Check for updates

DATA NOTE

The genome sequence of the Kent black arches, *Meganola albula* (Denis & Schiffermüller, 1775) [version 1; peer review: 2 approved, 1 approved with reservations]

Douglas Boyes[1+], Ian Sims[2], David Lees [3],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Natural History Museum Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Alexander Dove[4],
Darwin Tree of Life Consortium

[1]UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK
[2]British Entomological and Natural History Society, Reading, Berkshire, UK
[3]Natural History Museum, London, UK
[4]Wellcome Sanger Institute, Hinxton, UK

[+] Deceased author

Abstract

We present a genome assembly from an individual male *Meganola albula* (the Kent black arches; Arthropoda; Insecta; Lepidoptera; Nolidae). The genome sequence is 405 megabases in span. Most of the assembly (99.95%) is scaffolded into 30 chromosomal pseudomolecules with the Z sex chromosome assembled. The mitochondrial genome has also been assembled and is 15.4 kilobases in length.

Keywords

Meganola albula, Kent Black Arches, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ?

	1	2	3
version 1 22 Dec 2022	✓ view	✓ view	? view

1. **Raghuvir Viswanatha**, Blavatnik Institute - Harvard Medical School, Boston, USA

2. **Tobias Baril** , University of Neuchâtel, Neuchâtel, Switzerland

3. **Charles Robin**, The University of Melbourne, Melbourne, Australia

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: Boyes D: Investigation, Resources; **Sims I**: Investigation, Resources; **Lees D**: Investigation, Resources; **Dove A**: Writing – Original Draft Preparation;

How to cite this article: Boyes D, Sims I, Lees D *et al.* **The genome sequence of the Kent black arches, *Meganola albula* (Denis & Schiffermuller, 1775) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2022, **7**:310 https://doi.org/10.12688/wellcomeopenres.18672.1

First published: 22 Dec 2022, **7**:310 https://doi.org/10.12688/wellcomeopenres.18672.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Nolidae; Nolinae; *Meganola*; *Meganola albula* (Denis & Schiffermüller, 1775) (NCBI:txid987977).

Background

Kent Black Arches, or *Meganola albula*, belonging to the moth family *Nolidae* (Zahiri *et al.*, 2013) and one of the 80 species in the genus *Meganola* (Cha *et al.*, 2019), was first described in 1775 by Denis & Schiffermüller in Austria. The species bears close similarity to Scarce Black Arches (*Nola aerugula*), another species found in similar coastal UK habitats as Kent Black Arches (NBN Atlas; Õunap *et al.*, 2021). *M. albula* wings have a white ground colour with a curved brown medial band, white curved sub-terminal line and brown terminal field (Perveen & Khan, 2017) . The *M. albula* wingspan is 16–20 mm (Cha *et al.*, 2019; Kiss, 2017; Zahiri *et al.*, 2012). The head and thorax are white, and the abdomen is brown (Cha *et al.*, 2019).

This species is widely distributed across the Palearctic region (Tshistjakov, 2008); most observations have been reported across Western Europe, seen as far north as the southern coast of Finland. *M. albula* is sparsely distributed across the UK, primarily found in coastal regions. Early sightings within the UK were along the southern coast of England (Heath & Alford, 1973), primarily within counties of Kent and Sussex, but the species is more frequently seen further north along the south-east coastline (NBN Atlas). *M. albula* occupies a variety of coastal habitats, including coastal marshes (Zilli, 2014) and woodlands, with the larva feeding on *Rubus caesius* (Seymour, 2018), *Fragaria* (strawberry), *Potentilla*, *Vaccinium*, *Lotus*, *Trifolium* and *Mentha* (Tshistjakov, 2008).

The adult *M. albula* is most commonly seen within the UK and western Europe between June and August, with most sightings occurring during the month of July (NBN Atlas), flying in a single generation. In eastern regions (far eastern Russian, eastern China) multiple flight generations are seen, from towards the end of May till June and then from July into early August (Tshistjakov, 2008). This species overwinters as small larvae (Skinner & Wilson, 2009).

Genome sequence report

The genome was sequenced from one male *M. albula* specimen (Figure 1) collected from Wytham Woods, Berkshire, UK (latitude 51.769, longitude –1.33). A total of 43-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 108-fold coverage in 10X Genomics read clouds was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 10 missing or misjoins, reducing the scaffold number by 15%.

The final assembly has a total length of 405.2 Mb in 34 sequence scaffolds with a scaffold N50 of 14.3 Mb (Table 1).

Figure 1. The *Meganola albula* (ilMegAlbu1) specimen used for genome sequencing.

Most (99.95%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome (Figure 2–Figure 5; Table 2). Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 98.4% (single 98.1%, duplicated 0.4%) using the lepidoptera_odb10 reference set.

Methods

Sample acquisition and nucleic acid extraction

Two *M. albula* specimens (ilMegAlbu1 and ilMegAlbu2) of undetermined sex were collected using a light trap in Wytham Woods, Berkshire, UK (latitude 51.769, longitude –1.33) by Douglas Boyes (University of Oxford), who also identified the species. The specimens were identified by Douglas Boyes and snap-frozen on dry ice. A second specimen (ilMegAlbu4) of undetermined sex was collected using an aerial net from Hartslock Nature Reserve (latitude 51.51, longitude –1.11) by Ian Sims (British Entomological and Natural History Society). The specimen was identified by Ian Sims and David Lees (Natural History Museum, London) and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The ilMegAlbu1 sample was weighed and dissected on dry ice. Whole organism tissue was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 20 ng aliquot of extracted DNA using 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3

Table 1. Genome data for *Meganola albula*, ilMegAlbu1.1.

Project accession data	
Assembly identifier	ilMegAlbu1.1
Species	Meganola albula
Specimen	ilMegAlbu1
NCBI taxonomy ID	987977
BioProject	PRJEB46332
BioSample ID	SAMEA7701294
Isolate information	ilMegAlbu1: male, whole organism (genome assembly); ilMegAlbu2: unknown sex, (Hi-C); ilMegAlbu4: head/thorax (RNA-Seq)
Assembly metrics*	
Consensus quality (QV)	60.6 (Benchmark: ≥50)
k-mer completeness	100% (Benchmark: ≥95%)
BUSCO**	C:98.4%[S:98.1%,D:0.4%],F:0.5%,M:1.1%,n:5,286 (Benchmark: C ≥ 95%)
Percentage of assembly mapped to chromosomes	99.95% (Benchmark: ≥ 95%)
Sex chromosomes	ZZ (Benchmark: localised homologous pairs)
Organelles	Mitochondrial genome 15.4 kb (Benchmark: complete single alleles)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6808014
10X Genomics Illumina	ERR6688590–ERR6688593
Hi-C Illumina	ERR6688594
PolyA RNA-Seq Illumina	ERR10123654
Genome assembly	
Assembly accession	GCA_936450015.1
Accession of alternate haplotype	GCA_936440415.1
Span (Mb)	405.2
Number of contigs	47
Contig N50 length (Mb)	13.3
Number of scaffolds	34
Scaffold N50 length (Mb)	14.3
Longest scaffold (Mb)	25.7

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from (Rhie *et al.,* 2021).

** BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in compa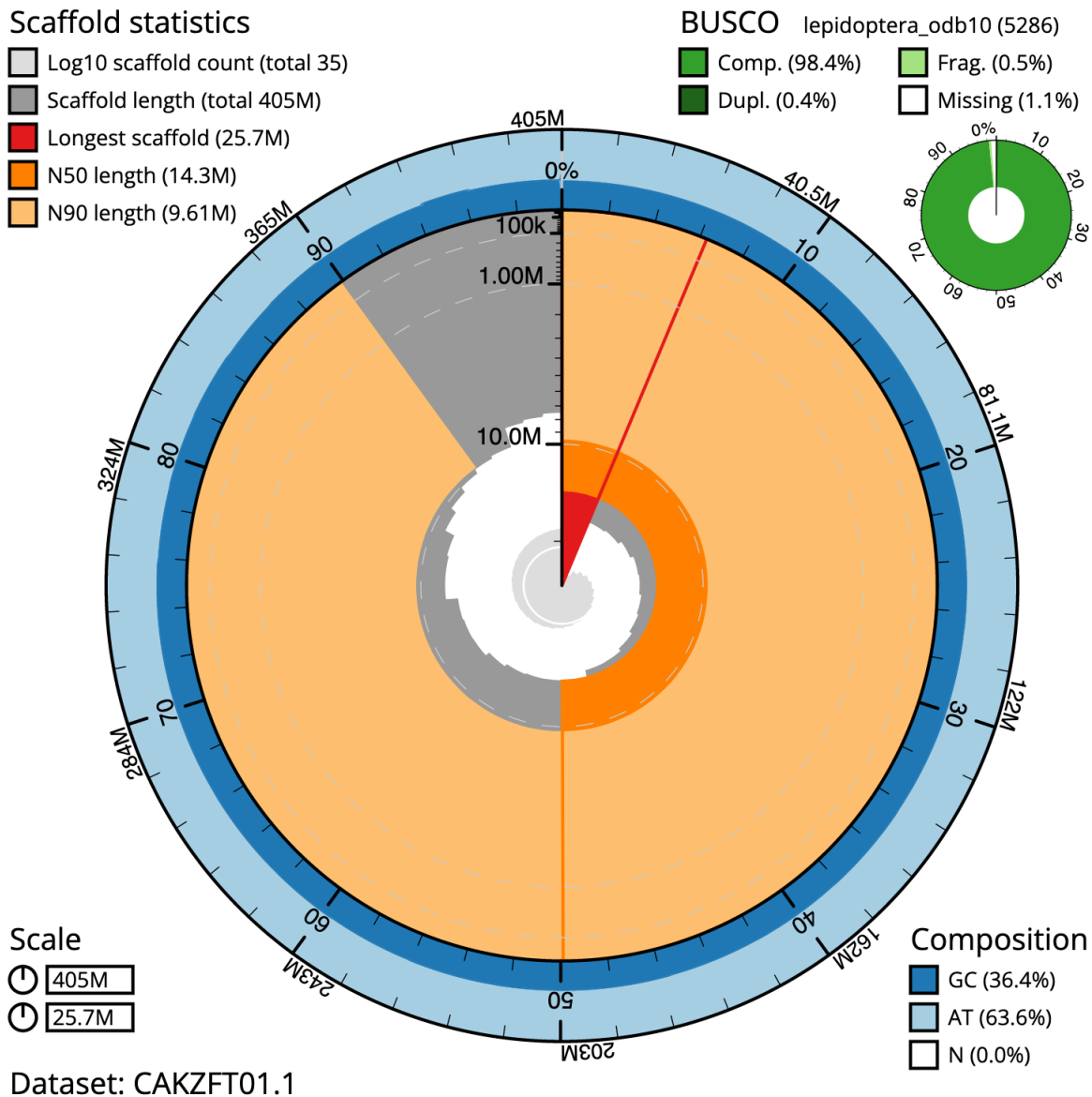rison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/ilMegAlbu1.1/dataset/CAKZFT01.1/busco.

Figure 2. Genome assembly of *Meganola albula*, ilMegAlbu1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 405,261,716 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (25,687,960 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (14,254,146 and 9,608,783 bp), respectively. The pale grey spiral shows the cumulative chromosome count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the lepidoptera_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilMegAlbu1.1/dataset/CAKZFT01.1/snail.

system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from head/thorax tissue of ilMegAlbu4 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 µl RNAse-free water and its concentration was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.
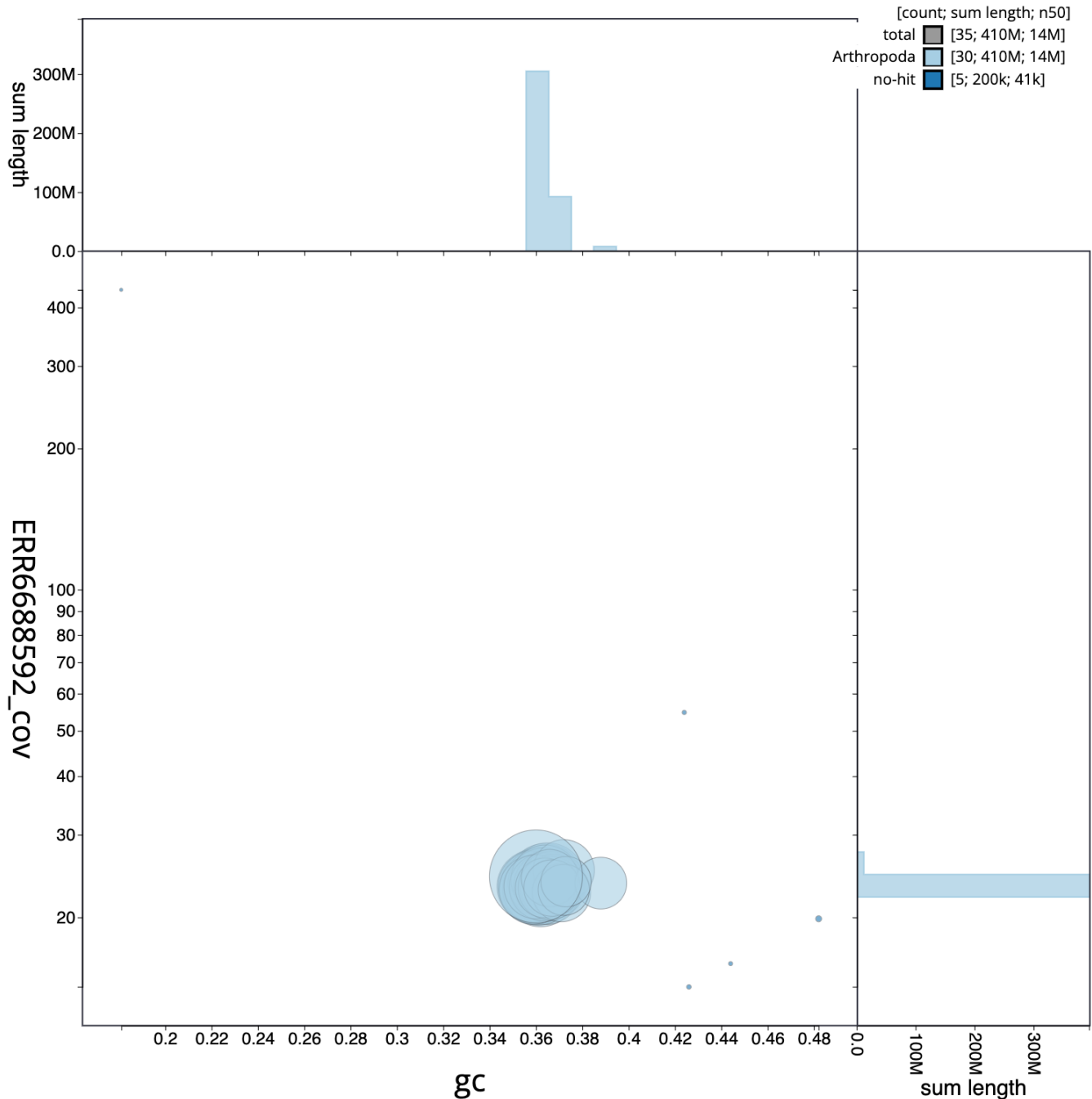
Figure 3. Genome assembly of *Meganola albula*, ilMegAlbu1.1: GC coverage. BlobToolKit GC-coverage plot. Chromosomes are coloured by phylum. Circles are sized in proportion to chromosome length. Histograms show the distribution of chromosome length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilMegAlbu1.1/dataset/CAKZFT01.1/blob.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq and

10X) instruments. Hi-C data were also generated from whole organism tissue of ilMegAlbu2 using the Arima v2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing
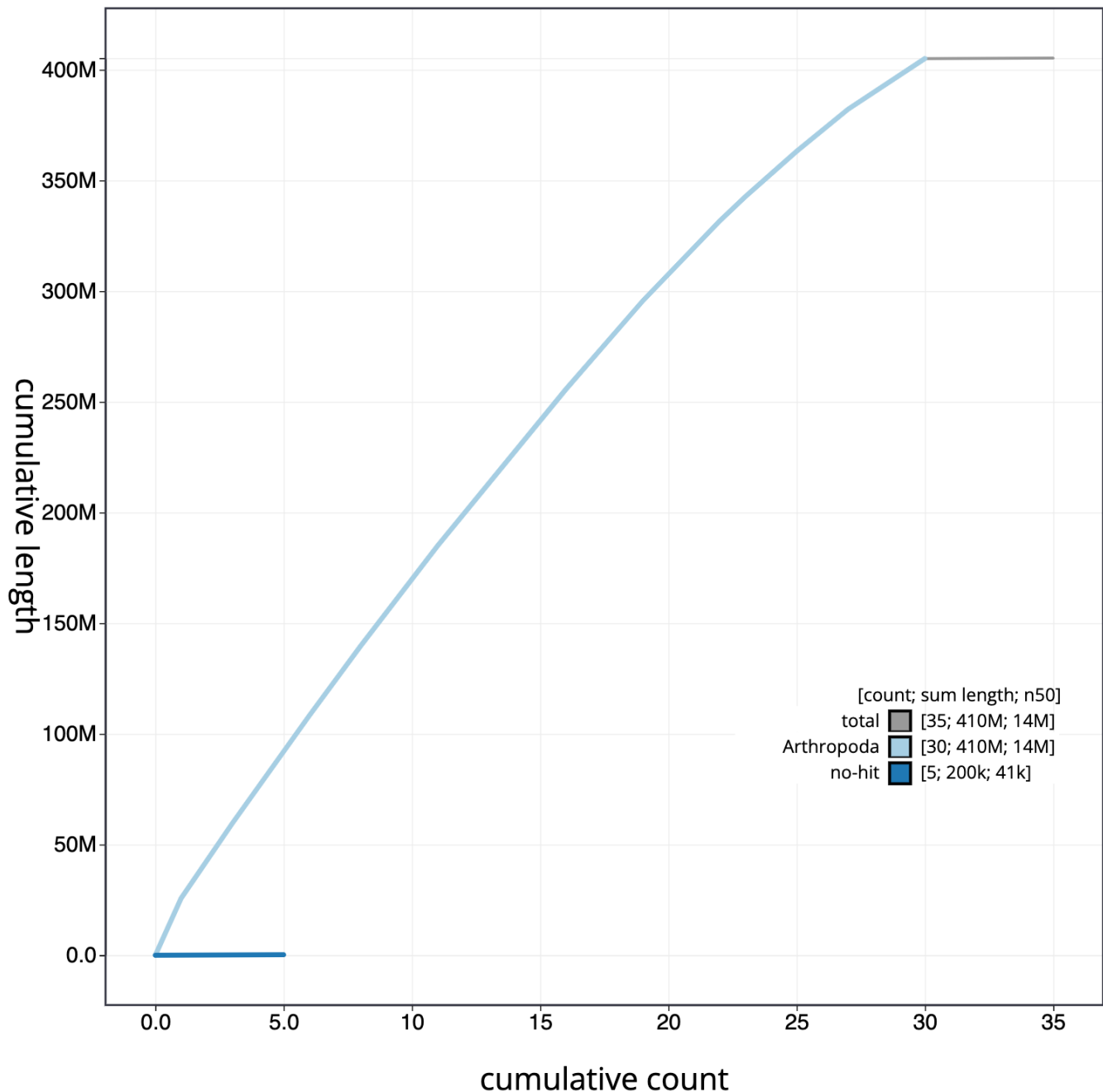
Figure 4. Genome assembly of *Meganola albula*, ilMegAlbu1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all chromosomes. Coloured lines show cumulative lengths of chromosomes assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilMegAlbu1.1/dataset/CAKZFT01.1/cumulative.

was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018)

and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), which performed annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.
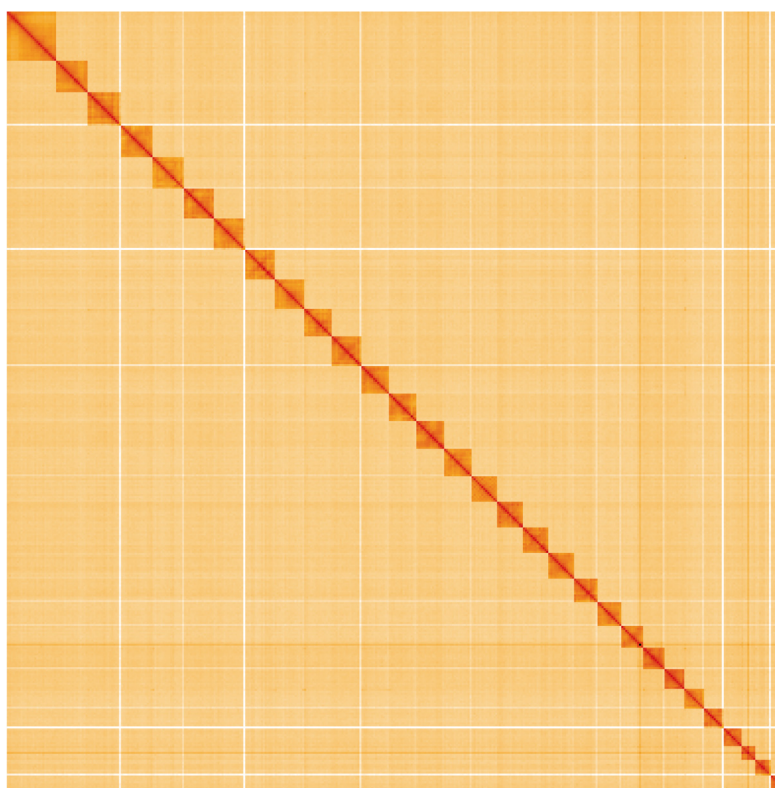
Figure 5. Genome assembly of *Meganola albula*, ilMegAlbu1.1: Hi-C contact map. Hi-C contact map of the ilMegAlbu1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=G0f3dXHkQLiKTZCRbaZhOQ.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Meganola albula*, ilMegAlbu1.

INSDC accession	Chromosome	Size (Mb)	GC%
OW388143.1	1	17.05	36.1
OW388144.1	2	16.89	36.4
OW388145.1	3	16.48	36.5
OW388146.1	4	16.13	36.2
OW388147.1	5	16.03	36.2
OW388148.1	6	15.8	35.9
OW388149.1	7	15.62	36
OW388150.1	8	15.28	36.4
OW388151.1	9	14.97	36.1
OW388152.1	10	14.96	35.9
OW388153.1	11	14.26	36.4
OW388154.1	12	14.25	36.4
OW388155.1	13	14.23	36.3
OW388156.1	14	14.13	35.9

INSDC accession	Chromosome	Size (Mb)	GC%
OW388157.1	15	13.83	36.4
OW388158.1	16	13.39	36.5
OW388159.1	17	13.35	36.3
OW388160.1	18	13.25	36.7
OW388161.1	19	12.09	36.7
OW388162.1	20	12.09	36
OW388163.1	21	11.92	36.6
OW388164.1	22	10.99	37.2
OW388165.1	23	10.47	36.3
OW388166.1	24	9.99	37.1
OW388167.1	25	9.61	36.6
OW388168.1	26	9.4	36.7
OW388169.1	27	7.83	38.8
OW388170.1	28	7.64	37.2
OW388171.1	29	7.42	37.3
OW388142.1	Z	25.69	36
OW388172.1	MT	0.02	18.3

Table 3. Software tools and versions used.

Software tool	Version	Source
BlobToolKit	3.4.0	(Challis *et al.*, 2020)
freebayes	1.3.1-17-gaa2ace8	(Garrison & Marth, 2012)
Hifiasm	0.15.3-r339)	(Cheng *et al.,* 2021)
HiGlass	1.11.6	(Kerpedjiev *et al.,* 2018)
Long Ranger ALIGN	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
MitoHiFi	2.0	(Uliano-Silva *et al.,* 2021)
PretextView	0.2	(Harry, 2022)
purge_dups	1.2.3	(Guan *et al.,* 2020)
SALSA	2.2	(Ghurye *et al.,* 2019)

Ethics/compliance issues

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. 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. 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: *Meganola albula* (Kent black arches). Accession number PRJEB46332; https://identifiers.org/ena.embl/PRJEB46332 (Wellcome Sanger Institute, 2022).

The genome sequence is released openly for reuse. The *Meganola albula* 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.

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4789928.

Members of the Natural History Museum Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4790042.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893703.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.4783585.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790455.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

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

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**

Cha YB, Oh SH, Bayarsaikhan U, *et al.*: **Review of the genus Meganola Dyar, 1898 (Lepidoptera: Nolidae, Nolinae) *from* Korea, with the description of a new species.** *Zootaxa.* 2019; **4686**(2): 215–240.
PubMed Abstract | **Publisher 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**

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012.
Publisher Full Text

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
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; (Accessed: 19 October 2022).
Reference Source

Heath J, Alford DV: **Provisional atlas of the insects of the British Isles.** Biological Records Centre, 1973.

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

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kiss Á: **Taxonomic review of the *Craniophora* s. l. (Lepidoptera, Noctuidae, Acronictinae) generic complex with description of 8 new genera and 13 new species.** *Zootaxa.* 2017; **4355**(1): 1–90.
PubMed Abstract | **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**

NBN Atlas: ***Meganola albula* ([Denis & Schiffermüller], 1775).** (Accessed: 27 November 2022).
Reference Source

Õunap E, Choi SW, Matov A, *et al.*: **Description of *Nola estonica* sp. nov., with comparison to *N. aerugula* and *N. atomosa* stat. rev. (Lepidoptera, Nolidae, Nolinae).** *Zootaxa.* 2021; **5082**(5): 401–424.
PubMed Abstract | **Publisher Full Text**

Perveen FK, Khan A: **Introductory Chapter: Lepidoptera.** In: *Lepidoptera.* InTech, 2017.
Publisher Full Text

Rao SS, 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**

Seymour SJJ: **Larva Collecting and Breeding: A Handbook to the Larvae of the British Macro-Lepidoptera and Their Food Plants; Both in Nature and in Confinement, with Authorities.** Fb&c Limited, 2018.
Reference Source

Skinner B, Wilson D: **Colour Identification Guide to the Moths of the British Isles.** Leiden, Netherlands: Brill, 2009.

Tshistjakov YA: **A Review of the Subfamily Nolinae (Lepidoptera, Nolidae) of the Russian Far East.** *Far Eastern Entomologist.* 2008; **185**: 1–19.
Reference Source

Uliano-Silva M, *et al.*: **MitoHiFi.** 2021; (Accessed: 19 October 2022).
Reference Source

Wellcome Sanger Institute: **The genome sequence of the Kent Black Arches, Meganola albula (Denis & Schiffermüller, 1775).** *European Nucleotide Archive.* PRJEB46332; [Dataset]. 2022.
https://identifiers.org/ena.embl/PRJEB46332

Zahiri R, Holloway JD, Kitching IJ, *et al.*: **Molecular phylogenetics of Erebidae (Lepidoptera, Noctuoidea).** *Systematic Entomology.* 2012; **37**(1): 102–124.
Publisher Full Text

Zahiri R, Lafontaine JD, Holloway JD, *et al.*: **Major lineages of Nolidae (Lepidoptera, Noctuoidea) elucidated by molecular phylogenetics.** *Cladistics.* 2013; **29**(4): 337–359.
PubMed Abstract | **Publisher Full Text**

Zilli A: **The Macromoths of a coastal marsh habitat in Central Italy.** Monitoring of rock partridge (Alectoris graeca) in Latium Lepidoptera italica. 2014.
Reference Source

Open Peer Review

Current Peer Review Status: ✔ ✔ ❓

❓   **Charles Robin**

The University of Melbourne, Melbourne, Victoria, Australia

This manuscript describes the genome sequencing and assembly of a Lepidopteran species that goes by the quaint common name of 'Kent black arches'.

The affiliations imply that the rationale for this project is part of the wonderful Darwin Tree of Life project. Such a rationale is fine by me but I am inferring it from the affiliations and so I have ticked No on the box above.

I have experience with Lepidopteran genomes and this one seems to stack up well. The length of the assembly and the number of implied chromosomes are consistent with my expectations. The other quality metrics, particularly those in Figure 2, look good too.

In general, I am disappointed (and don't really understand why) the heterogametic sex isn't routinely preferred in genome sequencing projects. We need more W chromosomes! And the consequent ratio of Z to autosomes reads may also help assemblies.

There is confusion over the first/second/third specimens. The first sentence of the Methods refers to collection of two specimens. But the third sentence refers to ' A second specimen' (isn't this the third?)

Figure 3 is too cryptic. What is the Y axis exactly? The key shows grey box but that cannot be identified in the picture. What do the small dots indicate? This is not a self-standing figure.

Is the rationale for creating the dataset(s) clearly described?
No

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: Insect genetics

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.

Reviewer Report 12 June 2023

https://doi.org/10.21956/wellcomeopenres.20707.r58649

✔️ **Tobias Baril** iD
Institute of Biology, University of Neuchâtel, Neuchâtel, Canton of Neuchâtel, Switzerland

The authors provide a high-quality genome assembly for the Kent black arches moth, which is available publicly and for reuse immediately.

Concise background information on the species is provided, and the introduction is well-written. The sequencing and assembly protocols are clear and well-described, and accessions are provided to access the data generated for this project. Version numbers are present where appropriate for bioinformatics workflows.

It would be useful for more clarification to be provided on the expert identification of this species to confirm this is indeed the Kent black arches, particularly as the authors state this bares "*close similarity to Scarce Black Arches (Nola aerugula), another species found in similar coastal UK habitats*" and the mentioned scarcity of this species in the UK.

Minor Points
  ○ I would suggest changing 'til' to something more appropriate such as "until" or "to" in this context.

  ○ Adding images of the other specimens used in this study would be a nice touch, and also potentially provide visual support for the identification of the species.

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, Bioinformatics tools development, Transposable elements, Comparative Genomics

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

Reviewer Report 05 June 2023

https://doi.org/10.21956/wellcomeopenres.20707.r56894

✔ **Raghuvir Viswanatha**
Department of Genetics, Blavatnik Institute - Harvard Medical School, Boston, Massachusetts, USA

The authors provide an accessible, high-quality genome for the Kent black arches, a species of moth.

The authors should provide more detail about the processes used by the experts to authenticate the species. In particular, they should outline the observations they used to avoid misidentification.

Based on comparison of the genome sequence of ilMegAlbu1 to the RNA sequence of ilMegAlbu4, the authors should state their confidence that these samples are of the same species.

Minor points
- At least one instance of 'till' should be changed to 'until'.

- The authors might present images of the other specimens used in the study (ilMegAlbu2 and ilMegAlbu4).

- The authors might make a statement about how improved coverage could be obtained and how chromosome phasing could be achieved.

○ The authors might include a sentence introducing outstanding biological questions that could be addressed with the data presented (e.g., to understand the mechanism of recent changes in species abundance, etc...)

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: Genetics and functional genomics. I cannot assess the details of genome assembly methods.

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.