

GENOME SEQUENCING

The Complete Genome of Three At-Risk Florida Butterflies: *Cyclargus thomasi bethunebakeri*, *Eumaeus atala* and *Heraclides ponceana*

Deborah A. Triant¹, Stacy Pirro², Akito Y. Kawahara³, Jaret C. Daniels³

¹ Data Analytics Center, Research Computing, University of Virginia, ² Iridian Genomes, ³ McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida

<https://doi.org/10.56179/001c.144799>

Biodiversity Genomes

We present short-read genome assemblies of three butterfly species native to the state of Florida: the Atala butterfly (*Eumaeus atala*), the Miami blue (*Cyclargus thomasi bethunebakeri*), and the Schaus' swallowtail (*Heraclides ponceana*). All species are of conservation concern with the latter two listed as federally endangered. Genome assemblies recovered 77-92% of single-copy orthologous insect genes, providing a valuable resource for advancing lepidopteran genomic research.

Introduction

Butterfly populations across the contiguous United States recently experienced a marked decline, with abundance decreasing by 22% between 2000 and 2020 (Edwards et al. 2025). In Florida, which is home to several federally listed butterfly species, active conservation efforts are underway to safeguard these vulnerable taxa. The Miami blue (*Cyclargus thomasi bethunebakeri*, [Figure 1.A](#)) and the Schaus' swallowtail (*Heraclides ponceana*, [Figure 1.B](#)) are two prominent examples of federally endangered species. In contrast, the Atala butterfly (*Eumaeus atala*, [Figure 1.C](#)) was once considered extinct and never listed but has seen a resurgence in after its rediscovery (Koi and Daniels 2015).

The Miami Blue butterfly was once widespread across southern Florida, but extensive habitat loss led to severe population fragmentation and near extirpation. A remnant breeding population was later discovered, prompting reevaluation of its conservation status and its subsequent listing as endangered (Calhoun, Slotten, and Salvato 2002). Similarly, Schaus' swallowtail suffered drastic declines due to habitat degradation and was among the first butterflies protected under the Endangered Species Act. In this study, we present the genome assemblies of these three taxa providing critical resources for their conservation.

Methods

Tissue used for this study was isolated from individuals that were lab-reared at the McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, in Gainesville, Florida. For *C. t. bethunebakeri*, DNA was extracted from 5-pooled females, while *E. atala* was extracted from a single female and *H. ponceana* from a 2-pooled



Figure 1.A. Miami Blue butterfly, *Cyclargus thomasi bethunebakeri*. Photo by Jaret C. Daniels, Ph.D.



Figure 1.B. Schaus' swallowtail, *H. ponceana*. Photo by Jaret C. Daniels, Ph.D.

female sample. DNA extractions were performed with the Qiagen DNAeasy genomic extraction kit according to standard protocols. A paired-end sequencing library was constructed using the Illumina TruSeq kit following the manufacturer's instructions. The library was sequenced on an Illumina Hi-Seq platform in paired-end, 2 × 150bp format. The resulting fastq files were trimmed with Trimmomatic v0.33 (Bolger, L, and Usadel 2014) to



Figure 1.C. Atala butterfly, *Eumaeus atala*. Photo by Jaret C. Daniels, Ph.D.

remove adapter/primer sequences and low-quality regions with validation using fastqc (Andrews 2014). The trimmed sequences were then assembled by SPAdes v2.5 (Bankevich et al. 2012) followed by a finishing step using Zanfona (Kieras, O'Neill, and Pirro 2021), both set to default parameters. To evaluate the genomes' completeness, we used Busco v5.0.0 (Seppey, Manni, and Zdobnov 2019) to measure the proportion of conserved single-copy orthologs using the insecta_odb10 dataset. A portion of our analyses were conducted using the ACCESS system (Boerner et al. 2023).

Results and Data Availability

Genome assemblies yielded total sequence lengths ranging from 355-522 bp with scaffold N50s spanning 3-6.2Mb. BUSCO scores indicated between 77-92% of the benchmarked, conserved, single-copy orthologous genes within Insecta (Table 1). All raw read data and assembled genomes are available on NCBI (Table 2).

Species	Genome Size (Mb)	Number scaffolds	Scaffold N50	Genome coverage	Complete BUSCO (%)	Complete single-copy BUSCO (%)
<i>Cyclargus thomasi bethunebakeri</i>	363.4	94,323	14.3 kb	120x	92.6	92.4
<i>Eumaeus atala</i>	522.7	52,842	20.5 kb	140x	91.6	91.4
<i>Heracles ponceana</i>	335	270,595	3 Mb	160x	77.4	77.0

Table 1. Genome assembly statistics and Busco scores for butterfly genome assemblies

Species	Common Name	Family	BioProject	Raw reads	GenBank	Genome assembly
<i>Cyclargus thomasi bethunebakeri</i>	Miami blue butterfly	Lycaenidae	PRJNA434121	SRR6727422	GCA_026546445.1	ASM2654644v1
<i>Eumaeus atala</i>	Atala butterfly	Lycaenidae	PRJNA434129	SRR6727440	GCA_026546695.1	ASM2654669v1
<i>Heracles ponceana</i>	Schaus' swallowtail	Papilionidae	PRJNA434126	SRR6727429	GCA_016277805.2	ASM1627780v2

Table 2. Sample identities and NCBI accessions

Despite using exclusively short-reads for these assemblies, two of the three genomes contained over 90% of orthologous genes found with insects (*C. t. bethunebakeri* and *E. atala*). Moreover, these two genomes were among the top 50 highest-quality genomes across a comprehensive dataset of 822 butterfly genomes (Ellis, Storer, and Kawahara 2021). We are confident that these genomes will all be an asset for future Lepidopteran conservation efforts.

.....

Funding

Funding was provided by Iridian Genomes, grant # IRGEN_RG_2021-1345 Genomic Studies of Eukaryotic Taxa.

Acknowledgements

This work used Bridges-2 at Pittsburgh Supercomputing Center through allocation [MCB130187] from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing computational resources that have contributed to the research results reported within this paper. URL: <https://www.tacc.utexas.edu>

Submitted: September 24, 2025 EDT. Accepted: September 26, 2025 EDT. Published: September 26, 2025 EDT.

REFERENCES

- Andrews, S. 2014. “FastQC.” <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” *Journal of Computational Biology* 19 (5): 455–77. <https://doi.org/10.1089/cmb.2012.0021>.
- Boerner, T. J., S. Deems, T. R. Furlani, S. L. Knuth, and J. Towns. 2023. “ACCESS: Advancing Innovation: NSF’s Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support.” In *Practice and Experience in Advanced Research Computing (PEARC '23)*. New York, NY, USA: ACM. <https://doi.org/10.1145/3569951.3597559>.
- Bolger, A. M., Marc L, and B. Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Calhoun, J. V., J. R. Slotten, and M. H. Salvato. 2002. “The Rise and Fall of Tropical Blues in Florida: *Cyclargus Ammon* and *Cyclargus Thomasi Bethunebakeri* (Lepidoptera: Lycaenidae).” *Holarctic Lepidoptera* 7 (1): 13–20.
- Edwards, C. B., E. F. Zipkin, E. H. Henry, N. M. Haddad, M. L. Forister, K. J. Burls, S. P. Campbell, et al. 2025. “Rapid Butterfly Declines across the United States during the 21st Century.” *Science* 387 (6738): 1090–94. <https://doi.org/10.1126/science.adp4671>.
- Ellis, E. A., C. G. Storer, and A. Y. Kawahara. 2021. “De Novo Genome Assemblies of Butterflies.” *Gigascience* 10 (6): giab041. <https://doi.org/10.1093/gigascience/giab041>.
- Kieras, M., K. O’Neill, and S. Pirro. 2021. “Zanfona, a Genome Assembly Finishing Tool for Paired-End Illumina Reads.” <https://github.com/zanfona734/zanfona>.
- Koi, S., and J. C. Daniels. 2015. “New and Revised Life History of the Florida Hairstreak *Eumacis Atala* (Lepidoptera: Lycaenidae) with Notes on Its Current Conservation Status.” *Florida Entomologist* 98:1134–47. <https://doi.org/10.1653/024.098.0418>.
- Seppy, M., M. Manni, and E. M. Zdobnov. 2019. “BUSCO: Assessing Genome Assembly and Annotation Completeness.” *Methods in Molecular Biology* 1932:227–45. https://doi.org/10.1007/978-1-4939-9173-0_14.