DATA NOTE

The genome sequence of the Twenty-plume Moth, *Alucita hexadactyla* Linnaeus, 1758

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

Liam M. Crowley[1], Peter W. H. Holland[1], David C. Lees[2], Caroline Eve Mitchell[3],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Natural History Museum Genome Acquisiiton 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]University of Oxford, Oxford, England, UK
[2]Natural History Museum, London, England, UK
[3]Long Read Sequencing Team, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 13 Feb 2025, **10**:70
https://doi.org/10.12688/wellcomeopenres.23736.1
Latest published: 13 Feb 2025, **10**:70
https://doi.org/10.12688/wellcomeopenres.23736.1

Abstract

We present a genome assembly from a specimen of *Alucita hexadactyla* (Twenty-plume Moth; Arthropoda; Insecta; Lepidoptera; Alucitidae). The genome sequence has a total length of 878.53 megabases. Most of the assembly (99.74%) is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.32 kilobases in length.

Keywords

Alucita hexadactyla, Twenty-plume Moth, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ? ✓ ✓ ✓

	1	2	3	4
version 1 13 Feb 2025	? view	✓ view	✓ view	✓ view

1. **Jesper Boman**, Uppsala University, Uppsala, Sweden
2. **Daniel Berner**, University of Basel, Basel, Switzerland
3. **Annabel Whibley**, The University of Auckland, Auckland, New Zealand
Bragato Research Institute, Lincoln, New Zealand
4. **Toni De-Dios**, University of Tartu, Tartu, Estonia

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: Crowley LM: Investigation, Resources; **Holland PWH**: Investigation, Resources; **Lees DC**: Investigation, Resources; **Mitchell CE**: 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: Crowley LM, Holland PWH, Lees DC *et al.* **The genome sequence of the Twenty-plume Moth, *Alucita hexadactyla* Linnaeus, 1758 [version 1; peer review: 3 approved, 1 approved with reservations]** Wellcome Open Research 2025, **10**:70 https://doi.org/10.12688/wellcomeopenres.23736.1

First published: 13 Feb 2025, **10**:70 https://doi.org/10.12688/wellcomeopenres.23736.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Alucitoidea; Alucitidae; *Alucita*; *Alucita hexadactyla* Linnaeus, 1758 (NCBI:txid753150)

Background

Alucita hexadactyla is a common moth in the UK, belonging to the family Alucitidae and the group microlepidoptera (Davis, 2012). Commonly called the Many-plumed or Twenty-plume moth, the wings of *A. hexadactyla* have an unusual morphology. Unlike a more typical moth with four distinct scaled wings, each wing of *A. hexadactyla* is made from six feather-like fronds, hence *'hexadactyla'* meaning six-fingered in Greek. One study proposes that this may help the moth evade bats as the wing structure reduces the intensity of simulated ultrasonic bat calls (Kovalev, 2016). The wings have a white, brown and black chevron pattern and have a 14–16 mm wingspan (Butterfly Conservation, no date).

A. hexadactyla is the only member of the family in the UK and can be found widely across the UK and Ireland (Manley, 2021). The species is commonly found in gardens and woodlands where honeysuckle (*Lonicera caprifolium, periclymenum, xylosteum*) grows because the larvae feed on the buds, flowers, and leaves of the plant (Ellis, 2021). The moths fly throughout the year and are drawn to light so are often seen at windows or lights around buildings (Butterfly Conservation, no date; Wood, 1870).

The Alucitidae family is relatively small for a lepidopteran family, with roughly 180 species (Watkins, 2005), but is still expanding. Various new species in the family have been discovered across the world in the past 20 years (Byun, 2006; Landry & Landry, 2004; Ustjuzhanin & Kovtunovich, 2016), with a notable recent burst of discovery in Cameroon (Ustjuzhanin *et al.*, 2018; Ustjuzhanin *et al.*, 2020). With recent discoveries of new species, and the potential for more, reference genomes for the family will be key to understanding the evolutionary history and phylogeny of Alucitidae and, more widely, lepidoptera. Here we present a chromosomally complete genome sequence for *Alucita hexadactyla*, based on a male specimen from Olton, West Midlands, United Kingdom (Figure 1). This reference genome for *Alucita hexadactyla* is the first of the family and of the Alucitoidea superfamily.

Genome sequence report

Sequencing data

The genome of a specimen of *Alucita hexadactyla* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 30.72 Gb from 2.62 million reads. GenomeScope analysis of the PacBio HiFi data estimated the haploid genome size at 871.06 Mb, with a heterozygosity of 0.79% and repeat content of 44.71%. These values provide an initial assessment of genome complexity and the challenges anticipated during assembly. Based on this estimated genome size, the sequencing data provided approximately 34.0x coverage of the genome. Chromosome conformation Hi-C data produced 137.24 Gb from 908.89 million reads. Table 1 summarises the specimen and sequencing information, including the BioProject, study name, BioSample numbers, and sequencing data for each technology.

Figure 1. Photograph of the *Alucita hexadactyla* (ilAluHexa2) specimen used for genome sequencing.

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 14 misjoins or missing joins and removed three haplotypic duplications. The final assembly has a total length of 878.53 Mb in 58 scaffolds, with 99 gaps, and a scaffold N50 of 31.03 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 (99.75%) was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3). During curation, chromosome Z was identified based on the Hi-C signal.

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 in GenBank.

Table 1. **Specimen and sequencing data for *Alucita hexadactyla*.**

Project information			
Study title	Alucita hexadactyla (twenty-plume moth)		
Umbrella BioProject	PRJEB73441		
Species	*Alucita hexadactyla*		
BioSpecimen	SAMEA112232562		
NCBI taxonomy ID	753150		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	ilAluHexa2	SAMEA112233010	whole organism
Hi-C sequencing	ilAluHexa3	SAMEA112233015	whole organism
RNA sequencing	ilAluHexa4	SAMEA114806037	whole organism
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq X	ERR12743807	9.09e+08	137.24
PacBio Sequel IIe	ERR12721083	2.62e+06	30.72
RNA Illumina NovaSeq X	ERR13493920	8.48e+07	12.81

Table 2. **Genome assembly data for *Alucita hexadactyla*.**

Genome assembly		
Assembly name	ilAluHexa2.1	
Assembly accession	GCA_964059475.1	
Alternate haplotype accession	*GCA_964059515.1*	
Assembly level for primary assembly	chromosome	
Span (Mb)	878.53	
Number of contigs	157	
Number of scaffolds	58	
Longest scaffold (Mb)	57.39	
Assembly metrics	**Measure**	***Benchmark***
Contig N50 length	11.55 Mb	*≥ 1 Mb*
Scaffold N50 length	31.03 Mb	*= chromosome N50*
Consensus quality (QV)	Primary: 63.5; alternate: 62.6; combined 63.0	*≥ 40*
k-mer completeness	Primary: 81.84%; alternate: 79.84%; combined: 99.02%	*≥ 95%*
BUSCO*	C:98.1%[S:97.4%,D:0.7%], F:0.3%,M:1.6%,n:5,286	*S > 90%, D < 5%*
Percentage of assembly mapped to chromosomes	99.75%	*≥ 90%*
Sex chromosomes	Z	*localised homologous pairs*
Organelles	Mitochondrial genome: 15.32 kb	*complete single alleles*

* BUSCO scores based on the lepidoptera_odb10 BUSCO set using version 5.5.0. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

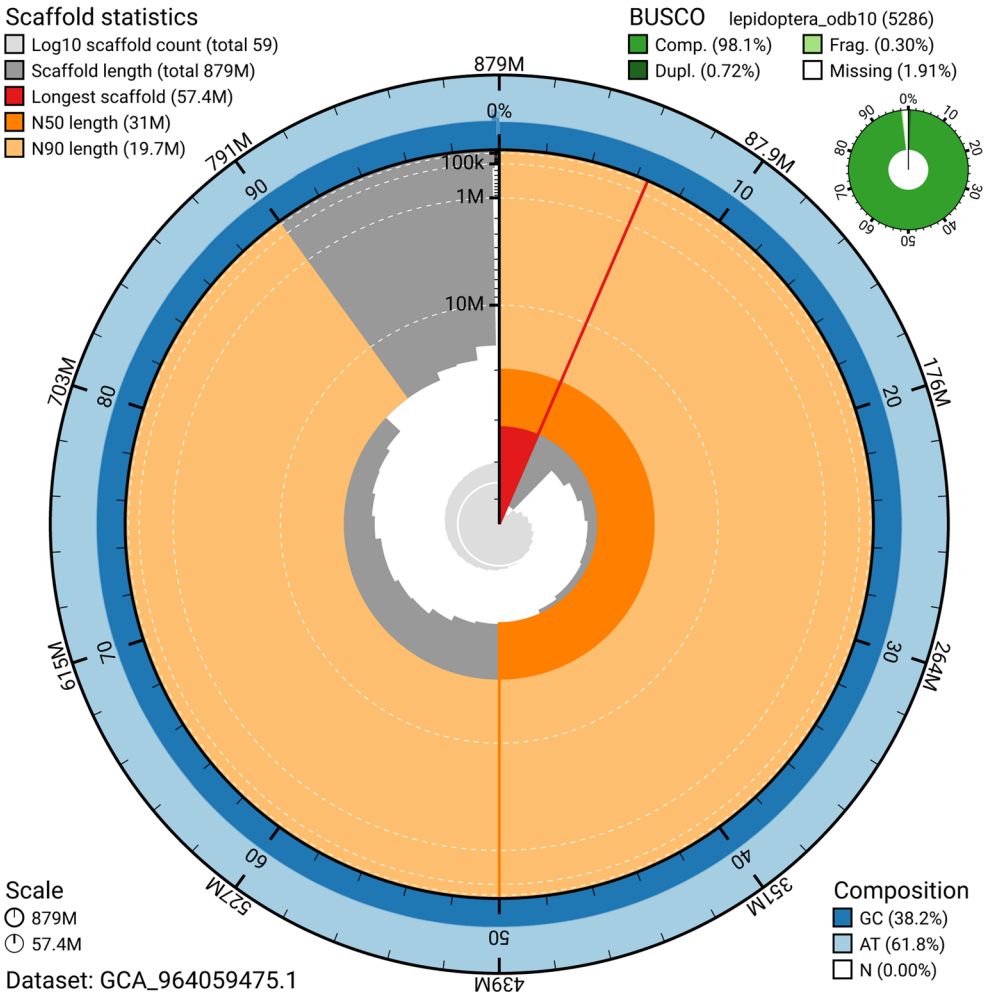

Figure 2. Genome assembly of *Alucita hexadactyla*, : metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. 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 lepidoptera_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964059475.1/dataset/GCA_964059475.1/snail.

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 primary haplotype has a QV of 63.5, and the combined primary and alternate assemblies achieve an estimated QV of 63.0. The *k*-mer completeness for the primary haplotype is 81.84%, and for the alternate haplotype it is 79.84%. The combined primary and alternate assemblies achieve a *k*-mer completeness of 99.02%. BUSCO analysis using the lepidoptera_odb10 reference set (n = 5,286) indicated a completeness score of 98.1% (single = 97.4%, duplicated = 0.7%).

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.63.

Methods

Sample acquisition and DNA barcoding

An adult male *Alucita hexadactyla* (specimen ID Ox002335, ToLID ilAluHexa2) was collected from Olton, West Midlands,

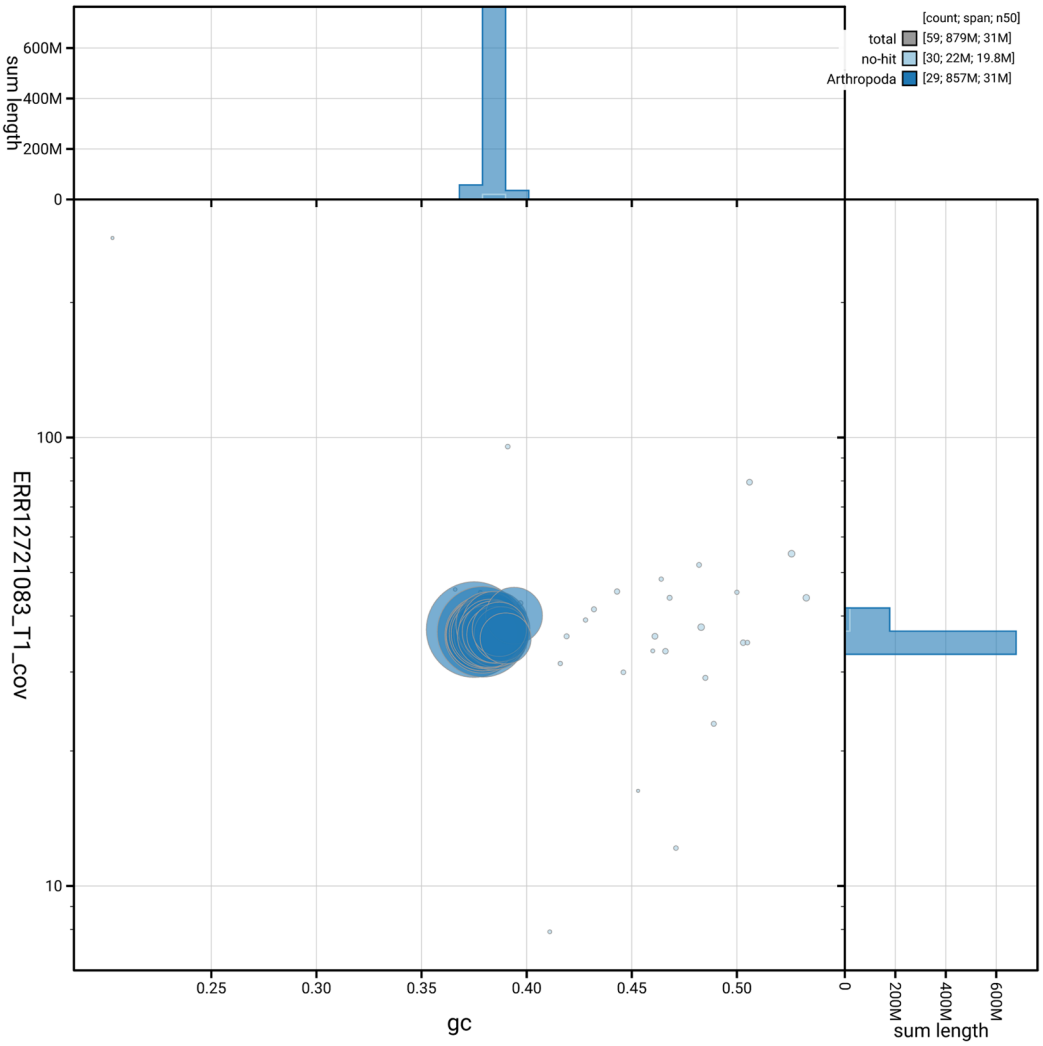

Figure 3. Genome assembly of *Alucita hexadactyla*, : BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms 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_964059475.1/blob.

United Kingdom (latitude 52.44, longitude –1.81) on 2022-07-16 by potting. The specimen was collected and identified by Liam Crowley (University of Oxford) and preserved on dry ice. This specimen was used for genome sequencing.

The specimen used for Hi-C sequencing (specimen ID Ox002338, ToLID ilAluHexa3) was a adult specimen collected from Wallingford, Oxfordshire, United Kingdom (latitude 51.6, longitude –1.14) on 2022-07-11, using a light trap. The specimen was collected and identified by Peter Holland (University of Oxford) and preserved on dry ice.

The specimen used for RNA sequencing (specimen ID NHMUK014584894, ToLID ilAluHexa4) was collected from Lucas Road, High Wycombe, England, United Kingdom (latitude 51.63, longitude –0.74) on 2022-07-08. The specimen was collected and identified by David Lees (Natural History Museum) and preserved by dry freezing (–80 °C).

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) (Pereira *et al.*, 2022). 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

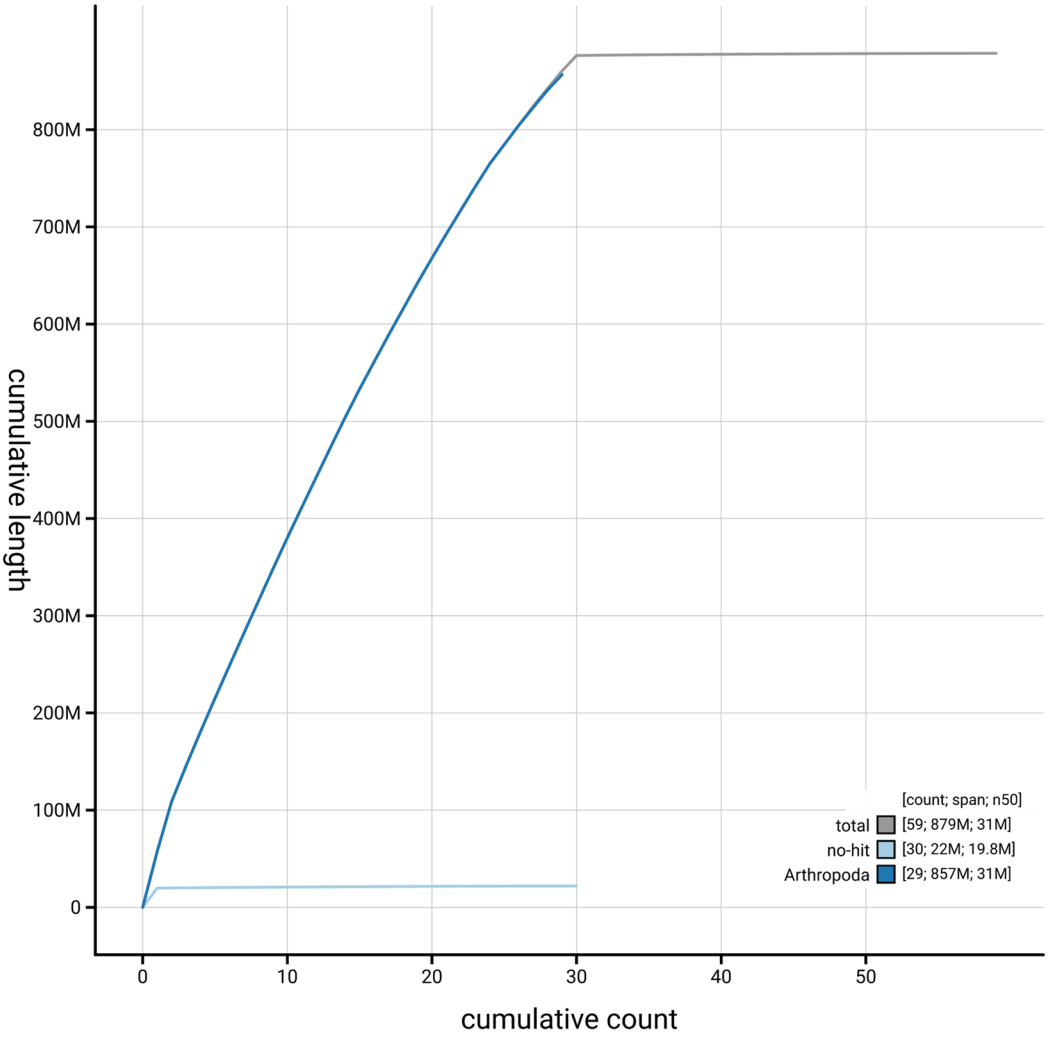

Figure 4. Genome assembly of *Alucita hexadactyla* : BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964059475.1/dataset/GCA_964059475.1/cumulative.

the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding have been deposited on protocols.io (Beasley *et al.*, 2023).

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 ilAluHexa2 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). The 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 (Strickland *et al.*, 2023). 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 whole organism tissue of ilAluHexa4 in the Tree of Life Laboratory at the WSI using the RNA

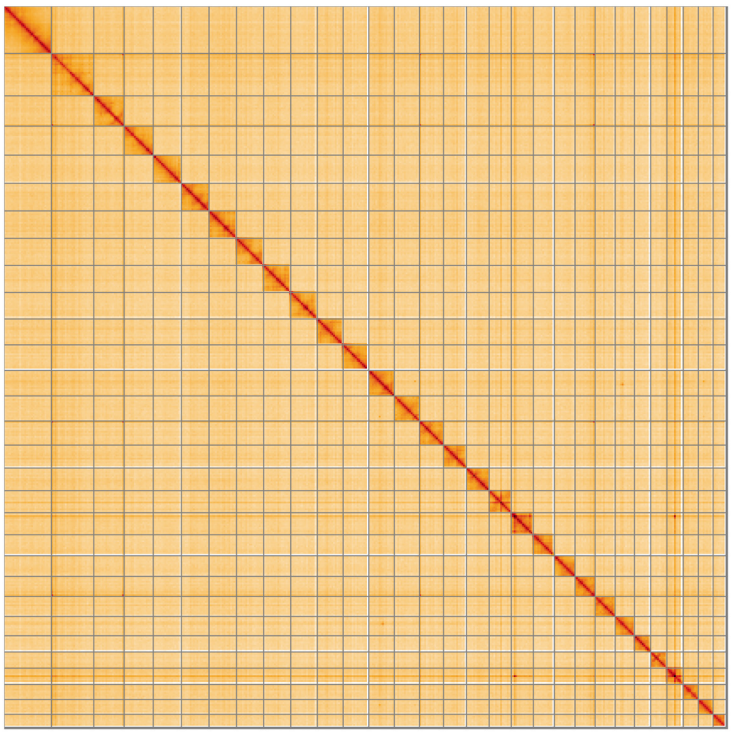

Figure 5. Genome assembly of *Alucita hexadactyla* : Hi-C contact map of the assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=IJtVqGMSTHuQ0wxO9542LA.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Alucita hexadactyla*, ilAluHexa2.

INSDC accession	Name	Length (Mb)	GC%
OZ060715.1	1	51.42	38
OZ060716.1	2	36.8	38
OZ060717.1	3	35.23	38
OZ060718.1	4	34.24	38
OZ060719.1	5	33.58	38
OZ060720.1	6	33.37	38
OZ060721.1	7	32.95	38
OZ060722.1	8	32.92	38
OZ060723.1	9	32.24	38
OZ060724.1	10	31.39	38
OZ060725.1	11	31.03	38
OZ060726.1	12	30.97	38.5
OZ060727.1	13	30.49	38
OZ060728.1	14	29.24	38.5
OZ060729.1	15	27.9	38.5
OZ060730.1	16	27.57	38
OZ060731.1	17	26.77	38
OZ060732.1	18	26.76	38.5
OZ060733.1	19	25.64	38.5
OZ060734.1	20	25.03	38.5
OZ060735.1	21	24.42	38.5
OZ060736.1	22	24.2	38
OZ060737.1	23	23.43	38.5
OZ060738.1	24	19.75	39
OZ060739.1	25	19.73	38.5
OZ060740.1	26	19.72	39.5
OZ060741.1	27	18.3	39
OZ060742.1	28	18.04	38.5
OZ060743.1	29	15.8	39
OZ060714.1	Z	57.39	37.5
OZ060744.1	MT	0.02	20.5

Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Hi-C sample preparation

Tissue from the whole organism of the ilAluHexa3 sample was processed for Hi-C sequencing at the WSI Scientific Operations core, using the Arima-HiC v2 kit. In brief, 20–50 mg of frozen tissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde concentration. After crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the Arima-HiC v2 kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation. Additionally, the biotinylation percentage was estimated using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit and 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, California, USA), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA) 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. Following the manufacturer's instructions, size selection and clean up was carried out using diluted AMPure PB beads (Pacific Biosciences, California, USA). DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher 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 gDNA 55kb BAC analysis kit.

Samples were sequenced using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

Hi-C

For Hi-C library preparation, DNA was fragmented using the Covaris E220 sonicator (Covaris) and size selected using SPRISelect beads to 400 to 600 bp. The DNA was then enriched using the Arima-HiC v2 kit Enrichment beads. Using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) for end repair, a-tailing, and adapter ligation. This uses a custom protocol which resembles the standard NEBNext Ultra II DNA Library Prep protocol but where library preparation occurs while DNA is bound to the Enrichment beads. For library amplification, 10 to 16 PCR cycles were required, determined by the sample biotinylation percentage. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq X instrument.

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 were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) 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 (article in preparation). Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were

corrected, and duplicate sequences were tagged and removed. Sex chromosomes were identified by Hi-C coverage analysis. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation (article in preparation).

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) that were computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

A Hi-C contact map was produced for the final version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools 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.

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'**,

Table 4. Software tools: versions and sources.

Software tool	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	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	https://github.com/thegenemyers/FASTK
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/chhylp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84a a44357826c0b6753eb28de	https://github.com/higlass/higlass
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi

Software tool	Version	Source
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextView	0.2	https://github.com/sanger-tol/PretextView
purge_dups	None	https://github.com/dfguan/purge_dups
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.5.1	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

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: Alucita hexadactyla (twenty-plume moth). Accession number PRJEB73441; https://identifiers.org/ena.embl/PRJEB73441. The genome sequence is released openly for reuse. The *Alucita hexadactyla* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. 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.

References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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 Megaruptor®3 for LI PacBio.** *protocols.io.* 2023.
Publisher Full Text

Beasley J, Uhl R, Forrest LL, *et al.*: **DNA barcoding SOPs for the Darwin Tree of Life project.** *protocols.io.* 2023; [Accessed 25 June 2024].
Publisher 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

Butterfly Conservation: **Twenty-plume moth**. [no date].
Reference Source

Byun BK: **Alucitidae (Lepidoptera) of Korea: description of a new species and records of two previously unrecorded species.** *Zootaxa.* 2006; **1188**(1): 37–47.
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

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

da Veiga Leprevost F, Grüning BA, Alves Aflitos 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

Davis T: **A review of the status of microlepidoptera in Britain**. 2012.
Reference Source

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

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

Ellis WN: **Parasites**. 2021.
Reference Source

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

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 TEXTure 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

Kovalev IS: **Stealth moths: the multi-plumed wings of the moth *Alucita hexadactyla* may decrease the intensity of their echo to simulated bat echolocation cries.** *Entomol News.* 2016; **126**(3): 204–212.
Publisher 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

Landry B, Landry JF: **The genus *Alucita* in North America, with description of two new species (Lepidoptera: Alucitidae).** *Can Entomol.* 2004; **136**(4): 553–579.
Publisher 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

Manley C: **Alucitidae and Pterophoridae, British and Irish Moths: Third Edition: a Photographic Guide.** 3rd ed. Bloomsbury Publishing, 2021.

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

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

Pereira L, Sivell O, Sivess L, *et al.*: **DToL Taxon-specific Standard Operating Procedure for the terrestrial and freshwater arthropods working group**. 2022.
Publisher Full Text

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

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
PubMed Abstract | Publisher Full Text | Free 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.*: **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

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023.
Publisher 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**

Ustjuzhanin P, Kovtunovich V, Maicher V, *et al.*: **Even hotter hotspot: description of seven new species of many-plumed moths (Lepidoptera, Alucitidae) from Mount Cameroon.** *Zookeys.* 2020; **935**: 103–119.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ustjuzhanin P, Kovtunovich V, Sáfián S, *et al.*: **A newly discovered biodiversity hotspot of many-plumed moths in the Mount Cameroon area: first report on species diversity, with description of nine new species (Lepidoptera, Alucitidae).** *Zookeys.* 2018; 777: 119–139.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ustjuzhanin PY, Kovtunovich VN: **Two new species of the many-plumed moth genus *Alucita* Linnaeus (Lepidoptera, Alucitidae) from Madagascar.** *Entomol Rev.* 2016; **96**(8): 1115–1118.
Publisher 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

Watkins RA: **Pterophoroidea & Alucitoidea (Lepidoptera).** *Ann Entomol Soc Am.* 2005; **98**(5): 754.
Publisher Full Text

Wood JG: **Chapter 5. The common moths of England**. G. Routledge, 1870.
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 1

Reviewer Report 13 March 2025

https://doi.org/10.21956/wellcomeopenres.26184.r118753

✓ **Toni De-Dios**
Institute of Genomics, University of Tartu, Tartu, Estonia

In the present article, Crowley and colleagues report a new genome assembly for *Alucita hexadactyla*, a common moth that represents the only member of the Alucitidae family in the UK. The resultant assembled genome is of high quality, with and average depth of 34X, with 99.75% of the sequences contained in the 30 assembled chromosomes (autosomes and sexual chromosomes), a quality value of 63.5%, combined *k-mer* completeness of 99.02%, and BUSCO of 98.1%. The protocols used to achieve these results are state-of-the-art, and they are explained in the detail in the methods sections. All software used is properly detailed and cited. The assembled genome is publicly available at ENA, and the raw data at INSDC.

The results of this piece of work are clear and solid, with this high-quality assembly setting a new precedent for the study of the population genomics of the moth *Alucita hexadactyla*.

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: Ancient DNA, population genetics, metagenomics, phylogenetics.

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 04 March 2025

https://doi.org/10.21956/wellcomeopenres.26184.r118751

Annabel Whibley
[1] The University of Auckland, Auckland, New Zealand
[2] Grapevine Improvement, Bragato Research Institute, Lincoln, Canterbury, New Zealand

Crowley and colleagues report the genome assembly of the twenty plume moth, *Alucita hexadactyla* , a reference that serves as an important "first" for the Alucitoidea superfamily. The analysis follows standardised Darwin Tree of Life sampling, verification, sequencing and reporting protocols. As usual, these are comprehensive, clearly structured and using current best-practice tools. That said, this is the first data note that I have reviewed that decomposes kmer completeness into primary and alternative haplotypes, and more information on methods has been incorporated, such as the reporting of kmer lengths used for analyses, and I welcome these changes to the template. A very high quality reference is reported, starting with >100x HiFi data and Illumina HiC data. RNAseq data was obtained but an annotation has not been reported. Public data accession links are active.

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

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 01 March 2025

https://doi.org/10.21956/wellcomeopenres.26184.r118756

Daniel Berner
University of Basel, Basel, Switzerland

This paper describes the assembly and annotation of a reference genome in chromosome quality for the Twenty-plume Moth *Alucita hexadactyla*. Based on samples from the UK, the work involves PacBio HiFi long-read sequencing and Hi-C link information to assemble 30 chromosomes, including the Z (the assembly is based on a male specimen lacking the W chromosome). The overall strategy follows the Darwin Tree of Life pipeline. Based on different completeness and contiguity checks, the genome can be considered of top quality. In addition, whole-organism RNA sequencing is performed for a de novo annotation, although this annotation is not included in this paper.
The samples, wet laboratory procedures and analytical pipelines are comprehensively described, and accession numbers for the output are provided, allowing the community to use this genome as a resource. I do not have any comments to the authors apart from my recommendation to add the annotation to this paper if possible, to make this genomic resource more complete. Since the RNA seq data are already generated, I feel this would not be too much extra work.

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 ecology; Population genomics; Speciation

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 19 February 2025

https://doi.org/10.21956/wellcomeopenres.26184.r118757

?

Jesper Boman
Uppsala University, Uppsala, Sweden

Crowley et al. presents a high-quality genome assembly of the twenty-plume moth (*Alucita hexadactyla*). As the authors state, this assembly is useful for further evolutionary analysis of Alucitidae. Background is overall well-written. See below for some comments.

1. First, I appreciate the thoroughness in barcoding the samples to make sure of species status.

2. In the Background section you write "*Lonicera caprifolium, periclymenum, xylosteum*". Based on my knowledge it is more common to write: *Lonicera caprifolium, L. periclymenum, L. xylosteum.*

3 You write: "The Alucitidae family is relatively small for a lepidopteran family, with roughly 180 species (Watkins, 2005), but is still expanding." It would be more appropriate to write that more species in Alucitidae are being described rather than that it is expanding, which gives connotations of a biological process such as a diversfication rate larger than zero. The latter might be true but I think it is the former that the authors are trying to say.

I also want to reiterate point 1), 2) and 4) from a previous peer review:

Boman J. Peer Review Report For: The genome sequence of the Bright-line Brown-eye moth, Lacanobia oleracea Linnaeus, 1758 [version 1; peer review: 1 approved with reservations]. Wellcome Open Res 2024, 9:515 (https://doi.org/10.21956/wellcomeopenres.25228.r98875)

They are valid for this report as well and addressing them would improve clarity and reproducibility.

Best regards,
Jesper

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

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

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?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Speciation genomics, Population genomics, Butterflies, Gene regulation, Transposons, 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, however I have significant reservations, as outlined above.