DATA NOTE

The genome sequence of the small skipper, *Thymelicus sylvestris* (Poda, 1761) [version 1; peer review: 1 approved, 1 approved with reservations]

Alex Hayward¹, Ryan Biscocho¹, 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, Darwin Tree of Life Consortium

¹Department of Biosciences, University of Exeter, Penryn, TR10 9FE, UK

V1 First published: 01 Feb 2022, 7:35
<https://doi.org/10.12688/wellcomeopenres.17577.1>
 Latest published: 01 Feb 2022, 7:35
<https://doi.org/10.12688/wellcomeopenres.17577.1>

Abstract

We present a genome assembly from an individual male *Thymelicus sylvestris* (the small skipper; Arthropoda; Insecta; Lepidoptera; Hesperidae). The genome sequence is 471 megabases in span. The majority of the assembly (99.97%) is scaffolded into 27 chromosomal pseudomolecules, with the Z sex chromosome assembled. The mitochondrial genome was also assembled and is 17.1 kilobases in length.

Keywords

Thymelicus sylvestris, small skipper, genome sequence, chromosomal, Lepidoptera

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

Open Peer Review

Approval Status ✓ ?

	1	2
version 1	✓	?
01 Feb 2022	view	view
1. Francois Olivier Hebert , Université de Montréal, Montréal, Canada Université de Moncton, Moncton, Canada		
2. Xueyan Li , Chinese Academy of Sciences, Kunming, China		
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: **Hayward A:** Investigation, Resources, Supervision, Writing – Review & Editing; **Biscocho R:** Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328). AH is supported by a Biotechnology and Biological Sciences Research Council (BBSRC) David Phillips Fellowship (BB/N020146/1).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Hayward A *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Hayward A, Biscocho R, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the small skipper, *Thymelicus sylvestris* (Poda, 1761) [version 1; peer review: 1 approved, 1 approved with reservations]** Wellcome Open Research 2022, 7:35 <https://doi.org/10.12688/wellcomeopenres.17577.1>

First published: 01 Feb 2022, 7:35 <https://doi.org/10.12688/wellcomeopenres.17577.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Hesperioidea; Hesperidae; Hesperinae; Hesperini; *Thymelicus*; *Thymelicus sylvestris* (Poda, 1761) (NCBI:txid272628).

Background

The small skipper (*Thymelicus sylvestris*) is a butterfly within the skipper family Hesperidae. The skippers are named for their characteristic quick, darting flight. The common name of *T. sylvestris* is a clear reference to its small size: the adult wingspan ranges from 27–34 mm (Tomlinson & Still, 2002). However, it is not the smallest of the skippers, with four other British species being an equivalent size or smaller. Similar to other skippers, *T. sylvestris* has golden-orange wings with clear sex brands on males, but it can be distinguished by a lack of coloured patches on its wings and a dull brown or orange colouration to its antennae (Tomlinson & Still, 2002).

Thymelicus sylvestris is widespread across the European continent with a habitat range paralleling that of other skipper species. This range encompasses the northernmost reaches of Morocco and Algeria all the way to the bordering regions between the Baltic states and Russia (Tolman & Lewington, 2008). However, it is noticeably absent from northern Scandinavia, Corsica and Sardinia (Tolman & Lewington, 2008). In the British Isles the small skipper is found across most of Wales and England with recent trends showing a northward expansion in range, beyond the England-Scotland border.

Recently, *T. sylvestris* individuals have also been observed in Ireland, where they had not been reported previously (Harding & Jacob, 2013). *Thymelicus sylvestris* populations appear stable and it is listed as a species of least concern by the IUCN (van Swaay *et al.*, 2010).

The small skipper is a habitat generalist (Louy *et al.*, 2007) and can be found in open areas with long grass, such as rough grasslands and roadside verges (Tomlinson & Still, 2002). It is most associated with Yorkshire fog (*Holcus lanatus*), its main food plant, on which it often basks and lays eggs from June to July. Females are known to be meticulous with their choice of oviposition sites, spending up to 15 minutes inspecting potential host plants prior to laying eggs (Tolman & Lewington, 2008). After approximately a month, eggs hatch into caterpillars which develop through 5 instar stages. Come winter, caterpillars spin cocoons within which they undergo diapause. The caterpillars re-emerge in spring, constructing a 'leaf tube' by joining together the ends of a leaf, where they live and feed, moving to new leaves as necessary. Small skipper caterpillars usually pupate by June, with adult butterflies emerging in July, to spend their remaining days in tall grassland until the summer's end in September.

Genome sequence report

The genome was sequenced from a single male *T. sylvestris* collected from Ruan Minor, Cornwall, UK (latitude 49.9942295, longitude -5.1974720) (Figure 1). A total of 40-fold coverage in Pacific Biosciences single-molecule long reads and 63-fold coverage in 10X Genomics read clouds were

Figure 1. Fore and hind wings of the *Thymelicus sylvestris* specimen from which the genome was sequenced. Dorsal (left) and ventral (right) surface view of wings from specimen EN_RM_1378 (ilThySylv1) from Ruan Minor, Cornwall, UK, used to generate Pacific Biosciences and 10X genomics data.

generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 9 missing/misjoins and removed 3 haplotypic duplications, reducing the assembly size by 0.06% and scaffold number by 20.00%.

The final assembly has a total length of 471 Mb in 32 sequence scaffolds with a scaffold N50 of 17 Mb (Table 1). Of the assembly sequence, 99.97% was assigned to 28 chromosomal-level scaffolds, representing 27 autosomes (numbered by sequence length), and the Z sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO (Simão *et al.*, 2015) v5.1.2

completeness of 98.5% (single 98.1%, duplicated 0.5%) using the lepidoptera_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods

Specimen acquisition and nucleic acid extraction

Three male *T. sylvestris* (ilThySylv1, ilThySylv2 and ilThySylv3) specimens were collected from Ruan Minor, Cornwall, UK (latitude 49.9942295, longitude -5.1974720) using a net by Alex Hayward in May 2019. The samples were identified by the same individual and snap-frozen on dry ice.

Table 1. Genome data for *Thymelicus sylvestris*, ilThySylv1.1.

Project accession data	
Assembly identifier	ilThySylv1.1
Species	<i>Thymelicus sylvestris</i>
Specimen	ilThySylv1/EN_RM_1378 (genome assembly); ilThySylv2 (RNA-Seq); ilThySylv3 (Hi-C)
NCBI taxonomy ID	NCBI:txid272628
BioProject	PRJEB41953
BioSample ID	SAMEA7523279
Isolate information	Male, whole organisms
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6608659
10X Genomics Illumina	ERR6363310-ERR6363313
Hi-C Illumina	ERR6363315
Illumina PolyA RNA-Seq	ERR6363314
Genome assembly	
Assembly accession	GCA_911387775.1
Accession of alternate haplotype	GCA_911387695.1
Span (Mb)	471
Number of contigs	46
Contig N50 length (Mb)	16.6
Number of scaffolds	32
Scaffold N50 length (Mb)	17.3
Longest scaffold (Mb)	21.0
BUSCO* genome score	C:98.5%[S:98.1%,D:0.5%],F:0.3%,M:1.1%,n:5286

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.1.2. 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/ilThySylv1.1/dataset/CAJVQR01/busco>.

Scaffold statistics

- Log10 scaffold count (total 33)
- Scaffold length (total 470M)
- Longest scaffold (37M)
- N50 length (17M)
- N90 length (12M)

BUSCO

lepidoptera_odb10 (5286)

- Complete (98.5%)
- Fragmented (0.3%)
- Duplicated (0.5%)
- Missing (1.1%)

Scale

Dataset: CAJVQR01

Composition

- GC (36.1%)
- AT (63.9%)
- N (0.0%)

Figure 2. Genome assembly of *Thymelicus sylvestris*, ilThySylv1.1: metrics. 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 470,727,450 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 (37,236,842 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (17,253,319 and 11,732,099 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/ilThySylv1.1/dataset/CAJVQR01/snail>.

DNA was extracted from the whole organism of ilThySylv1 (specimen ID: EN_RM_1378) at the Wellcome Sanger Institute (WSI) Scientific Operations core from the whole organism

using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions. RNA from whole organism tissue of ilThySylv2 was extracted in the Tree of Life Laboratory at

Figure 3. Genome assembly of *Thymelicus sylvestris*, ilThySylv1.1: 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/ilThySylv1.1/dataset/CAJVQR01/blob>.

the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 µl RNase-free water and its concentration 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 the 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. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific

Figure 4. Genome assembly of *Thymelicus sylvestris*, ilThySylv1.1: 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/ilThySylv1.1/dataset/CAJVQR01/cumulative>.

Biosciences SEQUEL II (HiFi), Illumina HiSeq X (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated from head tissue of ilThySylv3 in the Tree of Life Laboratory using the Arima Hi-C+ kit and sequenced on an Illumina NovaSeq 6000 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

Figure 5. Genome assembly of *Thymelicus sylvestris*, ilThySylv1.1: Hi-C contact map. Hi-C contact map of the ilThySylv1.1 assembly, visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom.

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 was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021). 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.

Data availability

European Nucleotide Archive: *Thymelicus sylvestris* (small skipper). Accession number PRJEB45673; <https://identifiers.org/ena.embl/PRJEB45673>.

The genome sequence is released openly for reuse. The *T. sylvestris* 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 the RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics

Table 2. Chromosomal pseudomolecules in the genome assembly of *Thymelicus sylvestris*, ilThySylv1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
OU426886.1	1	20.99	36.0
OU426887.1	2	20.74	36.9
OU426888.1	3	20.31	36.3
OU426889.1	4	19.76	36.3
OU426890.1	5	19.60	36.5
OU426891.1	6	19.09	35.7
OU426892.1	7	18.50	35.9
OU426893.1	8	18.16	35.9
OU426894.1	9	17.98	35.8
OU426895.1	10	17.90	36.0
OU426896.1	11	17.25	35.6
OU426897.1	12	17.04	35.8
OU426898.1	13	16.99	36.2
OU426899.1	14	16.61	36.0
OU426900.1	15	16.11	36.2
OU426901.1	16	16.10	35.9
OU426902.1	17	15.35	35.9
OU426903.1	18	15.05	36.0
OU426904.1	19	14.87	36.2
OU426905.1	20	14.87	36.6
OU426906.1	21	14.38	36.2
OU426907.1	22	13.89	36.3
OU426908.1	23	11.73	36.0
OU426909.1	24	11.38	36.7
OU426910.1	25	10.65	36.1
OU426911.1	26	9.84	36.9
OU426912.1	27	8.15	36.4
OU426885.1	Z	37.24	35.7
OU426913.1	MT	0.02	16.7
-	Unplaced	0.18	41.7

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.15	Cheng et al., 2021
purge_dups	1.2.3	Guan et al., 2020
SALSA2	2.2	Ghurye et al., 2019
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
MitoHiFi	2.0	Uliano-Silva et al., 2021
gEVAL	N/A	Chow et al., 2016
HiGlass	1.11.6	Kerpedjiev et al., 2018
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	2.6.4	Challis et al., 2020

Institute. Raw data and assembly accession identifiers are reported in [Table 1](#).

Author information

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.5744972>.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: <https://doi.org/10.5281/zenodo.5744840>.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: <https://doi.org/10.5281/zenodo.5746904>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.5743293>.

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.5638618>.

References

- Challis R, Richards E, Rajan J, et al.: **BlobToolKit - Interactive Quality Assessment of Genome Assemblies**. *G3 (Bethesda)*. 2020; **10**(4): 1361–74. [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): 170–75. [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): 2508–10. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing**. arXiv: 1207.3907. 2012. [Reference Source](#)
- 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): 2896–98. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Harding J, Jacob M: **Addition of Small Skipper Butterfly (*Thymelicus sylvestris*) to the Irish List and Notes on the Essex Skipper (*Thymelicus lineola*) (Lepidoptera: Hesperidae).** *The Irish Naturalists' Journal*. 2013; **32**(2): 142–44.

[Reference Source](#)

Howe K, Chow W, Collins J, *et al.*: **Significantly Improving the Quality of Genome Assemblies through Curation.** *GigaScience*. 2021; **10**(1): gaaa153.

[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](#)

Louy D, Habel JC, Schmitt T, *et al.*: **Strongly Diverging Population Genetic Patterns of Three Skipper Species: The Role of Habitat Fragmentation and Dispersal Ability.** *Conserv Genet.* 2007; **8**(3): 671–81.

[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): 1665–80.

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

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.** *Bioinformatics*. 2015; **31**(19): 3210–12.

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

Tolman T, Lewington R: **The Most Complete Guide to the Butterflies of Britain and Europe.** Harper Collins, London, 2008.

Tomlinson D, Still R: **Britain's Butterflies.** WILDGuides, 2002.

[Reference Source](#)

Uliano-Silva M, Nunes JGF, Krashenninnikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0.** 2021.

[Publisher Full Text](#)

van Swaay C, Wynhoff I, Verovnik R, *et al.*: **IUCN Red List of Threatened Species: *Thymelicus sylvestris*.** *IUCN Red List of Threatened Species*. 2010.

[Reference Source](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 09 June 2022

<https://doi.org/10.21956/wellcomeopenres.19436.r50767>

© 2022 Li X. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Xueyan Li

State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

This paper "The genome sequence of the small skipper, *Thymelicus sylvestris* (Poda, 1761)" assembled the chromosomal-level genome of *Thymelicus sylvestris*. Due to the HiFi, 10X Genomics, Hi-C Illumina and RNA-seq, the assembly quality is relatively high. The high-quality reference genomes of more butterfly species will be helpful to investigate the evolution of butterfly diversity.

In general, this paper is well written with a concise description. Nevertheless, the authors didn't provide genome annotation (repeat and genes), which, if unavailable, will limit the use of the data, especially for non-bioinformatic specialists. So, please provide genome annotation information.

Several revisions are needed for the data.

- First, how many chromosomes for this species, please describe it and cite the related literature.
- Second, how the authors identified the Z sex chromosome?

Minor revisions

- Background: "The skippers are named for their characteristic quick, darting flight." to "The skippers are named for their characteristically quick, darting flight."

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?

Partly

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

Yes

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** Insect genome and evolution

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 21 February 2022

<https://doi.org/10.21956/wellcomeopenres.19436.r48488>

© 2022 Hebert F. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Francois Olivier Hebert ¹ CR-CHUM, Université de Montréal, Montréal, Canada² Cannabis Innovation and Research Center, Université de Moncton, Moncton, BC, Canada**General comments:**

This short and well written manuscript presents the genome sequence of a male *Thymelicus sylvestris* (small skipper) sample in the context of the Darwin Tree of Life project. The genome assembly in this work was generated with a combination of PacBio HiFi reads and additional 10X sequencing from a single male individual, and RNAseq and HiC sequencing for genome improvement. They report a final genome sequence of 471 Mb, which is 98.5% complete according to its BUSCO score. Although I am not an expert per say in entomology or specifically in skipper biology, I found this piece of work interesting, written in a straightforward manner and in a simple form. This work is totally worth publishing because it provides a very high-quality reference genome sequence for a species that was still lacking from public databases. It allows the research community to expand its general knowledge on butterfly genomics and it also provides additional data that can be used in the future to perform comparative analyses and better understand insect genome evolution on a broader sense. The authors did a good job at generating a complete and highly contiguous/complete genome sequence.

Although this manuscript is very clear and the assembly of high quality (props to the team for the cleanliness of this assembly), I would have liked to learn a little bit more about the reason(s) behind this genome sequence, i.e., the “what we know”, “what we don’t know”, and what it is that this genome sequence potentially brings to the table to better understand what we don’t know about this species or its related taxa. Just to put the work into context. Maybe on the form of just a sentence on what the ultimate goal of the Darwin Tree of Life project is? And maybe a little bit more details in the methods in terms of command lines and scripts used to run the programs

(which would augment the repeatability of this nice work). But apart from that, great work! Down there are some comments that could increase the quality of the paper if addressed by the authors:

Specific comments:

- Suggestion: just a tiny minor detail, but maybe specify in the abstract that it's 27 chromosomes + Z chromosome, for a total of 28 chromosomes? It is well mentioned in the main text, but it could be phrased differently in the abstract so that it does not create any ambiguity when just reading the abstract by itself.
- Impressive quality in terms of BUSCO score: little duplicated BUSCO genes and 98% single is very good, especially in animal species with (typically) large quantities of repeated elements.
- On that note, is there a report somewhere of the percentage of repeats in this assembly? Or maybe this information is not relevant in the context of the DToL project, but I would have liked to have a sense of how repeated that genome is. I presume that there is no gene annotation report so far? If yes, this would be valuable to provide details on the numbers of genes, gene densities on the chromosomes, and other relevant functional annotation metrics.
- Written in the "general comments" section, but again here: in the introduction, I think it would be interesting for the readership to learn a little bit more about the reasons behind this work. Why produce a genome sequence of this species in particular? How is it part of a larger initiative with a particular goal, for instance here, sequencing numerous species with a super large comparative genomics perspective in mind (i.e., understanding the evolutionary genomics / mapping the genome of all animals living in a common geographic location)? In other words, a brief rationale justifying the dataset.
- In the section Genome sequence report, first paragraph, it says that the genome was sequenced from a single individual male, but later in the Methods section, the authors report the capture of three samples. In Table 1, it looks like these three samples were used to generate different libraries of different sequencing technologies, all involved in the genome assembly method. I was thinking that the authors could specify this detail in the first paragraph of the "Genome sequence report" section when they mention the different technologies used in their study, i.e., which technology was used to sequence which sample (and maybe the reason why three samples)? Perhaps I'm missing something here in terms of assembly strategy, but in the end, if 3 different samples were used, the final genome assembly is not entirely from a single individual... or is it? Why not use these 3 technologies on the same sample? I know that extracting enough DNA material in a single individual insect can sometimes be quite difficult (low yields), so maybe this is the reason why the authors chose to use three samples? Just a short clarification on that note would elevate the quality of the paper... to my humble opinion.
- I am by no means an expert in Lepidoptera, although I've assembled a lepidopteran genome in the past, I am wondering, out of curiosity, how many chromosomes are expected in this species specifically? Is it 28 or 31? I know chromosome numbers are quite variable in Hesperidae, but isn't it supposed to be 31 usually (or ancestrally)? Maybe the authors could briefly comment on that topic in the manuscript, i.e., if this number of chromosomes is expected or not? If there are 98% complete single BUSCO group, we can presume that no chromosomes are missing though, but maybe there are 28 chromosomes assembled here (out of 32 scaffolds?), but there should be 30 or 31 and the rest is in the "unanchored" contigs and remaining scaffolds? Could the authors, please, clarify this?
- Data accessibility: I respect so much the fact that everything from this manuscript seems to be transparent and open source. Maybe I missed something in the data provided with this

work, but is there a mention somewhere of the scripts or command lines used by the authors to produce this assembly? Or maybe all the programs were used with default values (which is sometimes recommended anyways when reading these tools' manual)? The protocols are appropriate, and the work is technically sound, but it lacks a little bit of method details to be repeatable.

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

Partly

Are the protocols appropriate and is the work technically sound?

Yes

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

Partly

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Functional genomics, genome assembly, RNAseq, physiology

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.