





Chromosome-scale Genome Assembly of the Rough Periwinkle *Littorina saxatilis*

Aurélien De Jode ^{1,2,3,*}, Rui Faria^{4,5,6}, Giulio Formenti ⁷, Ying Sims⁸, Timothy P. Smith⁹, Alan Tracey⁸, Jonathan M.D. Wood ⁸, Zuzanna B. Zagrodzka⁶, Kerstin Johannesson^{1,†}, Roger K. Butlin^{1,6,†}, and Erica H. Leder ^{1,10,*†}

¹Department of Marine Sciences, Tjärnö Marine Laboratory, University of Gothenburg, SE 45296 Strömstad, Sweden

²Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, USA

³Dauphin Island Sea Lab, Dauphin Island, AL, USA

⁴CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal

⁵BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal

⁶Ecology and Evolutionary Biology, School of Biosciences, The University of Sheffield, Sheffield S10 2TN, UK

⁷The Vertebrate Genome Laboratory, The Rockefeller University, New York, NY 10065, USA

⁸Tree of Life, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1SA, UK

⁹USDA Agricultural Research Service, U.S. Meat Animal Research Center, Clay Center, NE 68933, USA

¹⁰Natural History Museum, University of Oslo, Oslo, Norway

†These authors contributed equally.

*Corresponding authors: E-mails: aurelien.dejode@laposte.net; erica.leder@gu.se.

Accepted: March 29, 2024

Abstract

The intertidal gastropod *Littorina saxatilis* is a model system to study speciation and local adaptation. The repeated occurrence of distinct ecotypes showing different levels of genetic divergence makes *L. saxatilis* particularly suited to study different stages of the speciation continuum in the same lineage. A major finding is the presence of several large chromosomal inversions associated with the divergence of ecotypes and, specifically, the species offers a system to study the role of inversions in this divergence. The genome of *L. saxatilis* is 1.35 Gb and composed of 17 chromosomes. The first reference genome of the species was assembled using Illumina data, was highly fragmented (N50 of 44 kb), and was quite incomplete, with a BUSCO completeness of 80.1% on the Metazoan dataset. A linkage map of one full-sibling family enabled the placement of 587 Mbp of the genome into 17 linkage groups corresponding to the haploid number of chromosomes, but the fragmented nature of this reference genome limited the understanding of the interplay between divergent selection and gene flow during ecotype formation. Here, we present a newly generated reference genome that is highly contiguous, with a N50 of 67 Mb and 90.4% of the total assembly length placed in 17 super-scaffolds. It is also highly complete with a BUSCO completeness of 94.1% of the Metazoa dataset. This new reference will allow for investigations into the genomic regions implicated in ecotype formation as well as better characterization of the inversions and their role in speciation.

Key words: chromosome-level assembly, speciation, gastropod, *Littorina*, chromosomal inversion, ecotype formation.

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Significance

The rough periwinkle, *Littorina saxatilis*, has become a model to study adaptation, evolutionary innovation, and speciation, including the role of chromosomal inversions in these processes. Chromosomal inversions have also been identified in other species of *Littorina* providing a valuable opportunity for investigating their origin and evolution. Here, we present a new chromosome-scale reference genome of *L. saxatilis* generated from long reads and Hi-C chromatin proximity mapping replacing an earlier highly fragmented genome. This will enable detailed investigations of how inversions contribute to adaptation and reproductive isolation in *L. saxatilis* and related taxa, including studies of the effects of individual loci within and outside inversions.

Introduction

Speciation and local adaptation are key evolutionary processes that have been successfully studied in the marine snail, *Littorina saxatilis* (Fig. 1A), and related taxa (Johannesson et al. 2010; 2017; Rolán-Alvarez et al. 2015; Johannesson 2016; Blakeslee et al. 2021). Multiple ecotypes have been described in this species and its close relatives (Reid 1996; Johannesson et al. 2024). In particular, two ecotypes are repeatedly found in close proximity on the North Atlantic shores from Portugal to Norway and have been studied in detail. The “crab” ecotype is typically found in areas with intense crab predation while the “wave” ecotype is present in shore areas battered by waves. Their distributions often partially overlap in contact zones. The two ecotypes differ in several traits such as shell size, shape, thickness, ornamentation, and behavior. Across the species distribution, gene flow is occurring at different rates between the two ecotypes. In Sweden, allelic frequencies present a clinal pattern from crab to wave habitat (Westram et al. 2021), while in Spain, two distinct genetic clusters are found in upper and lower shore areas (Raffini et al. 2023). This particular configuration enables the characterization of genomic, phenotypic, and organismal differences between pairs of populations from the same species at various stages of divergence, making *L. saxatilis* a most appropriate system to study the speciation process (Johannesson et al. 2024). Ecotypes also form in other species, providing opportunities for comparative studies of divergence mechanisms for population pairs along the speciation continuum (Johannesson et al. 2024).

Moreover, several genomic inversions showing allelic frequency differences between ecotypes (Faria et al. 2019a; Westram et al. 2021), sex association (Hearn et al. 2022), and association with adaptive traits (Koch et al. 2021; 2022) have been identified in *L. saxatilis* and other species of the genus (Le Moan et al. 2024; Reeve et al. 2023). Inversions can promote the process of local adaptation and speciation and are associated with ecotype divergence in several other natural systems, e.g. (Faria et al. 2019b; Merot 2020).

Furthermore, *L. saxatilis* and its near relatives have direct development, i.e. no planktonic stage (Reid 1996), resulting in low dispersal; however, *L. saxatilis* is ovoviviparous while

the other species lay eggs on the substrate. The shift in reproductive mode compared to other members of the genus provides an outstanding opportunity to study the evolution of this key innovation (Stankowski et al. 2024).

Therefore, *L. saxatilis* and the whole *Littorina* genus constitute a particularly interesting system to study the importance of chromosomal inversions in local adaptation and reproductive isolation. However, for this, a contiguous reference genome is needed.

The *L. saxatilis* genome is composed of 17 chromosomes for a total haploid size of around 1.35 Gbp (Janson 1983; Birstein and Mikhailova 1990; Rolán-Alvarez et al. 1996). The first reference genome of *L. saxatilis* (hereafter “*L. saxatilis* genome v.1”) was 1.6 Gbp in length and composed of 388,619 contigs placed in 116,262 scaffolds (Westram et al. 2018). This genome was highly fragmented with a N50 of ~44 kbp, with only 10% of its contigs (representing 50% of its total length) placed into 17 linkage groups by linkage mapping, with a BUSCO completeness of 80.1% against the Metazoa reference (Westram et al. 2018). This has constituted an obstacle to the precise characterization of the genomic architecture of, for example, the divergence between ecotypes. Such characterization is necessary to properly address questions regarding speciation, local adaptation, and to achieve a detailed understanding of the involvement of inversions in these processes.

Here, we present the first chromosome scale, annotated genome of *L. saxatilis* (hereafter “*L. saxatilis* genome v.2”). This new reference constitutes a major improvement compared to the existing one and will enable us to further enhance our understanding of evolution in this system.

Results and Discussion

A total of 24,891,710 PacBio CLR reads were generated representing 174,323,538,425 base pairs. Assembling with CANU (Koren et al. 2017) produced 23,755 contigs for a total length of 2,278,220,962 bases. After two rounds of haplotig removal with purge_dups (Guan et al. 2020) and long read polishing, the assembly contained 9,667 contigs for a total length of 1,261,606,186 bases.

While BUSCO (Manni et al. 2021) completeness remained at 95.2% for the Metazoa dataset after the two

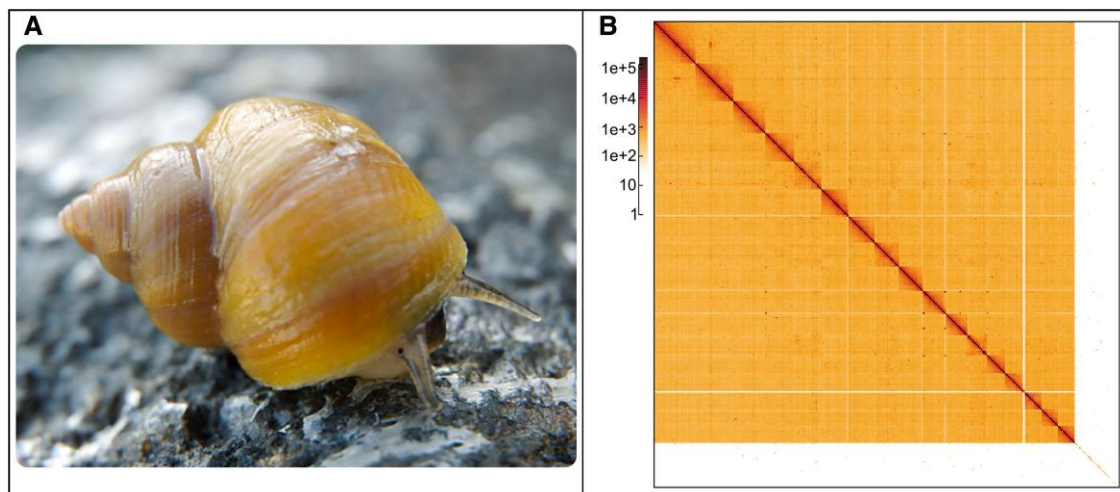


Fig. 1.—*Littorina saxatilis* “crab” ecotype photograph and Hi-C map. A) Photograph of *Littorina saxatilis* “crab” ecotype on Swedish shore (Photo: Patrik Larsson). B) Hi-C contact map of the new *Littorina saxatilis* assembly after manual curation, visualized in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom.

rounds of haplotig removal with `purge_dups`, the “complete and duplicated” score went from 72% to 2.9%. Merqury (Rhie et al. 2020) copy number spectra (`spectra_cn`) plots also showed a clear decrease in the amount of haplotypic duplication retained in the assembly (supplementary fig. S1, Supplementary Material online). Some level of haplotypic duplication could still be seen on the spectra graphs, but neither an additional round of `purge_dups` nor parameter tuning could remove the remaining haplotypic duplication. The high levels of haplotypic duplication obtained in the raw assembly are most likely due to the high heterozygosity of our genome (~1.5% based on k-mer analysis using GenomeScope (Vurture et al. 2017)) combined with potential polymorphic inversions detected in the *L. saxatilis* genome being heterokaryotypic in the individual we sequenced. The final k-mer spectra (supplementary fig. S1, Supplementary Material online) showed a high number of k-mers only found in the assembly and a high frequency of k-mers found only in the reads set (peak of the black line around 30x). This is likely due to the fact that the Illumina reads used for the k-mer analysis came from a different individual than the individual used to assemble the genome.

Using Blobtoolkit (Laetsch and Blaxter 2017; Challis et al. 2020), a total of 55 contigs identified as bacteria or Ciliophora (known commensals of *L. saxatilis* (Fenchel 1965)) were removed from the assembly prior to scaffolding.

A total of 280 million reads was obtained for Hi-C sequencing. The scaffolding produced a total of 4,322 scaffolds, and the scaffolded assembly had a N50 of around 82 Mb. Several errors made by the automatic scaffolding procedure implemented in YaHS (Zhou et al. 2023) were corrected by manual curation (Fig. 1B, supplementary fig. S2, Supplementary Material online).

The final assembly was composed of 4,070 scaffolds including 17 super-scaffolds (composed of contigs joined by gaps with a minimum arbitrary size of 200 bp) that contained 90.4% of the total base pairs in the assembly (Fig. 2A, supplementary fig. S3, Supplementary Material online). The hexamer telomere motif of TTAGGG was found on 16 scaffolds, but given the noisy nature of CLR long reads, the assembly of these regions is likely inaccurate and under-representing the length of these regions. The generated assembly was highly contiguous with an N50 of 67 Mb and a completeness BUSCO score of 98.4%, 94.1%, and 79.2% for the Eukaryota, Metazoa, and Mollusca databases, respectively (Fig. 2B, supplementary table S1, Supplementary Material online). The 17 super-scaffolds corresponded to the number of expected chromosomes for *L. saxatilis* (Janson 1983, Birstein and Mikhailova 1990, Rolán-Alvarez et al. 1996, Faria et al. 2019a) and were matched to the linkage groups from the *L. saxatilis* genome v.1 (supplementary table S2, Supplementary Material online). Using RepeatMasker (v 4.1.5) in quick search mode, simple repeats were identified comprising 7.81% of the genome, while low complexity regions comprised 0.89% of the genome. Additionally, 199 small RNAs were identified.

There were 25,144 protein coding genes predicted by the Braker3 pipeline (Gabriel et al. 2023). The majority of these genes had only one protein predicted, but 3,486 had two or more protein variants. The number of predicted proteins is slightly higher, but still quite similar to other gastropods: 21,438 *Gigantopelta aegis* (Lan et al. 2021), 21,533 *Pomacea canaliculata* (Liu et al. 2018), and 23,800 *Lottia gigantea* (Simakov et al. 2013). BUSCO scores of completeness for the predicted proteins were

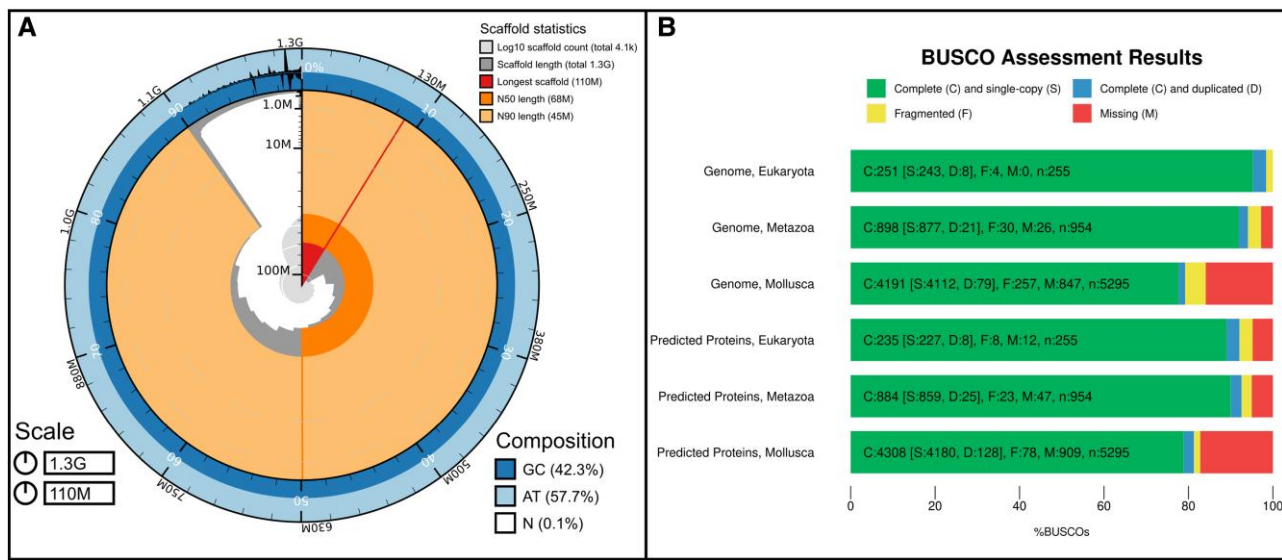


FIG. 2.—Genome assembly of *Littorina saxatilis*: metrics. A) BlobToolKit Snailplot shows N50 metrics. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,256,336,603 bp assembly. The distribution of scaffold lengths is shown in dark gray with the plot radius scaled to the longest scaffold present in the assembly (111,777,080 bp, the smallest of the three arcs, also shown in red). Next two sectors in size (orange and pale-orange arcs) show the N50 and N90 scaffold lengths (67,515,031 and 44,900,717 bp), respectively. The pale gray spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The dark-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. B) BUSCO scores for the genome and predicted proteins for the Eukaryota, Metazoa, and Mollusca datasets.

93.0% and 82.2% for the Metazoa and Mollusca databases, respectively (Fig. 2B, [supplementary table S1](#), [Supplementary Material](#) online).

A mitochondrial genome of *L. saxatilis* was previously assembled using the Illumina data (Marques et al. 2017) and is available on GenBank (NC_030595.1).

Conclusion

We combined PacBio CLR reads and Hi-C chromatin proximity mapping data to generate the first chromosome-level genome assembly for *L. saxatilis*. The *L. saxatilis* genome v.2 achieved a scaffold N50 of 67 Mb with more than 90% of the total assembly length (1.2 Gb) placed in 17 super-scaffolds. Moreover, this genome is highly contiguous with a 98% completeness BUSCO score on the Eukaryota dataset and contained 25,144 protein coding genes. The *L. saxatilis* genome v.2 constitutes a major improvement compared to the *L. saxatilis* genome v.1 and will underpin significant advances in the study of speciation and evolution, specifically the evolution of chromosomal inversions and their role in diversification within the *Littorina* genus.

Materials and Methods

High molecular weight DNA was extracted from fresh head and foot tissue of one female “crab” ecotype of *L. saxatilis* collected in Sweden (N58°52′28″, E11°6′59″) following the

protocol developed by Grohme et al. (2018) with minor adjustments. The protocol consisted of three parts: mucus removal using N-acetyl-L-cysteine, gDNA isolation using a phenol–chloroform–isoamyl alcohol solution and post-purification of gDNA using CTAB (see full protocol in [Supplementary Material](#)). The library was prepared with PacBio’s SMRTbell Express Template Prep Kit 2.0, and sequenced using the Sequel Binding Kit 3.0, Sequel Sequencing Plate 3.0, and Sequel DNA Internal Control 3.0. A total of 16 PacBio ZMW cells were sequenced using a Sequel instrument. Raw PacBio BAM files were converted to fastq files using bam2fastq v1.0.0 (<https://github.com/jts/bam2fastq>) from the SMRT suite. CLR PacBio raw reads were assembled using CANU v2.0 (Koren et al. 2017) with default parameters. Haplotypic duplications were removed from the raw assembly using the purge_dups v1.2.5 (Guan et al. 2020) pipeline two times in a row with default parameters. We used the short Illumina reads that were assembled to generate *L. saxatilis* genome v.1 (Westram et al. 2018) in the Merqury, Merfin, and Nextpolish analysis described hereafter. Levels of haplotypic duplication at the different steps were assessed using spectra-cn plot from Merqury (Rhie et al. 2020) based on Illumina reads. The obtained genome was polished first using the raw CLR long reads with the arrow_grid wrapper (Chin et al. 2013; Koren et al. 2017). The vcf file obtained was corrected using Merfin v1.1-development (Formenti et al. 2022a) based on Illumina reads. This polishing procedure was run two times.

Finally, raw Illumina reads were trimmed and filtered using fastp (Chen et al. 2018) with the following parameters --cut_front --cut_front_window_size 1 --cut_front_mean_quality 30 --cut_right --cut_right_mean_quality 3. The obtained quality filtered reads were used in a final round of polishing implemented in NextPolish v1.3.1 (Hu et al. 2020). Prior to scaffolding, contigs were examined for contaminants using the Blobtoolkit v2.6.4 pipeline (Laetsch and Blaxter 2017; Challis et al. 2020) and 55 of the contigs were removed from the assembly.

A different “crab” ecotype female snail from the same location in Sweden was used to build a Dovetail Hi-C library using the DpnII enzyme. Paired end reads (2 × 150 bp) were generated at SciLife on one lane of a NovaSeq 6000 SP instrument. Raw Hi-C reads were mapped to the contigs following the Arima mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline/tree/master), and contigs were placed into scaffolds using YaHS v1.1a-r3 (Zhou et al. 2023). The obtained scaffolded genome was further manually curated (Howe et al. 2021) using pretext (<https://github.com/wtsi-hpag/PretextView>), and HiGlass (Kerpedjiev et al. 2018).

Finally, assembly contiguity and completeness were assessed using gfastats v1.3.6 v5.0.2 (Formenti et al. 2022b) (supplementary table S3, Supplementary Material online), BUSCO v5.4.7 (Manni et al. 2021) (supplementary table S2, Supplementary Material online), and the Blobtoolkit v2.6.4 pipeline (Laetsch and Blaxter 2017; Challis et al. 2020) (supplementary fig. S3, Supplementary Material online).

To establish correspondence between the super-scaffolds from *L. saxatilis* genome v.1 and the linkage groups of *L. saxatilis* genome v.2, the following procedure was followed. For each SNP placed in the linkage map from the *L. saxatilis* genome v.1, a 200 bp sequence centered on the SNP was extracted and mapped to the *L. saxatilis* genome v.2 using minimap2 (Li 2018). Alignments were filtered by removing duplicates and partial mappings that did not include the SNP. Then for each 200 bp fragment, the best alignment was selected. RepeatMasker (v 4.1.5) was used to identify simple repeats using the quick search option (<http://www.repeatmasker.org>).

RNAseq data from GenBank BioProject PRJNA550990 along with RNAseq from mantle tissue from three “crab” ecotype females from Sweden were mapped to the genome assembly using HiSat2 v2.2.1 (Kim et al. 2019). Reads were sorted and filtered to include only mapped reads using SAMtools v 1.16.1 (Li et al. 2009). BRAKER3 (Gabriel et al. 2023) and the algorithms therein (Stanke et al. 2006; Gotoh 2008; Iwata and Gotoh 2012; Hoff et al. 2016; Hoff et al. 2019; Brūna et al. 2021, and Stanke et al. 2008) were used to predict protein coding genes in the assembly using both the RNAseq data from *L. saxatilis* and mollusk proteins from OrthoDBv11 (Kuznetsov et al.

2023). In addition to the algorithms employed by BRAKER3, the pipeline also uses various tools to gather evidence for protein sequences from the RNAseq data and the protein data using the following tools: DIAMOND (Buchfink et al. 2015), Stringtie2 (Kovaka et al. 2019), GFF utilities (Pertea and Pertea 2020), BamTools (Barnett et al. 2011), and BEDTools (Quinlan 2014). The protein evidence obtained was used for training GeneMark-EPT (Brūna et al. 2023) and later, AUGUSTUS (Stanke et al. 2006), and then the two sets of predictions were combined using TSEBRA (Gabriel et al. 2023). The Braker3 Docker container version v.1.0.4 was used for this analysis. BLASTP was then used to identify the predicted proteins using a custom database of mollusk proteins (taxid 6447) from OrthoDBv11 (Kuznetsov et al. 2023) (supplementary table S4, Supplementary Material online).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

The computations and data handling were enabled by resources in project NAISS 2023/6-270 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX, funded by the Swedish Research Council through grant agreement no. 2022-06725. We are very grateful to Sergey Koren for his help in getting CANU running on the cluster. Sequencing was performed at the Genomics Laboratory Facility, School of Biosciences, University of Sheffield. We thank Rachel Tucker and Helen Hipperson for their valuable input. The computations for the annotation were performed on resources provided by Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway. We also thank Alan Le Moan for proving the correspondence table between *L. saxatilis* genome v1 and *L. saxatilis* genome v2. This work was funded by the European Research Council (ERC-2015-AdG-693030- BARRIERS, R.K.B.), and the Swedish Research Council VR (2018-03695, R.K.B.; 2021-04191, K.J.; 2020-05385, E.H.L.). R.F. was funded by the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie (grant agreement No. 706376) and by the Fundação para a Ciência e a Tecnologia (2020.00275.CEECIND and PTDC/BIA-EVL/1614/2021). The use of trade names or commercial products in this manuscript is solely to provide specific information. It does not imply recommendation or endorsement by the U.S. Department of Agriculture. The USDA is an equal opportunity provider and employer.

Data Availability

The PacBio CLR sequencing reads, the sequences from the Hi-C library, and the Illumina short reads are deposited in NCBI under accession number PRJNA850123. The transcriptome short-read sequences were deposited under accession number PRJNA550990. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JBAMIC000000000. The version described in this paper is version JBAMIC010000000.

Literature Cited

- Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*. 2011;27(12):1691–1692. <https://doi.org/10.1093/bioinformatics/btr174>.
- Birstein V, Mikhailova N. On the karyology of trematodes of the genus *Microphallus* and their intermediate gastropod host, *Littorina saxatilis* I. Chromosome analysis of three *Microphallus* species. *Genetica*. 1990;80(3):159–165. <https://doi.org/10.1007/BF00137320>.
- Blakeslee AMH, Miller AW, Ruiz GM, Johannesson K, André C, Panova M. Population structure and phylogeography of two North Atlantic *Littorina* species with contrasting larval development. *Mar Biol*. 2021;168(7):117. <https://doi.org/10.1007/s00227-021-03918-8>.
- Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. 2021;3(1):lqaa108. <https://doi.org/10.1093/nargab/lqaa108>.
- Brůna T, Lomsadze A, Borodovsky M. GeneMark-ETP: automatic gene finding in eukaryotic genomes in consistence with extrinsic data. *BioRxiv* 524024. <https://doi.org/10.1101/2023.01.13.524024>, 3 January 2023, preprint: not peer reviewed
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60. <https://doi.org/10.1038/nmeth.3176>.
- Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. *G3 (Bethesda)*. 2020;10(4):1361–1374. <https://doi.org/10.1534/g3.119.400908>.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013;10(6):563–569. <https://doi.org/10.1038/nmeth.2474>.
- Faria R, Chaube P, Morales HE, Larsson T, Lemmon AR, Lemmon EM, Rafajlović M, Panova M, Ravinet M, Johannesson K, et al. Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Mol Ecol*. 2019a;28(6):1375–1393. <https://doi.org/10.1111/mec.14972>.
- Faria R, Johannesson K, Butlin RK, Westram AM. Evolving inversions. *Trends Ecol Evol*. 2019b;34(3):239–248. <https://doi.org/10.1016/j.tree.2018.12.005>.
- Fenchel T. Ciliates from Scandinavian molluscs. *Ophelia*. 1965;2(1):71–174. <https://doi.org/10.1080/00785326.1965.10409598>.
- Formenti G, Rhie A, Walenz BP, Thibaud-Nissen F, Shafin K, Koren S, Myers EW, Jarvis ED, Phillippy AM. Merfin: improved variant filtering, assembly evaluation and polishing via k-mer validation. *Nat Methods*. 2022a;19(6):696–704. <https://doi.org/10.1038/s41592-022-01445-y>.
- Formenti G, Abueg L, Brajuka A, Brajuka N, Gallardo-Alba C, Giani A, Fedrigo O, Jarvis ED. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*. 2022b;38(17):4214–4216. <https://doi.org/10.1093/bioinformatics/btac460>.
- Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *bioRxiv* 544449. <https://doi.org/10.1101/2023.06.10.544449>, 3 September 2023, preprint: not peer reviewed
- Gotoh O. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. *Nucleic Acids Res*. 2008;36(8):2630–2638. <https://doi.org/10.1093/nar/gkn105>.
- Grohme MA, Vila-Farré M, Rink JC. Small- and large-scale high molecular weight genomic DNA extraction from planarians. *Methods Mol Biol*. 2018;1774:267–275. https://doi.org/10.1007/978-1-4939-7802-1_7.
- Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. 2020;36(9):2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>.
- Hearn KE, Koch EL, Stankowski S, Butlin RK, Faria R, Johannesson K, Westram AM. Differing associations between sex determination and sex-linked inversions in two ecotypes of *Littorina saxatilis*. *Evol Lett*. 2022;6(5):358–374. <https://doi.org/10.1002/evl3.295>.
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*. 2016;32(5):767–769. <https://doi.org/10.1093/bioinformatics/btv661>.
- Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. *Methods Mol Biol*. 2019;1962:65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.
- Howe K, Chow W, Collins J, Pelan S, Pointon D-L, Sims Y, Torrance J, Tracey A, Wood J. Significantly improving the quality of genome assemblies through curation. *Gigascience*. 2021;10(1):giaa153. <https://doi.org/10.1093/gigascience/giaa153>.
- Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*. 2020;36(7):2253–2255. <https://doi.org/10.1093/bioinformatics/btz891>.
- Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res*. 2012;40(20):e161–e161. <https://doi.org/10.1093/nar/gks708>.
- Janson K. Chromosome number in two phenotypically distinct populations of *Littorina saxatilis* Olivii, and in specimens of the *Littorina obtusata* (L.) species-complex. *J. Molluscan Stud*. 1983;49(3):224–227. <https://doi.org/10.1093/oxfordjournals.mollus.a065716>.
- Johannesson K. What can be learnt from a snail? *Evol Appl*. 2016;9(1):153–165. <https://doi.org/10.1111/eva.12277>.
- Johannesson K, Butlin RK, Panova M, Westram AM. Mechanisms of adaptive divergence and speciation in *Littorina saxatilis*: integrating knowledge from ecology and genetics with new data emerging from genomic studies. In: Oleksiak M Rajora O, editors. *Population genomics: marine organisms*. Population genomics. Cham: Springer; 2017. p. 277–301.
- Johannesson K, Faria R, Le Moan A, Rafajlović M, Westram AM, Butlin RK, Stankowski S. Diverse pathways to speciation revealed by marine snails. *Trends Genet*. 2024;40(4):337–351. <https://doi.org/10.1016/j.tig.2024.01.002>.
- Johannesson K, Panova M, Kempainen P, André C, Rolán-Alvarez E, Butlin RK. Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Philos Trans R Soc Lond B Biol Sci*. 2010;365(1547):1735–1747. <https://doi.org/10.1098/rstb.2009.0256>.

- Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Luber JM, Ouellette SB, Azhir A, Kumar N, et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* 2018;19(1):125. <https://doi.org/10.1186/s13059-018-1486-1>.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
- Koch EL, Morales HE, Larsson J, Westram AM, Faria R, Lemmon AR, Lemmon EM, Johannesson K, Butlin RK. Genetic variation for adaptive traits is associated with polymorphic inversions in *Littorina saxatilis*. *Evol Lett.* 2021;5(3):196–213. <https://doi.org/10.1002/evl3.227>.
- Koch EL, Ravinet M, Westram AM, Johannesson K, Butlin RK. Genetic architecture of repeated phenotypic divergence in *Littorina saxatilis* ecotype evolution. *Evolution.* 2022;76(10):2332–2346. <https://doi.org/10.1111/evo.14602>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27(5):722–736. <https://doi.org/10.1101/gr.215087.116>.
- Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 2019;20(1):1–13. <https://doi.org/10.1186/s13059-019-1910-1>.
- Kuznetsov D, Tegenfeldt F, Manni M, Seppy M, Berkeley M, Kriventseva EV, Zdobnov EM. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* 2023;51(D1):D445–D451. <https://doi.org/10.1093/nar/gkac998>.
- Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. *F1000Res.* 2017;6:1287. <https://doi.org/10.12688/f1000research.12232.1>.
- Lan Y, Sun J, Chen C, Sun Y, Zhou Y, Yang Y, Zhang W, Li R, Zhou K, Wong WC, et al. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nat Commun.* 2021;12(1):1165. <https://doi.org/10.1038/s41467-021-21450-7>.
- Le Moan A, Stankowski S, Rafajlovic M, Martinez Ortega O, Faria R, Butlin R, Johannesson K. Coupling of twelve putative chromosomal inversions maintains a strong barrier to gene flow between snail ecotypes. *Evol. Lett.* 2024. in press.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Liu C, Zhang Yan, Ren Y, Wang H, Li S, Jiang F, Yin L, Qiao X, Zhang G, Qian W, et al. The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. *Gigascience.* 2018;7(9):gij101. <https://doi.org/10.1093/gigascience/gij101>.
- Manni M, Berkeley MR, Seppy M, Simão FA, Zdobnov EM. 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. <https://doi.org/10.1093/molbev/msab199>.
- Marques JP, Sotelo G, Larsson T, Johannesson K, Panova M, Faria R. Comparative mitogenomic analysis of three species of periwinkles: *Littorina fabalis*, *L. obtusata* and *L. saxatilis*. *Mar genomics.* 2017;32:41–47. <https://doi.org/10.1016/j.margen.2016.10.006>.
- Mérot C. Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Mol Ecol.* 2020;29(14):2513–2516. <https://doi.org/10.1111/mec.15500>.
- Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Res.* 2020;9:304. <https://doi.org/10.12688/f1000research.23297.1>.
- Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinformatics.* 2014;47(1):11–12. <https://doi.org/10.1002/0471250953.bi1112s47>.
- Raffini F, De Jode A, Johannesson K, Faria R, Zagrodzka ZB, Westram AM, Galindo J, Rolan-Alvarez E, Butlin RK. Phenotypic divergence and genomic architecture at two different stages of speciation in a marine snail. *bioRxiv* 558824. <https://doi.org/10.1101/2023.09.21.558824>, September 2023, preprint: not peer reviewed
- Reeve J, Butlin RK, Koch EL, Stankowski S, Faria R. Chromosomal inversion polymorphisms are widespread across the species ranges of rough periwinkles (*Littorina saxatilis* and *L. arcana*). *Mol Ecol.* 2023;00:1–18. <https://doi.org/10.1111/mec.17160>.
- Reid D. Systematics and evolution of Littorina. London, UK: Ray Society; 1996.
- Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 2020;21(1):245. <https://doi.org/10.1186/s13059-020-02134-9>.
- Rolán-Alvarez E, Austin CJ, Boulding EG. The contribution of the genus *Littorina* to the field of evolutionary ecology. *Oceanography and marine biology.* Boca Raton: CRC Press; 2015. p. 166–223.
- Rolán-Alvarez E, Buño I, Gosálvez J. Sex is determined by sex chromosomes in *Littorina saxatilis* (Olivier) (Gastropoda, Prosobranchia). *Hereditas.* 1996;124(3):261–268. <https://doi.org/10.1111/j.1601-5223.1996.00261.x>.
- Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D-H, Larsson T, Lv J, Arendt D, et al. Insights into bilaterian evolution from three spiralian genomes. *Nature.* 2013;493(7433):526–531. <https://doi.org/10.1038/nature11696>.
- Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntetically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics.* 2008;24(5):637–644. <https://doi.org/10.1093/bioinformatics/btn013>.
- Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* 2006;7(1):62. <https://doi.org/10.1186/1471-2105-7-62>.
- Stankowski S, Zagrodzka ZB, Garlovsky MD, Pal A, Shipilina D, Castillo DG, Lifchitz H, Le Moan A, Leder E, Reeve J, et al. The genetic basis of a recent transition to live-bearing in *Littorina* snails. *Science.* 2024;383(6678):114–119. <https://doi.org/10.1126/science.adi2982>.
- Vurtture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics.* 2017;33(14):2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>.
- Westram AM, Faria R, Johannesson K, Butlin R. Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Mol Ecol.* 2021;30(15):3797–3814. <https://doi.org/10.1111/mec.15861>.
- Westram AM, Rafajlović M, Chaube P, Faria R, Larsson T, Panova M, Ravinet M, Blomberg A, Mehlig B, Johannesson K, et al. Clines on the seashore: the genomic architecture underlying rapid divergence in the face of gene flow. *Evol Lett.* 2018;2(4):297–309. <https://doi.org/10.1002/evl3.74>.
- Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics.* 2023;39(1):btac808. <https://doi.org/10.1093/bioinformatics/btac808>.

Associate editor: John Wang