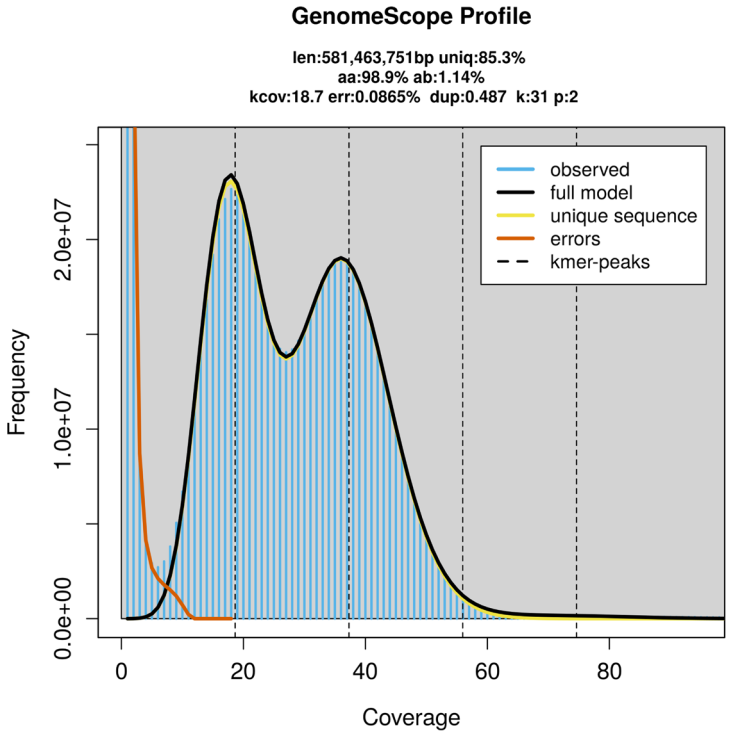

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

approximately 37× coverage. Hi-C sequencing produced 159.96 Gb from 1 059.36 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 615.75 Mb in 182 scaffolds, with 1 456 gaps, and a scaffold N50 of 29.5 Mb (Table 2).

Most of the assembly sequence (99.48%) was assigned to 21 chromosomal-level scaffolds, representing 20 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). The Z chromosome was identified by alignment to *Ceraclea dissimilis* (GCA_963576895.1).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

For haplotype 1, the estimated QV is 61.0, and for haplotype 2, 61.0. When the two haplotypes are combined, the assembly achieves an estimated QV of 61.0. The *k*-mer completeness is 79.35% for haplotype 1, 79.24% for haplotype 2, and 99.67% for the combined haplotypes (Figure 4).

BUSCO analysis using the endopterygota_odb10 reference set (n = 2 124) identified 96.1% of the expected gene set (single = 94.8%, duplicated = 1.3%) for haplotype 1. The snail plot in

Table 1. Specimen and sequencing data for BioProject PRJEB79787.

Platform	PacBio HiFi	Hi-C
ToLID	iiTriBico1	iiTriBico1
Specimen ID	NHMUK015059315	NHMUK015059315
BioSample (source individual)	SAMEA112963035	SAMEA112963035
BioSample (tissue)	SAMEA112963128	SAMEA112963128
Tissue	whole organism	whole organism
Instrument	Revio	Illumina NovaSeq 6000
Run accessions	ERR13650042	ERR13654253
Read count total	2.10 million	1 059.36 million
Base count total	23.00 Gb	159.96 Gb

Table 2. Genome assembly statistics.

Assembly name	iiTriBico1.hap1.1	iiTriBico1.hap2.1
Assembly accession	GCA_964276685.1	GCA_964276555.1
Assembly level	chromosome	scaffold
Span (Mb)	615.75	613.86
Number of chromosomes	21	Scaffold-level
Number of contigs	1 638	1 637
Contig N50	0.71 Mb	0.67 Mb
Number of scaffolds	182	152
Scaffold N50	29.5 Mb	30.65 Mb
Longest scaffold length (Mb)	49.48	-
Sex chromosomes	Z	-
Organelles	Mitochondrion: 16.2 kb	-

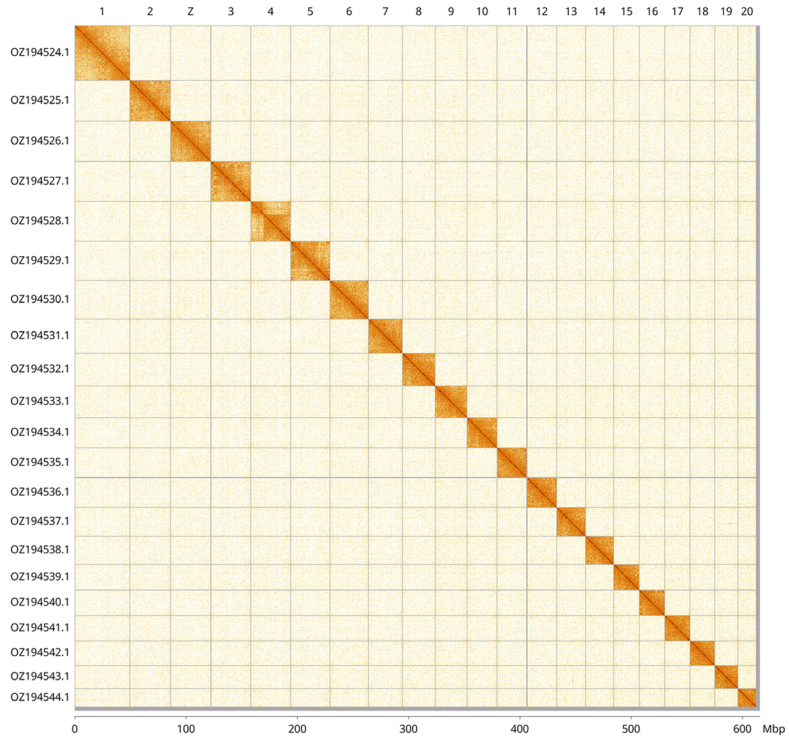

Figure 3. Hi-C contact map of the *Triaenodes bicolor* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Triaenodes bicolor* iiTriBico1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ194524.1	1	49.48	32
OZ194525.1	2	36.60	31
OZ194527.1	3	35.81	31.50
OZ194528.1	4	35.76	32.50
OZ194529.1	5	35.36	31.50
OZ194530.1	6	34.59	31.50
OZ194531.1	7	30.48	31.50
OZ194532.1	8	29.50	31.50
OZ194533.1	9	28.58	31.50
OZ194534.1	10	27.05	31.50
OZ194535.1	11	27.01	32
OZ194536.1	12	26.58	31.50
OZ194537.1	13	25.87	32
OZ194538.1	14	25.29	31.50
OZ194539.1	15	23.08	32
OZ194540.1	16	22.84	32
OZ194541.1	17	22.73	32
OZ194542.1	18	22.18	32
OZ194543.1	19	20.85	32
OZ194544.1	20	16.51	33
OZ194526.1	Z	36.38	31

Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **5.C.Q61**.

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

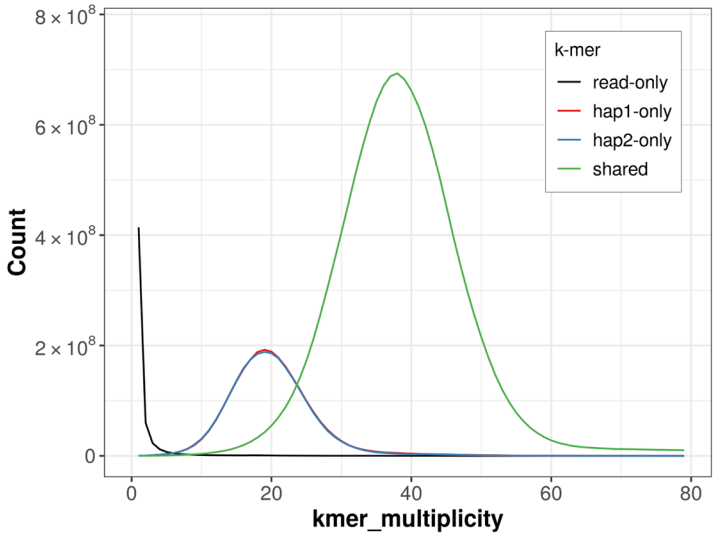

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.

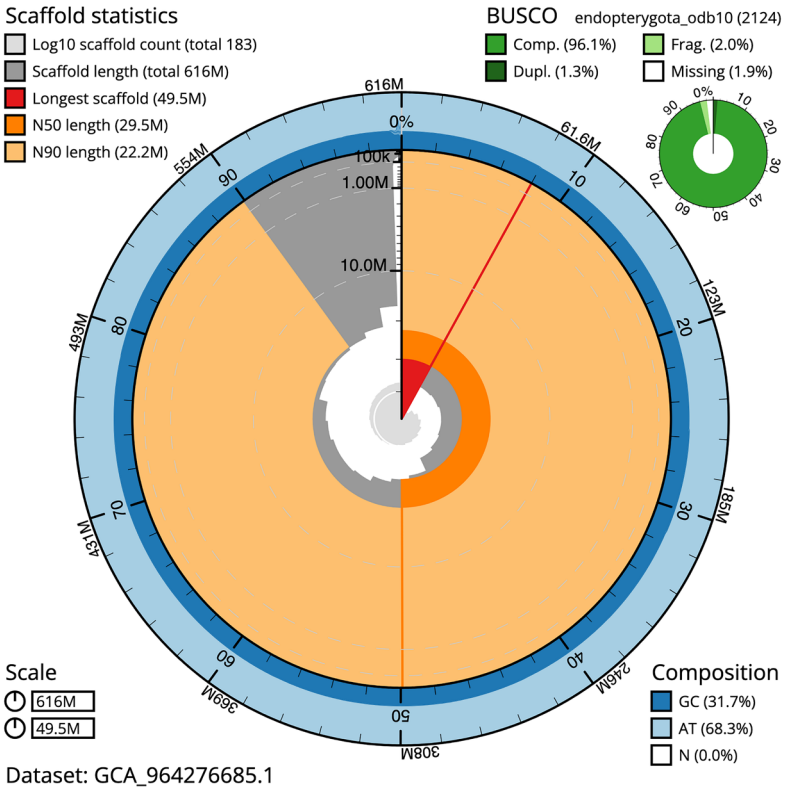

Figure 5. Assembly metrics for iiTriBico1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

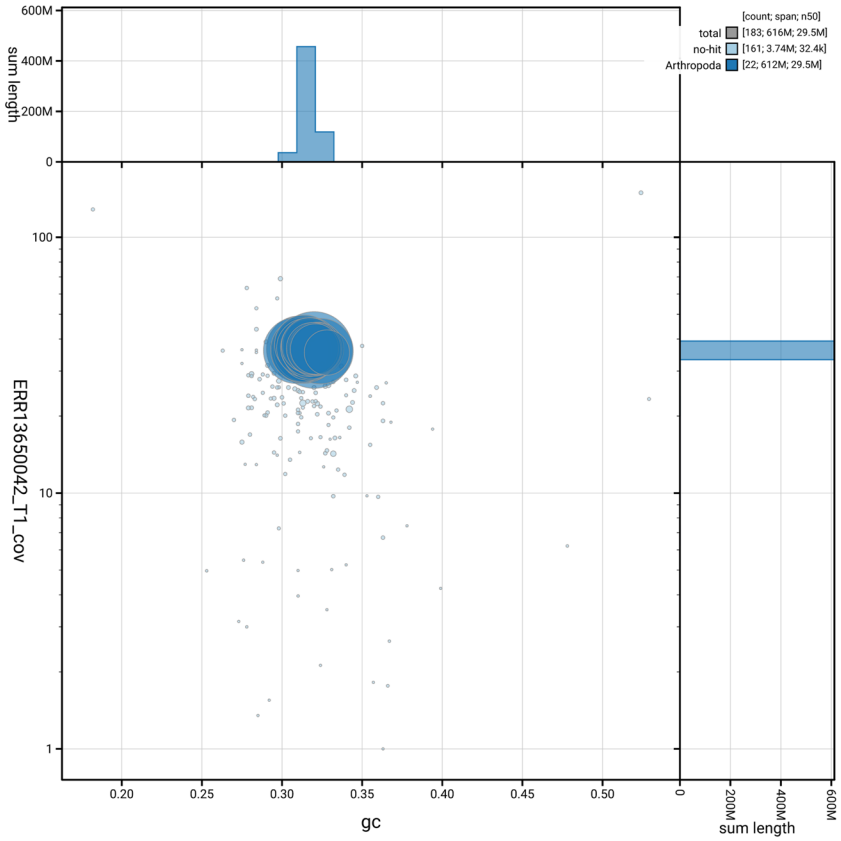

Figure 6. BlobToolKit GC-coverage plot for iiTriBico1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Triaenodes bicolor* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	5.C.Q61	6.C.Q40
Contig N50 length	0.71 Mb	≥ 1 Mb
Scaffold N50 length	29.50 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 61.0; haplotype 2: 61.0; combined: 61.0	≥ 40
k-mer completeness	Haplotype 1: 79.35%; Haplotype 2: 79.24%; combined: 99.67%	≥ 95%
BUSCO	C:96.1% [S:94.8%; D:1.3%]; F:2.0%; M:1.9%; n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.48%	≥ 90%

to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Further, the Wellcome Sanger Institute employs a process whereby

due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

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

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

Data availability

European Nucleotide Archive: Triaenodes bicolor. Accession number PRJEB79787. The genome sequence is released openly for reuse. The *Triaenodes bicolor* 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 Table 1 and Table 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Author information

Contributors are listed at the following links:

- Members of the Natural History Museum Genome Acquisition Lab
- 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 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
Hifiasm	0.19.8-r603	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

Software	Version	Source
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	-	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
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
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Barnard PC, Ross E: **The adult Trichoptera (Caddisflies) of Britain and Ireland.** Royal Entomological Society, 2012.
Reference Source

Bateman A, Martin MJ, 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 HG: **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 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Crofts SM: **Caddisfly adults (Trichoptera) of Britain and Ireland.** Freshwater Biological Association, 2019.
Reference Source

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

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

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

GBIF Secretariat: ***Triaenodes bicolor* (Curtis, 1834).** In: *GBIF Backbone Taxonomy*. 2025.
Reference Source

Hickin NW: **Larvae of the British Trichoptera.** *Transactions of the Royal Entomological Society of London.* 1946; **97**(8): 187–212.
Publisher Full Text

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

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

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

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

Lawniczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2022; **7**: 187.
Publisher Full Text

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
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

McLachlan R: **Trichoptera britannica: *a monograph of the British species of caddis-flies*.** *Transactions of the Royal Entomological Society of London.* 1867; **15**(1): 1–184.
Publisher Full Text

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Reference Source

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

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

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

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2024; **9**: 339.
PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
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.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Wallace I, Flint S, Flint P: **Adult Caddis (Trichoptera) of Britain and Ireland: a practical guide.** Field Studies Council, 2022.
Reference Source

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

Open Peer Review

Current Peer Review Status: ? ✓ ✓ ✓

Version 2

Reviewer Report 04 May 2026

https://doi.org/10.21956/wellcomeopenres.28591.r147622

✓ **Panagiotis Ioannidis**
Foundation for Research & Technology - Hellas, Crete, Greece

The present manuscript describes the sequencing and assembly of Triaenodes bicolor, a trichopteran insect.

The genome assembly is of high quality since virtually all generated contigs have been assigned to one of the 21 chromosome-level scaffolds. All assembly statistics are excellent; BUSCO score is at 96.1% (complete), QV is 61, and k-mer completeness is at 99.67%. Moreover, the blob plot is showing that the vast majority of the assembled contigs have a GC content around 32%, while the sequencing coverage is centered around 39x. Both of these values are the expected ones.

The authors report that "the genome will be annotated using available RNA-Seq data...". After checking the submitted data using the provided BioProject ID, I could not find any RNAseq data associated with this particular BioProject ID. So, do the authors mean that they will use RNAseq data from other, (presumably) related species? Because that would greatly limit the sensitivity of gene prediction. It would be nice if the authors could clarify this point. And also, please don't forget to run BUSCO on the predicted gene set and compare it with the BUSCO score from the genome, in order to understand how well the gene prediction pipeline worked.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: bioinformatics, insect genomics

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.

Reviewer Report 16 March 2026

https://doi.org/10.21956/wellcomeopenres.28591.r148667

Ricardo Rodriguez-de-la-vega
Université Paris-Saclay, Gif-sur-Yvette, France

Comments on Coleman and Hrynowiecka "The genome sequence of the Bicolour Sedge..."

In this Darwin Tree of Life (DToL) project genome note, authors describe the chromosome-level fully phased genome assemblies of a male *Triaenodes bicolor*. Raw data included HiFi and Arima-HiC v2 reads obtained with standard protocols of the Wellcome Sanger Institute (WSI) from the same specimen. The overall assembly protocol followed established methods of the WSI Tree of Life programme, including HiC-integrated hifiasm assembly, bwa-mem2 HiC mapping and YaHS scaffolding. The reported assemblies are among the highest quality among DToL releases, particulary in terms of the almost perfect assignment to chromosomal-level scaffolds for both haplotypes.

At the time of publication, this was the 13th Trichoptera genome released by the DToL project and the first after a hiatus of over one year. The assembly presented attained a chromosome level while the other haplotype remains at the scaffold level, though the quality metrics are nearly identical. Haplotype 2 have some very large scaffolds, likely due to misjoins that may have been corrected in Haplotype 1 (manual curation is not explicitly mentioned but is common in DToL assemblies).

One final observation: most chromosomes in the primary assembly align to a single scaffold in the alternate haplotype assembly (using minimap2 -cx asm5), with over 90% of their length included in matching blocks. The alternate haplotype is of comparable quality to the primary assembly, though the misjoins should be addressed for downstream applications.

I am including the links to Genomes on a Tree and Tree of Life quality check pages as I think are useful points to access the numerous underlying data analyses.

(refer to 1,2)

References

1. https://goat.genomehubs.org/record?recordId.
2. https://tolqc.cog.sanger.ac.uk/darwin/insects/Triaenodes_bicolor/.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Bioinformatics, Evolution

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.

Reviewer Report 06 March 2026

https://doi.org/10.21956/wellcomeopenres.28591.r149462

Michael Hiller
Senckenberg Research Institute, Frankfurt, Germany

The data note describes a high-quality chromosome-level assembly of Triaenodes bicolor.
All sequencing and assembly methods are described in great detail.

My only comment is that the contig N50 is <1 Mb, which is likely the result of GC-dependent amplification bias of the polymerases used in the original ULI protocol. I wonder if it was necessary to use the ULI protocol on a species of that size?

As a note, the identity of the A and B polymerases in the original ULI protocol are now known: NEBNext® High-Fidelity 2X PCR Master Mix (PacBio Ultra-Low DNA Input A) and TaKaRa Bio PrimeSTAR® GXL DNA Polymerase (PacBio Ultra-Low DNA Input B).
This could be potentially added.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: comparative genomics, genome assembly

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.

Version 1

Reviewer Report 28 January 2026

https://doi.org/10.21956/wellcomeopenres.27758.r144952

?

Christophe Klopp
INRAE, Castanet-Tolosan, France

The authors *provide the first assembly of Bicolour Sedge, Triaenodes bicolor.* They have used strandard procotols mixing HifI and Hi-C data and have achieved a very contiguous haplotyped assemblies.

The busco scores presented in the snail plot seem strange because usually complete + fragmented + missing = 100%. This is not the case here. This should be checked.

The genomescope2 and merqury kmers histograms do not present the same balance between haplotype specific and shared kmers. This is not expected. This should be checked and explained.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly and annotation

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, however I have significant reservations, as outlined above.

Author Response 28 Jan 2026

Tree of Life Team Sanger

Thank you for reviewing this article. The BUSCO values reported in the text and shown in the hosted (interactive) version of the snail plot are C: 96.1% [S: 94.8%; D: 1.3%]; F: 2.0%; M: 1.9%; *n* = 2,124, which sum to 100%.
We identified that the static image of the snail plot originally included in the manuscript did not reflect the current BUSCO values for this dataset. This figure has now been replaced with a corrected version. GenomeScope2 analyses *read-derived* k-mer multiplicity and does not use the assembly. It models k-mer frequencies from the raw sequencing reads to infer genome properties such as heterozygosity and repeat content statistically. In contrast, MerquryFK evaluates *assembly-derived* haplotypes by partitioning read k-mers according to their presence in the assembled sequences. As these analyses use different inputs and address different questions, the relative proportions of inferred heterozygous k-mers in GenomeScope2 and haplotype-specific k-mers in MerquryFK are not expected to match. The correspondence between the heterozygous and homozygous coverage peaks inferred by GenomeScope2 and the haplotype-specific and shared peaks observed in MerquryFK nevertheless indicates consistency between the two analyses.

Competing Interests: No competing interests were disclosed.