Edinburgh Research Explorer

The genome sequence of the high brown fritillary, Fabriciana adippe (Dennis & Schiffermüller, 1775)

Check for updates

DATA NOTE

The genome sequence of the high brown fritillary, *Fabriciana adippe* (Dennis & Schiffermüller, 1775) [version 1; peer review: awaiting peer review]

Konrad Lohse [ID]1, Roger Vila [ID]2, Alex Hayward3, Dominik R Laetsch [ID]1,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Niklas Wahlberg [ID]4,
Darwin Tree of Life Consortium

1Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, UK
2Institut de Biologia Evolutiva (CSIC - Universitat Pompeu Fabra), Barcelona, Spain
3College of Life and Environmental Sciences, Department of Biosciences, University of Exeter, Exeter, UK
4Department of Biology, Lund University, Lund, Sweden

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the
article can be found at the end of the article.

Abstract
We present a genome assembly from an individual female *Fabriciana adippe* (the high brown fritillary; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The genome sequence is 485 megabases in span. Most of the assembly (99.98%) is scaffolded into 29 chromosomal pseudomolecules with the Z sex chromosome assembled. The complete mitochondrial genome was also assembled and is 15.1 kilobases in length. Gene annotation of this assembly in Ensembl identified 13,536 protein coding genes.

Keywords
Fabriciana adippe, high brown fritillary, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Lohse K: Investigation, Resources, Writing – Review & Editing; **Vila R**: Investigation, Resources, Writing – Review & Editing; **Hayward A**: Investigation, Writing – Review & Editing; **Laetsch DR**: Investigation, Writing – Review & Editing; **Wahlberg N**: Writing – Original Draft Preparation;

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Papilionoidea; Nymphalidae; Heliconiinae; Argynnini; *Fabriciana*; *Fabriciana adippe* (Dennis & Schiffermüller, 1775) (NCBI:txid405009).

Background

The high brown fritillary, *Fabriciana adippe* (Dennis and Schiffermüller, 1775), is a widespread and common species in Eurasia, but is one of Britain's most endangered butterflies (Fox *et al.*, 2015). It is found in open habitats such as forest clearings, which are often covered in bracken. *F. adippe* is a univoltine summer species that can be seen on the wing from May to the beginning of September. In the UK, larval host plants are violets (*Viola* sp.), for example the common dog-violet (*Viola riviniana*). The species was placed in the genus *Argynnis* in the past, but recent molecular work suggests that *F. adippe* and allies are more closely related to the genus *Speyeria* than to *Argynnis*, hence its placement in the genus *Fabriciana* (de Moya *et al.*, 2017).

While *F. adippe* is considered a species of Least Concern according to the IUCN Red List for Europe (van Swaay *et al.*, 2010), it is listed as endangered on the UK Red List (Fox *et al.*, 2022), with only 37 populations remaining in the UK (Ellis *et al.*, 2019). Decreased coppicing and local conservation efforts to implement changes in forest management have enjoyed some success in protecting this species (Ellis *et al.*, 2019). Today, *F. adippe* is restricted to a few localities in western England and Wales. It is a large, fast flying species, and can be confused with several other similar sized species,

including the dark green fritillary (*Speyeria aglaja*) and the niobe fritillary (*F. niobe*).

Fabriciana adippe has 29 chromosome pairs and a ZO sex-determination system (Federley, 1938; Lorković, 1941). The genome sequencing of the high brown fritillary may help in conservation efforts and in understanding its population structure, which is notably marked. In fact, Iberian populations are strongly differentiated from other populations in both nuclear and mitochondrial genomes, and admixed populations seem to exist at the contact zone (Dapporto *et al.*, 2022; Polic *et al.*, 2022).

Genome sequence report

The genome was sequenced from a single female *F. adippe* collected from Lupşa, Apuseni Mountains, Alba, Romania (Figure 1). A total of 49-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 79-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 26 missing/misjoins and removed six haplotypic duplications, reducing the assembly size by 1.90% and the scaffold number by 9.62%.

The final assembly has a total length of 485 Mb in 94 sequence scaffolds with a scaffold N50 of 16.7 Mb (Table 1). Most of the assembly sequence (99.98%) was assigned to 29 chromosomal-level scaffolds, representing 28 autosomes (numbered by sequence length) and the Z sex chromosome (Figure 2–Figure 5; Table 2).

The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 98.9% (single 98.8%, duplicated 0.1%)

Figure 1. Forewings and hindwings of the female *F. adippe* specimen from which the genome was sequenced. Dorsal (left) and ventral (right) surface view of wings from specimen RO_FA_930 (ilFabAdip1) from Lupşa, Alba, Romania, used to generate Pacific Biosciences, 10X genomics and Hi-C data.

Table 1. Genome data for *F. adippe*, ilFabAdip1.1.

Project accession data	
Assembly identifier	ilFabAdip1.1
Species	*Fabriciana adippe*
Specimen	ilFabAdip1 (genome assembly, Hi-C); ilFabAdip2 (RNA-Seq)
NCBI taxonomy ID	405009
BioProject	PRJEB43797
BioSample ID	SAMEA7523308
Isolate information	ilFabAdip1: female, abdomen tissue (genome assembly), thorax tissue (Hi-C). ilFabAdip2: male, abdomen tissue (RNA-Seq).
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6565939
10X Genomics Illumina	ERR6054608-ERR6054611
Hi-C Illumina	ERR6054612
PolyA RNA-Seq Illumina	ERR6054613
Genome assembly	
Assembly accession	GCA_905404265.1
Accession of alternate haplotype	GCA_905404255.1
Span (Mb)	485
Number of contigs	110
Contig N50 length (Mb)	15.9
Number of scaffolds	94
Scaffold N50 length (Mb)	16.7
Longest scaffold (Mb)	26.1
BUSCO* genome score	C:98.9%[S:98.8%,D:0.2%],F:0.2%,M:0.9%,n:5,286
Genome annotation	
Number of protein-coding genes	13,536
Number of noncoding genes	8,725

*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 comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/ilFabAdip1.1/dataset/CAJQFA01/busco.

using the lepidoptera_odb10 reference set (n = 5,286). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Genome annotation report

The *F. adippe* genome was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Fabriciana_adippe_GCA_905404265.1/). The resulting annotation
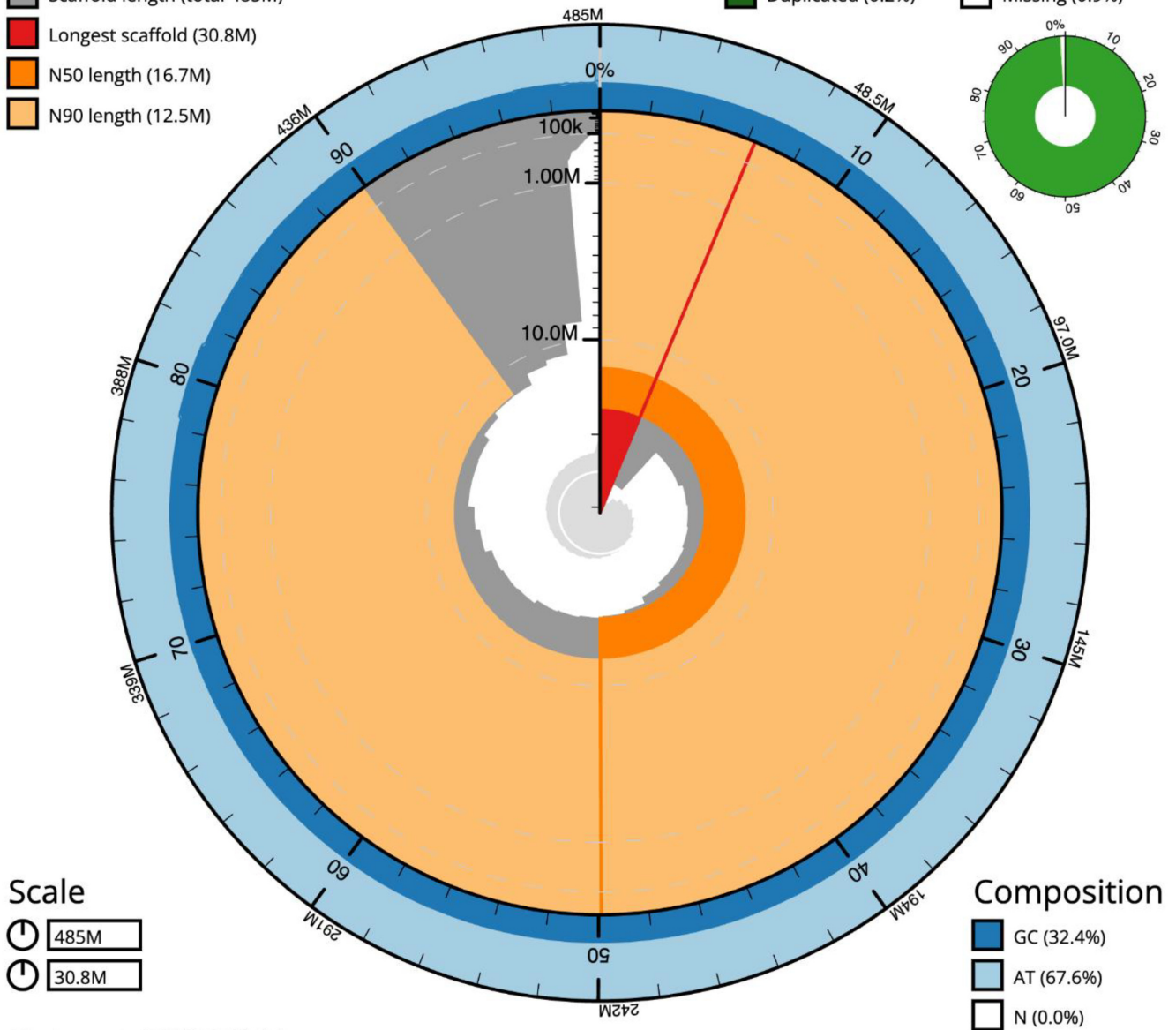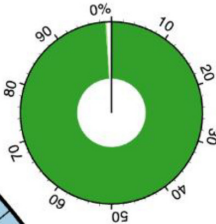
Figure 2. Genome assembly of *F. adippe*, ilFabAdip1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 484,826,487 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (30,838,137 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (16,739,368 and 12,513,400 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/ilFabAdip1.1/dataset/CAJQFA01/snail.
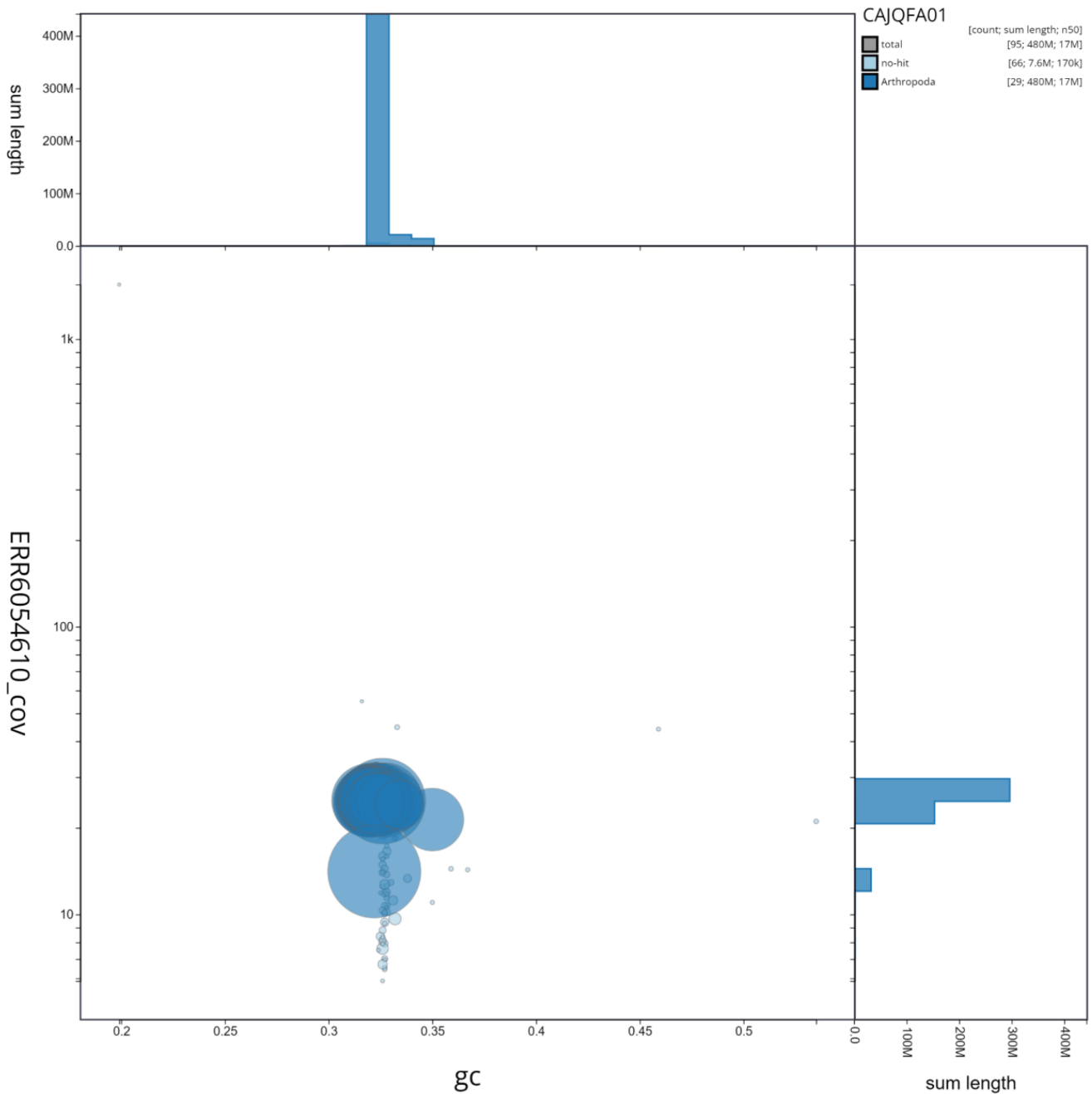
Figure 3. Genome assembly of *F. adippe*, ilFabAdip1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilFabAdip1.1/dataset/CAJQFA01/blob.

includes 35,064 transcribed mRNAs from 13,536 protein-coding and 8,725 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

Two *F. adippe* specimens (ilFabAdip1, female, genome assembly and Hi-C; ilFabAdip2, male, RNA-Seq) were collected using a handnet in Lupșa, Apuseni Mountains, Alba, Romania (latitude 46.416, longitude 23.192) by Konrad Lohse, Dominik Laetsch (University of Edinburgh), Alex Hayward (University of Exeter) and Roger Vila (Institut de Biologia Evolutiva, Barcelona). The specimens were identified by Roger Vila and snap-frozen from live in a dry shipper.

DNA was extracted at the Scientific Operations Core, Wellcome Sanger Institute. The ilFabAdip1 sample was weighed and
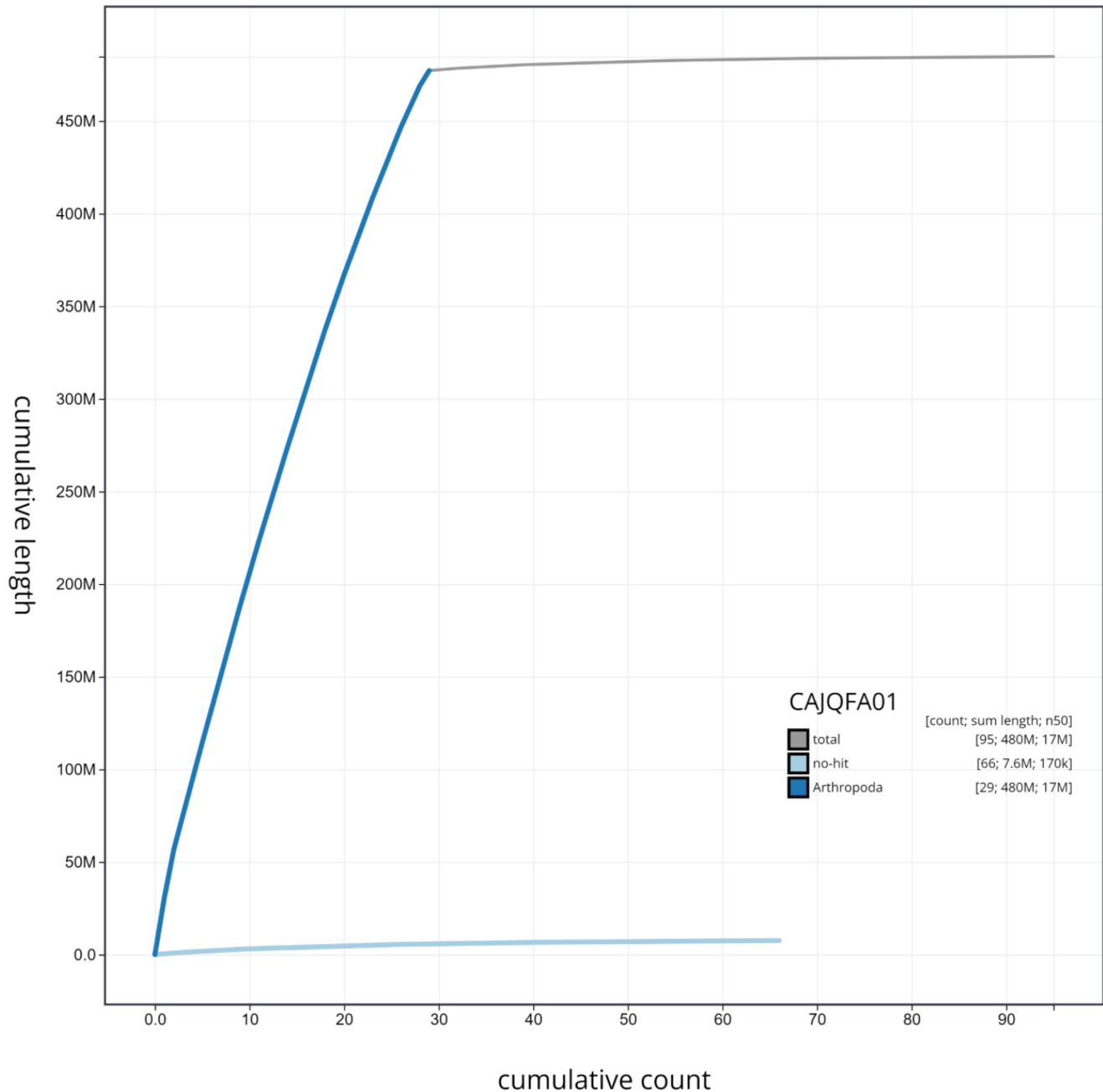
Figure 4. Genome assembly of *F. adippe*, ilFabAdip1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilFabAdip1.1/dataset/CAJQFA01/cumulative.

dissected on dry ice with thorax tissue set aside for Hi-C sequencing. Abdomen tissue was disrupted by manual grinding with a disposable pestle. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200 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 system with speed setting 30. Sheared DNA
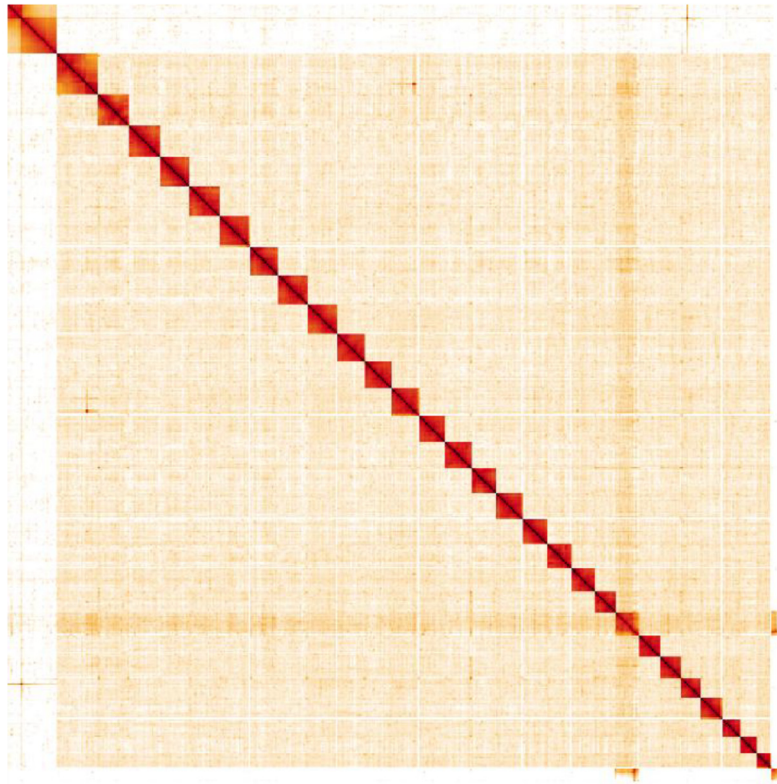
Figure 5. Genome assembly of *F. adippe*, ilFabAdip1.1: Hi-C contact map. Hi-C contact map of the ilFabAdip1.1 assembly, visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. The interactive Hi-C map can be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=G2lH9zghRgeH6ugDELRW7g.

Table 2. Chromosomal pseudomolecules in the genome assembly of *F. adippe*, ilFabAdip1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
FR989983.1	1	26.05	32.6
FR990003.1	2	13.83	35
FR989984.1	3	19.37	32.7
FR989985.1	4	19.29	32.2
FR989986.1	5	18.82	32.3
FR989987.1	6	18.7	31.9
FR989988.1	7	18.55	32.1
FR989989.1	8	18.33	32.5
FR989990.1	9	18.32	32.1
FR989991.1	10	17.98	32
FR989992.1	11	17.21	31.9
FR989993.1	12	16.91	32
FR989994.1	13	16.74	32.1
FR989995.1	14	16.56	32.1

INSDC accession	Chromosome	Size (Mb)	GC%
FR989996.1	15	16.42	32.4
FR989997.1	16	15.98	32.1
FR989998.1	17	15.9	32.3
FR989999.1	18	15.65	32.1
FR990000.1	19	14.94	32.5
FR990001.1	20	14.47	32.5
FR990002.1	21	13.83	32.4
FR990004.1	22	13.75	32.1
FR990005.1	23	13.08	33.1
FR990006.1	24	12.55	32.5
FR990007.1	25	12.51	32.2
FR990008.1	26	11.41	32.7
FR990009.1	27	10.96	32.3
FR990010.1	28	8.35	33.4
FR989982.1	Z	30.84	32.2
FR990011.1	MT	0.02	19.4
-	Unplaced	7.56	33

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. 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 abdomen tissue of ilFabAdip2 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 RNA 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.

Sequencing
Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq 10X and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated in the Tree of Life laboratory from thorax tissue of ilFabAdip1 using the Arima v2 kit and sequenced on a HiSeq 10X instrument.

Genome assembly
Assembly was carried out with Hifiasm (Cheng et al., 2021); haplotypic duplication was identified and removed with purge_dups (Guan et al., 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger 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 and corrected using the gEVAL system (Chow et al., 2016) as described previously (Howe et al., 2021). Manual curation (Howe et al., 2021) was performed using gEVAL, HiGlass (Kerpedjiev et al., 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva et al., 2021), which performs 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.

Genome annotation
The Ensembl gene annotation system (Aken et al., 2016) was used to generate annotation for the *F. adippe* assembly (GCA_905404265.1). Annotation was created primarily through alignment of transcriptomic data to the genome, with

Table 3. Software tools used.

Software tool	Version	Source
BlobToolKit	3.2.6	Challis et al., 2020
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
gEVAL	N/A	Chow et al., 2016
Hifiasm	0.12	Cheng et al., 2021
HiGlass	1.11.6	Kerpedjiev et al., 2018
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
MitoHiFi	1.0	Uliano-Silva et al., 2021
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	Guan et al., 2020
SALSA2	2.2	Ghurye et al., 2019

gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

Data availability
European Nucleotide Archive: *Fabriciana adippe* (high brown fritillary). Accession number PRJEB43797; https://identifiers.org/ena.embl/PRJEB43797 (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *F. adippe* 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. Raw data and assembly accession identifiers are reported in Table 1.

Author information
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

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
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**

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**

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL - a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Dapporto L, Menchetti M, Vodă R, *et al.*: **The atlas of mitochondrial genetic diversity for Western Palaearctic butterflies.** *Global Ecology and Biogeography.* 2022; **31**(11): 2184–2190.
Publisher Full Text

de Moya RS, Savage WK, Tenney C, *et al.*: **Interrelationships and diversification of *Argynnis* fabricius and *Speyeria* scudder butterflies.** *Systematic Entomology.* 2017; **42**(4): 635–649.
Publisher Full Text

Ellis S, Wainwright D, Dennis EB, *et al.*: **Are habitat changes driving the decline of the UK's most threatened butterfly: the High Brown Fritillary *Argynnis adippe* (Lepidoptera: Nymphalidae)?** *J Insect Conserv.* 2019; **23**(2): 351–367.
Publisher Full Text

Federley H: **Chromosomenzahlen Finnlän-Discher Lepidopteren: I. Rhopalocera.** *Hereditas.* 1938; **24**(4): 397–464.
Publisher Full Text

Fox R, Brereton TM, Asher J, *et al.*: **The State of the UK's Butterflies 2015.** Wareham, Dorset: Butterfly Conservation and the Centre for Ecology & Hydrology, 2015; (Accessed: 25 October 2022).
Reference Source

Fox R, Dennis EB, Brown AF, *et al.*: **A revised Red List of British butterflies.** *Insect Conservation and Diversity.* 2022; **15**(5): 485–495.
Publisher 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

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**

Lorković Z: **Die Chromosomenzahlen in Der Spermatogenese Der Tagfalter.** *Chromosoma.* 1941; **2**(1): 155–191.
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**

Polic D, Yıldırım Y, Lee KM, *et al.*: **Linking large-scale genetic structure of three Argynnini butterfly species to geography and environment.** *Mol Ecol.* 2022; **31**(16): 4381–4401.
PubMed Abstract | **Publisher Full Text** | **Free 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**

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

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

van Swaay C, Cuttelod A, Collins S, *et al.*: **European Red List of Butterflies.** IUCN (International Union for Conservation of Nature) and Butterfly Conservation Europe in collaboration with the European Union, 2010.
Publisher Full Text

Wellcome Sanger Institute: **The genome sequence of the high brown fritillary, Fabriciana adippe (Dennis & Schiffermüller, 1775), European Nucleotide Archive.** [dataset]. accession number PRJEB43797. 2022.