

DATA NOTE

Open Access



Draft genome of the Cuban Painted Landsnail *Polymita picta*, International Mollusc of the year 2022

Bernardo Reyes-Tur¹ , Zeyuan Chen^{2*} , Mario Juan Gordillo-Pérez³ , Alexander Ben Hamadou² , Charlotte Gerheim² , Carola Greve² and Julia D. Sigwart²

Abstract

Objective The Cuban Painted Landsnail is an iconic endemic tree snail species with distinctive colourful shells used in traditional handicrafts. This species won the International Mollusc of the Year 2022 competition in an open public vote. As the competition prize, we have assembled the draft genome of this species.

Data description Genomic DNA from *Polymita picta* (Born, 1778) was sequenced using PacBio HiFi sequencing with a yield of 5.3 million reads (41.4 Gb) and an N50 of 8.1 Kb. The genome size of *P. picta* was estimated to be 2.9 Gb, and the final assembly was 1.85 Gb, with a total of 22,619 contigs and a contig N50 of 124.2 Kb. BUSCO analysis of the genome assembly indicated a genome completeness of 88.4%, with 7% complete duplicated BUSCOs in metazoa_odb10. The draft genome will be a valuable resource for work on the endangered Cuban Painted Landsnail including monitoring genetic diversity and establishing captive breeding for conservation.

Keywords Mollusca, Adaptive radiation, Gastropoda, Stylommatophora, Cuba, Conservation genomics

Objective

The Cuban Painted Landsnail, *Polymita picta* (Born, 1778), is a visually iconic species recognised for its dramatic shell colour polymorphism [1, 2] (Table 1, Data file 1 [3]). *Polymita picta* is one member of a range-restricted genus, it is endemic to eastern Cuba, and threatened by habitat loss and illegal collection [4, 5]. This species is a focal subject for evolutionary biology, ecology, and conservation [5–8]. Despite its charisma and growing interest from both scientists and collectors [5, 6], *P.*

picta remains genetically understudied with only one short DNA fragment previously published [9], and one recently published mitochondrial genome [10]. Genomic resources for land snails (Stylommatophora) remain underrepresented in molluscan genomics compared to marine bivalves and cephalopods, despite the high species richness and economic importance of land snails [11]. Many available genomes are highly fragmented drafts, limited by the large genome sizes and high repeat content characteristic of gastropods [12], which complicate assembly. A high-quality *P. picta* reference genome will provide a valuable genetic resource to support genetic diversity assessment and conservation efforts, as well as to advance comparative genomic studies of land snails, particularly in the context of speciation, adaptation, and divergence.

*Correspondence:

Zeyuan Chen
zeyuan.chen@senckenberg.de

¹Department of Biology and Geography, Faculty of Natural and Exact Sciences, Universidad de Oriente, Santiago de Cuba, Cuba

²Senckenberg Research Institute and Museum, Frankfurt, Germany

³Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Image of the Cuban Painted Landsnail <i>Polymita picta</i>	Word (.docx)	Figshare [3]: https://doi.org/10.6084/m9.figshare.28790750
Data file 2	HiFi sequencing of <i>P. picta</i> and quality control	Excel (.xlsx)	Figshare [3]: https://doi.org/10.6084/m9.figshare.28790750
Data file 3	Genome survey of <i>P. picta</i>	Portable Document Format (.png)	Figshare [3]: https://doi.org/10.6084/m9.figshare.28790750
Data file 4	Genome assembly statistics at different stages of <i>P. picta</i>	Excel (.xlsx)	Figshare [3]: https://doi.org/10.6084/m9.figshare.28790750
Data file 5	Key statistics of the final genome assembly of <i>P. picta</i>	Excel (.xlsx)	Figshare [3]: https://doi.org/10.6084/m9.figshare.28790750
Data set 1	Pacbio sequence reads of <i>P. picta</i> genomic DNA	Fastq files (.fastq)	NCBI Sequence Read Archive [13]: SRX28390877 http://identifiers.org/insdc.sra:SRX28390877
Data set 2	Genome assembly of <i>P. picta</i>	Fasta file (.fna)	NCBI GenBank Database [14]: JBOZOB000000000 http://identifiers.org/nucleotide:JBOZOB000000000

The International Mollusc of the Year competition has been run annually through the Senckenberg Research Institute and Museum, Frankfurt, Germany, since 2021. In order to bring more attention to the need for increasing genomic resources for molluscs. The winning species is selected by popular vote from five finalists, and as a prize the genome of the winning species is sequenced and assembled, such as this draft for *P. picta*.

Data description

The sample used for genome sequencing was a brown shelled individual with supernumerary spiral bands (juvenile), collected by hand from El Diamante, Maisí municipality, Guantánamo province, Cuba (Table 1, data file 1 [3]). Genome size was estimated following a propidium iodide-stained nuclei flow cytometry (FCM) protocol [15], using the same *P. picta* individual used for PacBio sequencing. Neural tissue from *Acheta domesticus* (female, 1 C=2 Gb [16]) and chicken erythrocyte nuclei (1 C=1.2 Gb [17]) were used as internal reference standards. Here, 1 C refers to the haploid nuclear DNA content of a genome. The genome size (1 C) of *P. picta* was estimated by averaging measurements from three replicates taken on three different days to reduce random instrumental error. Genomic DNA was extracted using a CTAB-based method [18]. We prepared one PacBio

ultra-low input library including a long-range PCR amplification step using the SMRTbell® gDNA Sample Amplification Kit and the SMRTbell® Express Template Preparation Kit 2.0. In addition, to reduce potential PCR biases of the amplification polymerases, we prepared one further library using the KOD Xtreme™ Hot Start DNA Polymerase (Merck), optimized for amplification of long and GC-rich DNA templates. For this, we combined the buffer, dNTPs and KOD polymerase from the KOD Xtreme Hot Start DNA Polymerase Kit with the ultra-low input primers from the PacBio SMRTbell gDNA Sample Amplification Kit. Otherwise, we followed the PacBio ultra-low input protocol [19, 20]. These two ultra-low input libraries were then pooled in equal mass quantities and sequenced on a single Revio SMRTcell.

HiFi reads were called using a pipeline, which is running PacBio's tools ccs 6.4.0 [21], actc 0.3.1 [22], samtools 1.15 [23] and DeepConsensus 1.2.0 [24]. PCR adapter sequences and duplicates were removed using lima 2.9.0 [25] and pbmarkdup 1.0.3 [26], resulting in approximately 5.3 million HiFi reads (41.4 Gb) with an N50 of 8.1 kb (Table 1, Data file 2, Data set 1 [3, 13]), providing approximately 14-fold coverage of the genome of *P. picta* that was estimated at 2.9 Gb using flow cytometry. Genome size and heterozygosity were estimated from a k-mer profile of the HiFi reads using Jellyfish 2.3.0 [27] and GenomeScope 1.0 [28] using the k-mer size k=21, and maximum k-mer coverage of 1000 as suggested. The kmer-based genome size was estimated to be 3.06 Gb, with a heterozygosity about 2.41% (Table 1, Data file 3 [3]). The cleaned HiFi reads were assembled with hifiasm 0.16.1 [29], and Purge_haplotigs 1.1.0 was used to remove duplicated contigs from the genome [30] (Table 1, Data file 4 [3]). FCS-GX 0.5.5, was used to identify and remove contaminant sequences in the new genome [31] (Table 1, Data file 4 [3]). The final *P. picta* genome assembly consists of 22,619 contigs with a total length of 1.85 Gb, a contig N50 of 124.2 Kb (Table 1, Data file 5, Data set 2 [3, 14]). Genome completeness was assessed by BUSCO (Benchmarking Universal Single-Copy Orthologs) 5.4.3 [32], in euk_genome_met mode using the lineage dataset metazoa_odb10 and mollusca_odb10. The results showed 88.4% and 74.9% of complete BUSCOs in metazoa_odb10 (in total 954 genes) and in mollusca_odb10 (in total 5295 genes), respectively. Among them, complete single-copy genes accounted for 81.4% and 65.0%, respectively. Fragmented genes accounted for 7.3% and 4.9%, and missing genes accounted for 4.3% and 20.2%, respectively (Table 1, Data file 5 [3]).

The draft genome presents a valuable resource for monitoring genetic diversity, breeding, resilience to climate change and conservation of the endangered Cuban Painted Landsnail.

Limitations

The polymerase amplification process has yielded a promising amount of HiFi reads for *P. picta*; however, it also shortens the average HiFi read length, resulting in a more fragmented assembly. In addition, the relatively low sequencing coverage (14× relative to the genome size) leaves some genomic regions unassembled. Lower sequencing depths can also potentially lead to errors in kmer-based estimates of genome size. The high heterozygosity of *P. picta* further complicates the assembly process, contributing to a significant proportion of the genome representing alternative haplotypes. In addition, the discrepancy between the assembly and the estimated genome size may be due to highly repetitive sequences. Nonetheless, the assembly quality is comparable to that of other published genomes of land snails and slugs, and it now provides a valuable genetic resource for comparative genomics and the discovery of lineage-specific and adaptive genes.

Abbreviations

PacBio	Pacific Biosciences Sequel II
SMRT	Single Molecule Real Time
BUSCO	Benchmarking Universal Single-Copy Orthologs

Acknowledgements

We thank the Bioscientia Institut für Medizinische Diagnostik GmbH for providing the PacBio SMRT sequencing service on the PacBio Revio platform. BRT is grateful to J.M. Koene and A. Davison for the first advice about The International Mollusc of the Year competition and the 10092 participants that voted for *Polymita picta*. MJGP and BRT were partially financed by VLIR-UOS and Global Minds funds from Hasselt University, Belgium. Z.C. is supported by the Leibniz Association project PHENOME (P123/2021 to J.D.S.).

Authors' contributions

BRT, JDS, CGr conceived the project. BRT collected the samples. ABH and CGe performed DNA extraction and library preparation, ZC performed bioinformatic analysis, BRT, MJG performed investigation. ZC, JDS prepared the original draft, all authors reviewed, edited and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. Mollusc of the Year was a project funded under the LOEWE Translational Biodiversity Genomics initiative.

Data availability

The data described in this Data note can be freely and openly accessed on NCBI SRA under BioProject ID PRJNA1250545 [13], NCBI GenBank under accession number JBOZOB0000000000 [14], and figshare (<https://doi.org/10.6084/m9.figshare.28790750>) [3]. Please see Table 1 for details and links to the data.

Declarations

Ethics approval and consent to participate

Specimens were collected under permits to BRT and exported with authorisation from the relevant authorities in Cuba including prior informed consent and material transfer agreements as addenda to an established collaboration contract between Senckenberg Society and University of Oriente, Santiago de Cuba.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 April 2025 / Accepted: 7 August 2025

Published online: 03 September 2025

References

1. Torre de la C. El género *Polymita*. Memorias de la Sociedad Cubana de Historia Natural. Felipe Poe. 1950; 20:1–20.
2. Gordillo-Pérez MJ, Beenaerts N, Sánchez DL, Smeets K, Arias-Sosa YC, Reyes-Tur B. Shell colour luminance of Cuban painted snails, *Polymita picta* and *Polymita muscarum* (Gastropoda: Cepolidae). PLoS One. 2025;20:e0314008.
3. Reyes-Tur B, Chen Z, Gordillo-Pérez MJ, Hamadou AB, Gerheim C, Greve C, et al. Draft genome of the Cuban painted landsnail *Polymita picta*, international mollusc of the year 2022. Figshare Dataset. 2025. <https://doi.org/10.6084/m9.figshare.28790750>.
4. Milera J, Martínez JR. *Polymita*. 1st ed. La Habana: Editorial Científico Técnica; 1987.
5. González-Guillén A. *Polymita*: the most beautiful land snail of the world. Memphis: Fundcraft Publishing, Collierville; 2014.
6. Berovides V, Valdés G, Milera J. Polimorfismo Genético de *Polymita picta* roseolimbata torre, 1950 En La region de maisi, Cuba. Caribb J Sci. 1986;22:179–89.
7. Reyes-Tur B, Allen JA, Cuellar-Araujo N, Hernández N, Lodi M, Méndez-Hernández AA, et al. Mating behaviour, dart shape and spermatophore morphology of the Cuban tree snail *Polymita picta* (Born, 1780). J Molluscan Stud. 2015;81:187–95.
8. Cancio-Martínez E, Alfonso-Sánchez M, Espinosa-López G. Polimorfismo de Las Esterasas de músculo En Las especies de *Polymita*. Cien Biol. 1987;18:26–32.
9. Koene JM, Schulenburg H. Shooting darts: co-evolution and counter-adaptation in hermaphroditic snails. BMC Evol Biol. 2005;5:25.
10. Lewis A, Reyes-Tur B, Davison A. The mitochondrial genome and phylogenetic placement of the world's most beautiful snail, *Polymita picta*. J Molluscan Stud. 2025. <https://doi.org/10.1093/mollus/eyaf006>.
11. Chen Z. A glimpse into mollusk genomics from the de novo genome of *Arion vulgaris* MoquinTandon, 1855 (Gastropoda, Stylommatophora). Dissertation, LMU München: Faculty of Biology. <https://doi.org/2023.10.5282/edoc.31462>.
12. Chen Z, Baeza JA, Chen C, Gonzalez MT, González VL, Greve C, et al. A genome-based phylogeny for mollusca is concordant with fossils and morphology. Science. 2025;387:1001–7.
13. Reyes-Tur B, Chen Z, Gordillo-Pérez MJ, Hamadou AB, Gerheim C, Greve C, et al. Pacbio HiFi sequencing data of *Polymita picta*. NCBI GeneBank; 2025. <http://identifiers.org/insdc.sra:SRX28390877>.
14. Reyes-Tur B, Chen Z, Gordillo-Pérez MJ, Hamadou AB, Gerheim C, Greve C, et al. Draft genome of the Cuban painted landsnail *Polymita picta*. NCBI GeneBank; 2025. <http://identifiers.org/nucleotide:JBOZOB0000000000>.
15. Hare EE, Johnston JS. Genome size determination using flow cytometry of Propidium Iodide-Stained nuclei. In: Orgogozo V, Rockman MV, editors. MEGA. Totowa, NJ: Humana; 2012. pp. 3–12.
16. Johnston JS, Bernardini A, Hjelmén CE. Genome size estimation and quantitative cytogenetics in insects. Methods Mol Biol. 2019;1858:15–26.
17. Greilhuber J, Temsch EM, Loureiro JCM. Nuclear DNA content measurement. In: Doležel J, Greilhuber J, Suda J, editors. Flow Cytometry with plant cells: analysis of genes, chromosomes and genomes. Weinheim, Wiley-VCH Verlag; 2007. p. 67–101.
18. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucl Acids Res. 1980;8:4321–6.
19. Männer L, Schell T, Spies J, Galià-Camps C, Baranski D, Ben Hamadou A, et al. Chromosome-level genome assembly of the Sacoglossan sea slug *elysia timida* (Risso, 1818). BMC Genomics. 2024;25:941.
20. Bein B, Chrysostomakis I, Arantes LS, Brown T, Gerheim C, Schell T, et al. Long-read sequencing and genome assembly of natural history collection samples and challenging specimens. Genome Biol. 2025;26:25.
21. CCS - Generate. PacBio HiFi data. <https://github.com/PacificBiosciences/ccs>. Accessed 14 April 2025.
22. Align clr to ccs reads. <https://github.com/PacificBiosciences/actc>. Accessed 14 April 2025.
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics. 2009;25:2078–9.

24. Baid G, Cook DE, Shafin K, Yun T, Llinares-López F, Berthet Q, et al. DeepConsensus improves the accuracy of sequences with a gap-aware sequence transformer. *Nat Biotechnol*. 2023;41:232–8.
25. Lima. - Demultiplex PacBio data. <https://github.com/PacificBiosciences/barcoding>. Accessed 14 April 2025.
26. Mark duplicate reads from PacBio sequencing of an amplified library. <https://github.com/PacificBiosciences/pbmarkdup>. Accessed 14 April 2025.
27. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k -mers. *Bioinformatics*. 2011;27(6):764–70.
28. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432.
29. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18:170–5.
30. Roach MJ, Schmidt SA, Borneman AR. Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 2018;19:460.
31. Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol*. 2024;25:60.
32. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editor. *Gene prediction*. New York, NY: Springer New York; 2019. pp. 227–45.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.