



DATA NOTE

The genome sequence of the great crested newt, *Triturus cristatus* (Laurenti, 1768)

[version 1; peer review: 2 approved]

Jeffrey W. Streicher ¹, Stephanie Holt ^{1,2},
Natural History Museum Genome Acquisition Lab,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory
team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

¹Natural History Museum, London, England, UK

²University of Oxford, Oxford, England, UK

V1 First published: 23 May 2025, 10:270
<https://doi.org/10.12688/wellcomeopenres.24257.1>
Latest published: 23 May 2025, 10:270
<https://doi.org/10.12688/wellcomeopenres.24257.1>

Abstract

We present a genome assembly from a female specimen of *Triturus cristatus* (great crested newt; Chordata; Amphibia; Caudata; Salamandridae). The genome sequence has a total length of 22,324.62 megabases. Most of the assembly (98.78%) is scaffolded into 12 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 16.54 kilobases.

Keywords

Triturus cristatus, (Great or Northern) Crested Newt, genome sequence, chromosomal, Caudata



This article is included in the [Tree of Life](#) gateway.

Open Peer Review

Approval Status

	1	2
version 1		
23 May 2025	view	view

1. **R Alexander Pyron** , The George Washington University Museum and Textile Museum, Washington, USA
2. **Andrew J Crawford** , Universidad de Los Andes Bogotá, Bogotá, Colombia

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: **Streicher JW:** Conceptualization, Data Curation, Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Holt S:** Conceptualization, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

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.

Copyright: © 2025 Streicher JW *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: Streicher JW, Holt S, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the great crested newt, *Triturus cristatus* (Laurenti, 1768) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, 10:270 <https://doi.org/10.12688/wellcomeopenres.24257.1>

First published: 23 May 2025, 10:270 <https://doi.org/10.12688/wellcomeopenres.24257.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Pleurodelinae; *Triturus*; *Triturus cristatus* (Laurenti, 1768) (NCBI:txid8323)

Background

The great crested newt, *Triturus cristatus*, is widely distributed throughout northern Eurasia. It occurs in mixed and deciduous forests of Austria, Belarus, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Latvia, Liechtenstein, Lithuania, Luxembourg, Moldova, the Netherlands, Norway, Poland, Romania, the Russian Federation, Serbia, Slovakia, Sweden, Switzerland, Ukraine, and the United Kingdom ([AmphibiaWeb, 2021](#)). Also known as northern crested newt or great warty newt, the striking dorsal crest observed in breeding adult males has been used to study the evolutionary origin of phenotypic plasticity and novel traits ([Wiens et al., 2011](#)). Populations of *T. cristatus* in Great Britain are widely distributed in England and found in some parts of Wales and Scotland ([Arnold, 1995](#)). The species is protected by law in the United Kingdom with robust monitoring programmes in place ([Buxton & Griffiths, 2022](#); [Griffiths et al., 2010](#)). Known to be susceptible to the fungal pathogen *Batrachochytrium salamandrivans*, *T. cristatus* is a Biodiversity Action Plan priority species, listed on Appendix II of the Bern Convention and on Annexes II and IV of the EU Natural Habitats Directive in Europe ([Cunningham et al., 2019](#)).

The genome was sequenced from one female *T. cristatus* ([Figure 1A, B, C](#)) collected by net from a garden pond at Gilbert White's House and Gardens in Selbourne, Hampshire, England – a museum and site of historical significance in

British natural history. The museum, located at The Wakes – White's home until his death in 1793 – commemorates his life and work as a pioneering naturalist and educator, and hosts exhibitions on other figures in exploration and natural history. White was a pioneering English naturalist who is often credited as England's first ecologist, based on his book, *The Natural History and Antiquities of Selborne* ([White, 1789](#)), journals, and earlier publications in the *Philosophical Transactions of the Royal Society*. His work marked a step-change in observational natural history and influenced much contemporary work such as Thomas Pennant's groundbreaking *British Zoology* (1778), as well as many who followed him, including Charles Darwin and Sir David Attenborough. In *The Natural History and Antiquities of Selborne*, he discusses English newts in relation to the developing knowledge of complex and diverse amphibian lifecycles: 'I used to take it for granted that the *salamandra aquatica* [the newt] was hatched, lived, and died in the water. But John Ellis, Esq., F.R.S. (the coralline Ellis) asserts, in a letter to the Royal Society, dated June the 5th, 1766, in his account of the mud inguana [sic], an amphibious bipes [a reference to the obligate aquatic salamander *Siren lacertina*, which like the squamate genus *Bipes* possesses only forelimbs], from South Carolina, that the water-eft, or newt, is only the larva of the land-eft, as tadpoles are of frogs' ([White, 1789](#)). This early recognition of different life histories in caudatan amphibians, where some remain completely aquatic as adults, retaining larval features like external gills (e.g. *Siren lacertina*), while others undergo metamorphosis to become terrestrial (or semi-terrestrial) adults. (e.g. *Triturus cristatus*), marked an important advance in understanding the ecological diversity of amphibians.

The great crested newt and its close relative *T. marmoratus* have an atypical heteromorphy in their genomes; a chromosome 1 syndrome ([Macgregor & Horner, 1980](#); [Wielstra, 2020](#)). This syndrome leads to a so-called 'balanced lethal system' which



Figure 1. Female voucher specimen of *Triturus cristatus* (NHMUK 2022.7580, Field ID, JWS 916; Snout-vent length 82.6 mm) from which the genome was sequenced in lateral (**A**), dorsal (**B**), and ventral (**C**) views. The individual was collected from Gilbert White's House and Gardens in Selbourne, England. Two views of the garden pond where the specimen was collected via dip netting (**D-E**).

results in mortality of 50% of their offspring (Wielstra, 2020). The 50% mortality in *T. cristatus* is explained by each variant of chromosome 1 containing a lethal recessive allele. In offspring that are homozygous for either of the chromosome morphs, development ceases halfway through embryogenesis, whereas development completes in heterozygous individuals (Fahrbach & Gerlach, 2018). The evolutionary origins of (seemingly wasteful) balanced lethal systems remain enigmatic but could be related to ancestral sex chromosomes (Grossen *et al.*, 2012) or collapsed supergenes (Schwander *et al.*, 2014). In our assembly, we see evidence of chromosome 1 being heteromorphic with one arm of chromosome 1 having half the coverage expected.

Like other salamanders and newts (Amphibia: Caudata), *T. cristatus* has an extremely large genome (Liedtke *et al.*, 2018), with previous *c*-value estimates as large as 27.77 pg (Litvinchuk *et al.*, 2007). All cytogenetics studies of *Triturus* species (including *T. cristatus*) have reported a diploid $2n = 24$ (Baldwin & Macgregor, 1985), which matches the 12 chromosomes resulting from our assembly. Due to their extremely large size, five chromosomes were split into two pseudomolecules each, resulting in 17 pseudomolecules total. The estimated size of the genome (22 Gb) is larger than a previously sequenced salamandrid genome from *Pleurodeles waltii* (20 Gb, Brown *et al.*, 2025), but smaller than the genome sequenced from the axolotl, *Ambystoma mexicanum* (32 Gb, Nowoshilov *et al.*, 2018). The chromosome-level whole genome sequence for *T. cristatus* presented here was produced as part of the Darwin Tree of Life Project, Sanger Tree of life Project and Vertebrate Tree of Life Project.

Genome sequence report

Sequencing data

The genome of a specimen of *Triturus cristatus* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long

reads, generating 809.03 Gb from 65.40 million reads, which were used to assemble the genome. GenomeScope analysis estimated the haploid genome size at 21,641.98 Mb (22 Gb), with a heterozygosity of 0.28% and repeat content of 48.53%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 36 coverage. Hi-C sequencing produced 920.77 Gb from 6,097.79 million reads, and was used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The assembly was improved by manual curation, which corrected 324 misjoins or missing joins. These interventions reduced the total assembly length by 5.49%, decreased the scaffold count by 12.6%, and also decreased the scaffold N50 by 32.78%. The final assembly has a total length of 22,324.62 Mb in 1,303 scaffolds, with 4,698 gaps, and a scaffold N50 of 1248.77 Mb (Table 2).

The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (98.78%) was assigned to 12 chromosomal-level scaffolds, with each of chromosomes 1 to 5 split into two parts for submission to INSDC databases. These chromosome-level scaffolds, confirmed by Hi-C data, are named

Table 1. Specimen and sequencing data for *Triturus cristatus*.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	aTriCri1	aTriCri1	aTriCri1
Specimen ID	NH Muk014561658	NH Muk014561658	NH Muk014561658
BioSample (source individual)	SAMEA112468129	SAMEA112468129	SAMEA112468129
BioSample (tissue)	SAMEA112468181	SAMEA112468178	SAMEA112468179
Tissue	liver	muscle	ovary
Sequencing platform and model	Revio	Illumina NovaSeq 6000	Illumina NovaSeq X
Run accessions	ERR13033468; ERR13033469; ERR13033470; ERR13033471; ERR13033476; ERR13033467; ERR13033472; ERR13033475; ERR13033473; ERR13033474	ERR13063105	ERR13493948
Read count total	65.40 million	6 097.79 million	112.22 million
Base count total	809.03 Gb	920.77 Gb	16.95 Gb

Table 2. Genome assembly data for *Triturus cristatus*.

Genome assembly		
Assembly name	aTriCri1.1	
Assembly accession	GCA_964204655.1	
Alternate haplotype accession	GCA_964204665.1	
Assembly level for primary assembly	chromosome	
Span (Mb)	22,324.62	
Number of contigs	6,001	
Number of scaffolds	1,303	
Longest scaffold (Mb)	1950.67	
Assembly metric	Measure	Benchmark
Contig N50 length	8.5 Mb	≥ 1 Mb
Scaffold N50 length	1,248.77 Mb	= chromosome N50
Consensus quality (QV)	Primary: 65.4; alternate: 60.6; combined: 62.4	≥ 40
k-mer completeness	Primary: 94.04%; alternate: 45.11%; combined: 99.07%	≥ 95%
BUSCO*	C:92.7%[S:88.8%,D:3.8%], F:2.3%,M:5.0%,n:5,310	S > 90%; D < 5%
Percentage of assembly mapped to chromosomes	98.78%	≥ 90%
Sex chromosomes	Not identified	localised homologous pairs
Organelles	Mitochondrial genome: 16.54 kb	complete single alleles

* BUSCO scores based on the tetrapoda_odb10 BUSCO set using version 5.7.1. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

according to size (Figure 5; Table 3). During assembly curation, we noted that one arm of chromosome 1 (chromosome 1_1) has half-coverage, which is a completely heterozygous region related to the chromosome 1 heteromorphy of *T. cristatus* (Macgregor & Horner, 1980; Wielstra, 2020).

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.

Assembly quality metrics

The estimated Quality Value (QV) and k-mer completeness metrics, along with BUSCO completeness scores, were calculated for each haplotype and the combined assembly. The QV reflects the base-level accuracy of the assembly, while k-mer completeness indicates the proportion of expected k-mers identified in the assembly. BUSCO scores provide a measure of completeness based on benchmarking universal single-copy orthologues.

The combined primary and alternate assemblies achieve an estimated QV of 62.4. The k-mer recovery for the primary haplotype is 94.04%, and for the alternate haplotype

45.11%; the combined primary and alternate assemblies have a k-mer recovery of 99.07%. BUSCO v.5.7.1 analysis using the tetrapoda_odb10 reference set (n = 5,310) identified 92.7% of the expected gene set (single = 88.8%, duplicated = 3.8%).

Table 2 provides assembly metric benchmarks adapted from Rhie et al. (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The assembly achieves the EBP reference standard of 6.C.Q65.

Methods

Sample acquisition

The genome was sequenced from one female *T. cristatus* (specimen ID NHMUK014561658, ToLID aTriCri1; Figure 1A, B, C) collected by net from a garden pond at Gilbert White’s House and Gardens in Selbourne, England on 14 August 2022 (latitude 51.0975, longitude −0.9443; Figure 1D, E) at 13:45 hrs. The individual had a snout-vent length of 82.6 mm and a mass of 12.2 g. The specimen was collected and identified by Jeff Streicher and Stephanie Holt (Natural History Museum) and preserved dry freezing (−80 °C). The voucher specimen was deposited at the Natural History Museum, London (NHMUK 2022.7580, Field ID, JWS 916).

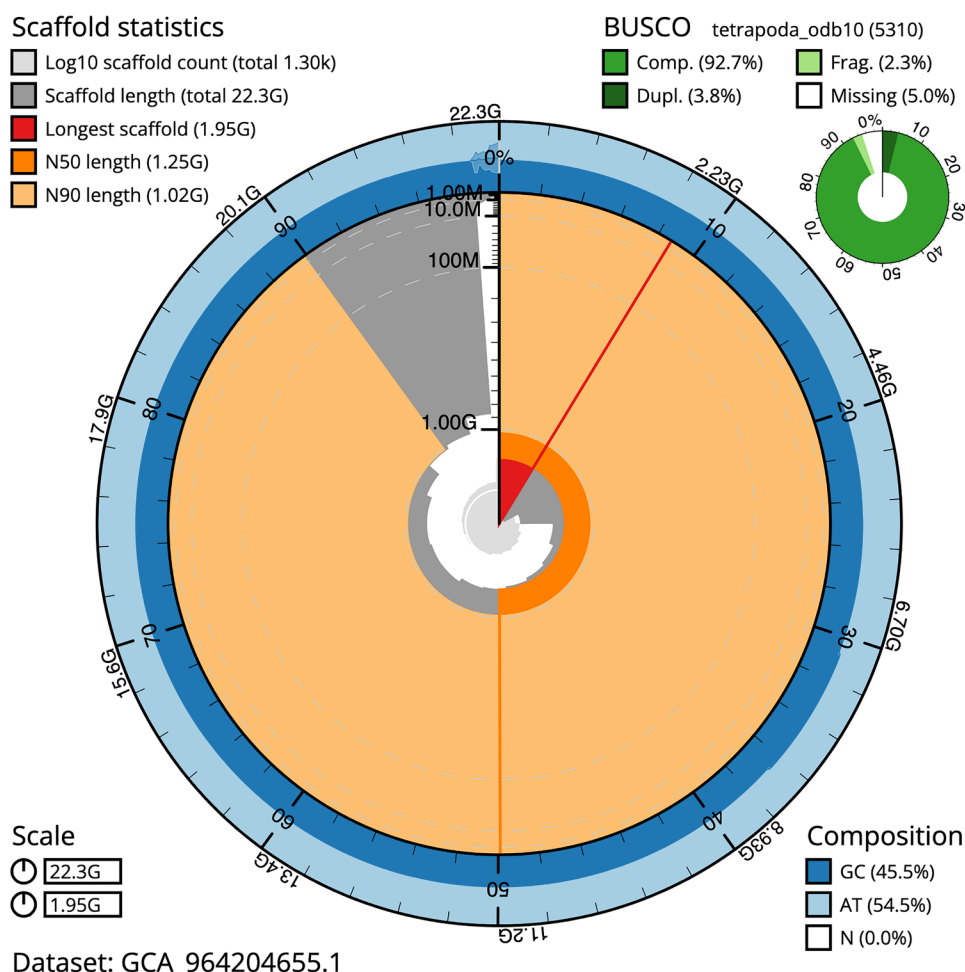


Figure 2. Genome assembly of *Triturus cristatus*, aTriCri1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. 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 tetrapoda_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964204655.1/dataset/GCA_964204655.1/snail.

Metadata collection for samples adhered to the Darwin Tree of Life project standards described by Lawniczak *et al.* (2022).

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The aTriCri1 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the liver was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the MagAttract v2 protocol (Oatley *et al.*, 2023a). DNA was sheared into an average fragment size

of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Nowoshilow *et al.*, 2018; Oatley *et al.*, 2023b). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from ovary tissue of aTriCri1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the

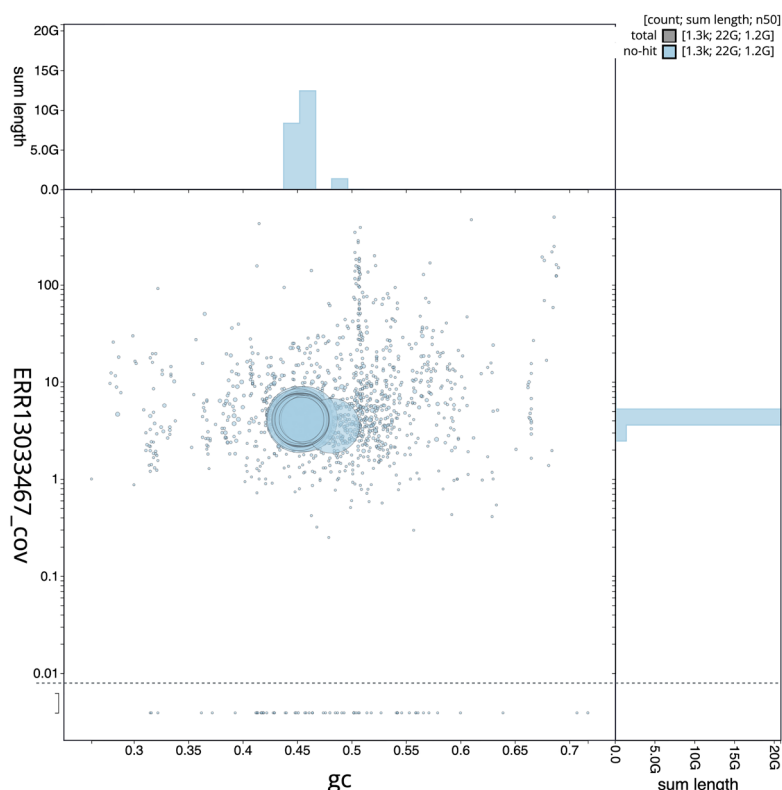


Figure 3. Genome assembly of *Triturus cristatus*, aTriCri1.1: BlobToolkit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. 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 at https://blobtoolkit.genomehubs.org/view/GCA_964204655.1/dataset/GCA_964204655.1/blob.

Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Hi-C sample preparation and crosslinking

Hi-C data were generated from the muscle of the aTriCri1 sample using the Arima-HiC v2 kit (Arima Genomics) with 20–50 mg of frozen tissue (stored at -80°C). As per manufacturer's instructions, tissue was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde concentration, and a final formaldehyde concentration of 2%. The tissue was then homogenised using the Diagenode Power Masher-II. The crosslinked DNA was digested using a restriction enzyme master mix, then biotinylated and ligated. A clean up was performed with SPRIselect beads prior to library preparation. DNA concentration was quantified using the Qubit Fluorometer v4.0 (ThermoFisher Scientific) and Qubit HS Assay Kit, and sample biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core.

PacBio HiFi

At a minimum, samples were required to have an average fragment size exceeding 8 kb and a total mass over 400 ng to proceed to the low-input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 as per the manufacturer's instructions. The kit includes the reagents required for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead cleanup, and nuclease treatment. Size-selection and clean-up were carried out using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using the Qubit Fluorometer v4.0 (ThermoFisher Scientific) with Qubit 1X dsDNA HS assay kit and the final library fragment size analysis was carried out using the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) and the gDNA 55kb BAC analysis kit.

Samples were sequenced on a Revio instrument (Pacific Biosciences, California, USA). Prepared libraries were normalised to 2 nM, and 15 μL was used for making complexes. Primers were annealed and polymerases were bound to create circularised complexes according to manufacturer's instructions. The complexes were purified with the 1.2X clean up with SMRTbell

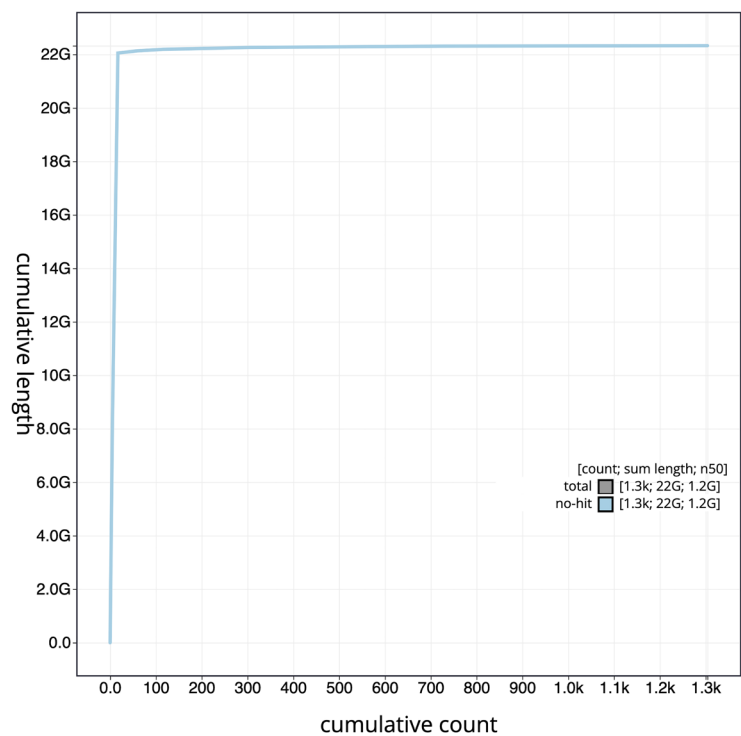


Figure 4. Genome assembly of *Triturus cristatus*, aTriCri1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964204655.1/dataset/GCA_964204655.1/cumulative.

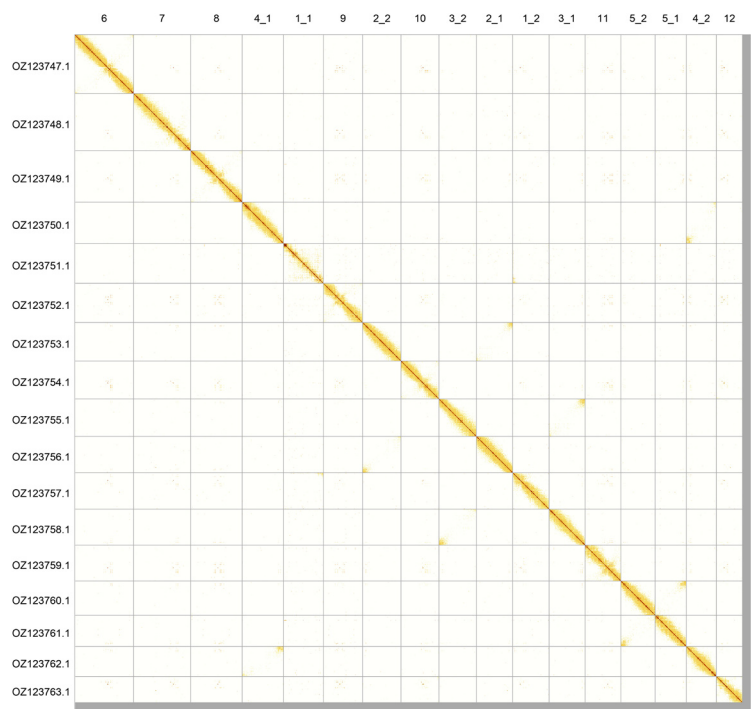


Figure 5. Genome assembly of *Triturus cristatus*: Hi-C contact map of the aTriCri1.1 assembly, produced in PretextView. Chromosomes are shown in order of size from left to right and top to bottom.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Triturus cristatus*, aTriCri1.

INSDC accession	Name	Length (Mb)	GC%
OZ123751.1	1_1	1,317.4	48
OZ123757.1	1_2	1,199.17	45.5
OZ123756.1	2_1	1,201.42	45.5
OZ123753.1	2_2	1,269.97	45
OZ123758.1	3_1	1,184.55	45.5
OZ123755.1	3_2	1,238.91	45.5
OZ123750.1	4_1	1,361.38	45.5
OZ123762.1	4_2	993.38	45.5
OZ123761.1	5_1	1,024.25	45
OZ123760.1	5_2	1,134.71	45
OZ123747.1	6	1,950.67	45
OZ123748.1	7	1,882.94	45.5
OZ123749.1	8	1,702.34	45
OZ123752.1	9	1,293.89	45.5
OZ123754.1	10	1,248.77	45
OZ123759.1	11	1,183.99	45.5
OZ123763.1	12	864.23	45.5
OZ123764.1	MT	0.02	41.5

beads. The purified complexes were then diluted to the Revio loading concentration (in the range 200–300 pM), and spiked with a Revio sequencing internal control. Samples were sequenced on Revio 25M SMRT cells (Pacific Biosciences, California, USA). The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

Hi-C

For Hi-C library preparation, the biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size-selected to 400–600 bp using SPRISelect beads. DNA was then enriched using Arima-HiC v2 Enrichment beads. The NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) was used for end repair, A-tailing, and adapter ligation, following a modified protocol in which library preparation is carried out while the DNA remains bound to the enrichment beads. PCR amplification was performed using KAPA HiFi HotStart mix and custom dual-indexed adapters (Integrated DNA Technologies) in a 96-well plate format. Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, samples were amplified for 10–16 PCR cycles. Post-PCR clean-up was carried out using SPRISelect beads. The libraries

were quantified using the Accuclear Ultra High Sensitivity dsDNA Standards Assay kit (Biotium) and normalised to 10 ng/μL before sequencing. Hi-C sequencing was performed on the Illumina NovaSeq 6000 instrument with 150 bp paired-end reads.

RNA

Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit, following the manufacturer's instructions. RNA sequencing was performed on the Illumina NovaSeq X instrument.

Genome assembly, curation and evaluation

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 first assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded using 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. Flat files and maps used in curation were generated via the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was conducted primarily in PretextView (Harry, 2022) and HiGlass (Kerpedjiev *et al.*, 2018), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were amended, and duplicate sequences were tagged and removed. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used to evaluate *k*-mer completeness and assembly quality for the primary and alternate haplotypes using the *k*-mer databases ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed in the blobtoolkit pipeline, a Nextflow (Di Tommaso *et al.*, 2017) port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools (Danecek *et al.*, 2021) and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoAT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using

DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these 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), relying on the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 4 contains a list of relevant software tool versions and sources.

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.4	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.7.1	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	666652151335353eef2fcd58880bcef5bc2928e1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoAT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhyllp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
MerquyFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQUERY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.04.2	https://github.com/nextflow-io/nextflow
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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 [here](#). 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: *Triturus cristatus* (warty newt). Accession number PRJEB75439; <https://identifiers.org/ena.embl/PRJEB75439>. The genome sequence is released openly for reuse. The *Triturus cristatus* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), the Sanger Institute Tree of Life Programme (PRJEB43745) and the

Vertebrate Genomes Project (PRJNA489243). 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](#).

Author information

Members of the Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.12159242>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.12158331>.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: <https://doi.org/10.5281/zenodo.12162482>.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: <https://doi.org/10.5281/zenodo.12165051>.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: <https://doi.org/10.5281/zenodo.12160324>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.12205391>.

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.4783558>.

Acknowledgements

We thank Natural England for the licence issued to S. Holt that enabled us to acquire the individual used for the genome sequencing. We are indebted to staff at the Gilbert White House and Gardens for permission to access the pond and collect the specimen. We are grateful to Inez Januszczak and Heather Allen for assistance with shipping the samples from the Natural History Museum to the Wellcome Sanger Institute. Genome size estimates from previous studies were obtained using Animal Genome Size Database (<https://www.genomesize.com/index.php>).

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](#)
- AmphibiaWeb: ***Triturus cristatus*: great crested newt**. 2021. [Reference Source](#)
- Arnold HR: **Atlas of amphibians and reptiles in Britain**. London: HMSO, 1995. [Reference Source](#)
- Baldwin L, Macgregor HC: **Centromeric satellite DNA in the newt *Triturus cristatus karelinii* and related species: its distribution and transcription on**

lampbrush chromosomes. *Chromosoma.* 1985; **92**(2): 100–107. [PubMed Abstract](#) | [Publisher Full Text](#)

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](#)

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor³ for LI PacBio**. *protocols.io.* 2023. [Publisher Full Text](#)

Brown T, Mishra K, Elewa A, *et al.*: **Chromosome-scale genome assembly reveals how repeat elements shape non-coding RNA landscapes active during newt limb regeneration**. *Cell Genom.* 2025; **5**(2): 100761. [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](#)

Buxton AS, Griffiths RA: **A revised system for interpreting great crested newt habitat suitability indices**. *Herpetol J*. 2022; **32**(3): 114–119.

[Publisher Full Text](#)

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved]**. *Wellcome Open Res*. 2023; **8**: 24.

[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](#)

Cunningham AA, Smith F, McKinley TJ, *et al.*: **Apparent absence of *Batrachochytrium salamandrivorans* in wild urodeles in the United Kingdom**. *Sci Rep*. 2019; **9**(1): 2831.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

da Veiga Leprevost F, Grüning BA, Alves Afritas S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization**. *Bioinformatics*. 2017; **33**(16): 2580–2582.

[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](#)

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash**. *protocols.io*. 2023a.

[Publisher Full Text](#)

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1**. *protocols.io*. 2023b.

[Publisher Full Text](#)

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows**. *Nat Biotechnol*. 2017; **35**(4): 316–319.

[PubMed Abstract](#) | [Publisher Full Text](#)

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation**. *Genome Biol*. 2023; **24**(1): 74.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana**. *protocols.io*. 2023.

[Publisher Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#)

Fahrback M, Gerlach U: **The genus *Triturus*: history, biology, systematics, captive breeding**. Frankfurt am Main, Germany: Edition Chimera, 2018.

[Reference Source](#)

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfstats: conversion, evaluation and manipulation of genome sequences using assembly graphs**. *Bioinformatics*. 2022; **38**(17): 4214–4216.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Grossen C, Neuenschwander S, Perrin N: **The balanced lethal system of crested newts: a ghost of sex chromosomes past?** *Am Nat*. 2012; **180**(6): E174–E183.

[PubMed Abstract](#) | [Publisher Full Text](#)

Griffiths RA, Sewell D, McCrea RS: **Dynamics of a declining amphibian metapopulation: survival, dispersal and the impact of climate**. *Biol Conserv*. 2010; **143**(2): 485–491.

[Publisher Full Text](#)

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences**. *Nat Methods*. 2018; **15**(7): 475–476.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies**. *Bioinformatics*. 2020; **36**(9): 2896–2898.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Harry E: **PretextView (Paired REad TETexture Viewer): a desktop application for viewing pretext contact maps**. 2022.

[Reference Source](#)

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](#)

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection**. *protocols.io*. 2023.

[Publisher 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](#)

Lawnczak MKN, Pereira-da-Conceicao LL, Blaxter ML, *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](#)

Liedtke HC, Gower DJ, Wilkinson M, *et al.*: **Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate**. *Nat Ecol Evol*. 2018; **2**(11): 1792–1799.

[PubMed Abstract](#) | [Publisher Full Text](#)

Litvinchuk SN, Rosanov JM, Borkin LJ: **Correlations of geographic distribution and temperature of embryonic development with the nuclear DNA content in the *Salamandridae* (Urodela, Amphibia)**. *Genome*. 2007; **50**(4): 333–342.

[PubMed Abstract](#) | [Publisher Full Text](#)

Macgregor HC, Horner H: **Heteromorphism for chromosome 1, a requirement for normal development in crested newts**. *Chromosoma*. 1980; **76**: 111–112.

[Publisher Full Text](#)

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes**. *Mol Biol Evol*. 2021; **38**(10): 4647–4654.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment**. *Linux J*. 2014; **2014**(239): 2. [Accessed 2 April 2024].

[Reference Source](#)

Nowoshilow S, Schloissnig S, Fei JF, *et al.*: **The axolotl genome and the evolution of key tissue formation regulators**. *Nature*. 2018; **559**(7712): E2.

[PubMed Abstract](#) | [Publisher Full Text](#)

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2**. *protocols.io*. 2023a.

[Publisher Full Text](#)

Oatley G, Sampaio F, Howard C: **Sanger Tree of Life fragmented DNA clean up: automated SPRI**. *protocols.io*. 2023b.

[Publisher Full Text](#)

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis**. 2023.

[Publisher Full Text](#)

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes**. *Nat Commun*. 2020; **11**(1): 1432.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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.*: **Merquary: reference-free quality, completeness, and phasing assessment for genome assemblies**. *Genome Biol*. 2020; **21**(1): 245.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Schwander T, Libbrecht R, Keller L: **Supergenes and complex phenotypes**. *Curr Biol*. 2014; **24**(7): R288–R294.

[PubMed Abstract](#) | [Publisher Full Text](#)

Uliano-Silva M, Ferreira JGRN, Krashenninnikova 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](#)

White G: **The natural history and antiquities of selborne**. London: B. White, 1789.

Wielstra B: **Balanced lethal systems**. *Curr Biol*. 2020; **30**(13): R742–R743.

[PubMed Abstract](#) | [Publisher Full Text](#)

Wiens JJ, Sparreboom M, Arntzen JW, *et al.*: **Crest evolution in newts: implications for reconstruction methods, sexual selection, phenotypic plasticity and the origin of novelties**. *J Evolution Biol*. 2011; **24**(10): 2073–2086.

[Publisher Full Text](#)

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool**. *Bioinformatics*. 2023; **39**(1): btac808.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 03 June 2025

<https://doi.org/10.21956/wellcomeopenres.26754.r124336>

© 2025 Crawford 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.



Andrew J Crawford 

Universidad de Los Andes Bogotá, Bogotá, Bogotá, Colombia

Summary:

This genome assembly report is remarkable in the level of detail and links to further resources, maximizing transparency and replicability of this work. The result is a massive genome assembled to remarkably high standards. The final assembly has a total length of 22,324.62 Mb in 1,303 scaffolds, with 4,698 gaps, and a scaffold N50 of 1248.77 Mb.

Minor comments:

Abstract:

Common name of amphibians are usually written with first letters in upper case (see AMNH Amph Spp World, AmphibiaWeb, and your keywords for examples).

"22,324.62 megabases" could be more clearly written as "22.3 Gb" here and throughout, but that's a personal opinion.

Background:

Paragraphs 1 and 3: see comment above on capitalizing common name.

Line 9: slight misspelling of "Batrachochytrium salamandrivorans".

Paragraph 3, line 1: I think the ';' should be a ':'.

Sentence 2: 'each variant' of how many? Just 2 variants or many are found in the population?

Paragraph 4: I actually prefer the ':' between higher taxonomic names, as you have here, although in Abstract & Taxonomy sections the authors used the ';' (as does NCBI).

I believe 'C-value' is more typically written with an uppercase 'C'. Perhaps also provide for the reader a conversion of pg to Gb?

There is also the 2019 *A. mexicanum* assembly by JJ Smith et al., and new *Ambystoma* assembly weighing in at 29 Gb https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_040938575.1/ 'Sanger Tree of life Project', capitalize the 'L'. Regarding 'Vertebrate Tree of Life Project', do you perhaps mean the "Vertebrate Genome Project"?

'Sequencing data' section:

'36 coverage', missing an 'X' or '-fold'?

Thanks for including RNAseq data. <https://www.ncbi.nlm.nih.gov/sra/ERR13493948> That's very helpful. The community looks forward to the subsequent annotation.

Results state "misjoins or missing joins" and the Methods say "missed joins and mis-joins". Perhaps homogenize the terminology?

Methods:

Line 1. Your specimen ID "NHMUK014561658" gets no hits at the Museum, <https://data.nhm.ac.uk/search>, but these 2 work: "NHMUK 014561658" or "NHMUK:014561658". Just to be kind, perhaps add the ':' in there. Ah, I see on Line 5 you have a different code "NHMUK 2022.7580" which also works. Just curious, why a 'specimen ID' and a museum voucher number?

While I absolutely support this statement, "Metadata collection for samples adhered to the Darwin Tree of Life project standards described by Lawniczak et al. (2022)," having this phrase comprising by itself a whole new paragraph is actually not clear to me. Does this mean the previous paragraph followed Lawniczak et al., or does this mean the metadata online follows these recommendations? Perhaps you could state what metadata provided and then cite this reference? Just a thought.

Denton et al. (2023b) now has a version 2 you could cite directly:

<https://www.protocols.io/view/sanger-tree-of-life-wet-laboratory-protocol-collec-8epv5xxy6g1b/v2>

PacBio HiFi

Add space in '55kb'.

Pacific Biosciences is cited 4 times, half the time with the location, half without. Most other citations to manufactures do not include their location. And some products had no citation to the manufacture. Perhaps be more consistent? 'analyses' plural.

"Wellcome Sanger Institute – Legal and Governance"

Line 4. Just curious, the whole animal went into -80°C, or it was ethanized first and the tissues went into -80°C? So, tissues were later taken out of a rock-solid frozen newt? (No liquid nitrogen

was involved?)

I am not seeing an ethics statement concerning Institutional Animal Care & Use approval. The section 'Wellcome Sanger Institute – Legal and Governance' points me to a document called 'Sampling Code of Practice' here <https://www.darwintreeoflife.org/project-resources/> which offers me a PDF here <https://www.darwintreeoflife.org/wp-content/uploads/2023/10/Template-DToL-Sampling-Code-Of-Practice.pdf> which has an Appendix that mentions "Sampling methods for relevant animals must be humane and consistent with current legal requirements and good practice." I admit I am unfamiliar with UK or English regulations around animal care and use, but in many countries one is expected to provide evidence of approval for euthanasia methods by an institutional committee, such as an IACUC in the USA. I would recommend providing the reader with some assurance that freezing a live newt in -80°C freezer is an approved method. Alternatively, if the animal was anesthetized first, one ought to state the method and provide evidence of committee approval and/or a current reference to veterinary recommendations for the chosen method of euthanasia.

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: Evolutionary biology and genomics of amphibians.

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 29 May 2025

<https://doi.org/10.21956/wellcomeopenres.26754.r124338>

© 2025 Pyron R. 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.



R Alexander Pyron 

The George Washington University Museum and Textile Museum, Washington, District of Columbia, USA

The Data Note by Streicher et al. describes one of only ~5 chromosome-scale salamander genomes, an important early achievement in urodelan genomics. I only have a few minor suggestions.

The section on the collection locality in “Background” – second paragraph – is very interesting and important natural-history and actual history. However, it feels a bit out of place, going from the newt in the first paragraph, to Gilbert White in the second, back to newts in the third. I would suggest moving it either to the Methods, or to the end of the Background.

In Genome sequence Report, Sequencing data, the sentence “Based on the estimated genome size, the sequencing data provided approximately 36 coverage” I believe should be “36x coverage.”

The authors do not seem to explain why they did not use the Hi-C data with the Hi-C mode of hifiasm to produce phased/haplotype-resolved assemblies for aTriCri1. Have they explored this?

The authors report sequencing RNA, but they don’t report anything about the quality or any attempted assemblies. It seems like even a basic de novo assembly in Trinity or similar, with BUSCO/compleasm scores, would be a good complement to the main genome. Also, was the sequencing strand-specific, for downstream analysis?

Perhaps this will be saved for a later manuscript or considered beyond the scope of a Data Note, but the authors do not really provide much context for how this genome compares to the other existing high-quality salamander assemblies such as Ambystoma mexicanum and Pleurodeles waltl.

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: Evolutionary genomics in salamanders

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.