

Letter to the editor

Open Access

Chromosome-level genome assembly of the freshwater snail *Bellamya purificata* (Caenogastropoda)

DEAR EDITOR,

Bellamya purificata (Gastropoda: Caenogastropoda; Architaenioglossa; Viviparidae: Bellamyinae; *Sinotaia*), a homotypic synonym of *S. purificata*, is widely distributed in freshwater habitats in Asia. It is an economically important edible snail and plays a vital function in freshwater wetland ecology. However, genomic resources for this snail are lacking and no reference genome has been released. In this study, we constructed the first chromosome-level genome of *B. purificata* using PacBio long-read sequencing and high-throughput chromosome conformation capture (Hi-C) technology. In total, 33.64 Gb of circular consensus sequencing reads were generated. The preliminary genome assembly was 1.01 Gb in size, with a contig N50 of 45.14 Mb. Using Hi-C data, the assembled sequences were anchored to eight pseudochromosomes. After Hi-C correction, the final genome was 984.33 Mb, with a contig N50 of 37.21 Mb and scaffold N50 of 141.97 Mb. The chromosome anchoring rate was 99.28%. A total of 22 125 protein-coding genes were predicted. Phylogenetic analyses indicated that *B. purificata* diverged from *Pomacea canaliculata* approximately 288 million years ago (Ma). We identified 34 expanded and 26 contracted gene families in *B. purificata* compared with its most recent common ancestor. Four protein-coding genes under positive selection in *B. purificata* were identified (false discovery rate (FDR)<0.05). These genomic data provide a valuable resource for ecological and evolutionary studies of the family Viviparidae, and for genetic improvement of *B. purificata*.

Bellamya purificata is the largest snail species in Bellamyinae, Viviparidae, Caenogastropoda, and is widely distributed in freshwater habitats across Asia. It feeds on microalgae and organic detritus, serving as an important consumer in wetland ecosystems (Zhao et al., 2014). During

gill respiration, *B. purificata* can capture *Microcystis* particles in water to form mucus-encapsulated *Microcystis* clusters. This behavior decreases *Microcystis* biomass suspended in water for a short period and temporarily reduces bloom states (Qu et al., 2010). In addition, *B. purificata* can scrape and consume suspended particles adsorbed on submerged plants (Li et al., 2007), thereby reducing total nitrogen and ammonia nitrogen levels in water (Zhao et al., 2014).

Bellamya purificata is also an economically important edible snail in Asia (Lydeard & Cummings, 2019). Over the past several years, “snail rice noodle” dishes (i.e., rice noodles with snails) have become a “net red food” (i.e., popular online) in China, creating more than 10 billion yuan in economic benefits each year. The annual demand for freshwater Viviparidae snails in China is about 1.5 million tons (Jin et al., 2022). However, according to the 2021 China Fishery Statistical Yearbook, annual yield of cultured Viviparidae snails in China is only 90 640 tons, indicating a huge gap between supply and demand. To date, most cultured freshwater snails are sourced from slow-growing wild populations. Therefore, to promote the development of snail aquaculture, freshwater snail breeding programs, especially for *B. purificata*, have become a priority.

Currently, genomic resources for *B. purificata* are lacking, with only a single transcriptome and proteome analysis performed to screen for shell color-related genes (Huang et al., 2021). However, a detailed reference genome of *B. purificata* is required for ecological and evolutionary research and genetic improvement. Third-generation sequencing technologies can expand sequencing reads and provide superior platforms to produce complete and high-quality genomes. In this study, PacBio long-read sequencing and high-throughput chromosome conformation capture (Hi-C) technology were used to assemble a high-quality chromosome-level genome of *B. purificata*. Comparative genomic analysis of *B. purificata* with eight other mollusk

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2022 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 27 May 2022; Accepted: 14 July 2022; Online: 15 July 2022

Foundation items: This work was supported by the Central Public-Interest Scientific Institution Basal Research Fund, CAFS (2021XT703) and Central Public-interest Scientific Institution Basal Research Fund, Freshwater Fisheries Research Center, CAFS (2021JBFM07)

species was performed to explore *B. purificata* evolution in Asia.

To estimate the genome size of *B. purificata*, we performed *k*-mer (*k*=17) frequency distribution analysis using 181.6 Gb of clean data (Figure 1A). In this process, 17 bp *k*-mers (17-mer) were extracted from the sequencing data and 17-mer frequency was calculated. Here, 17-mer analysis conformed to

Poisson distribution, with an estimated genome size of 939 Mb.

PacBio sequencing data were obtained from the PacBio Sequel II platform (Supplementary Table S1). Data were assembled using Hifiasm (v0.16.1), then polished. Redundancies and haplotigs were removed, producing an assembly with a genome size of 1 006 Mb. Genome length

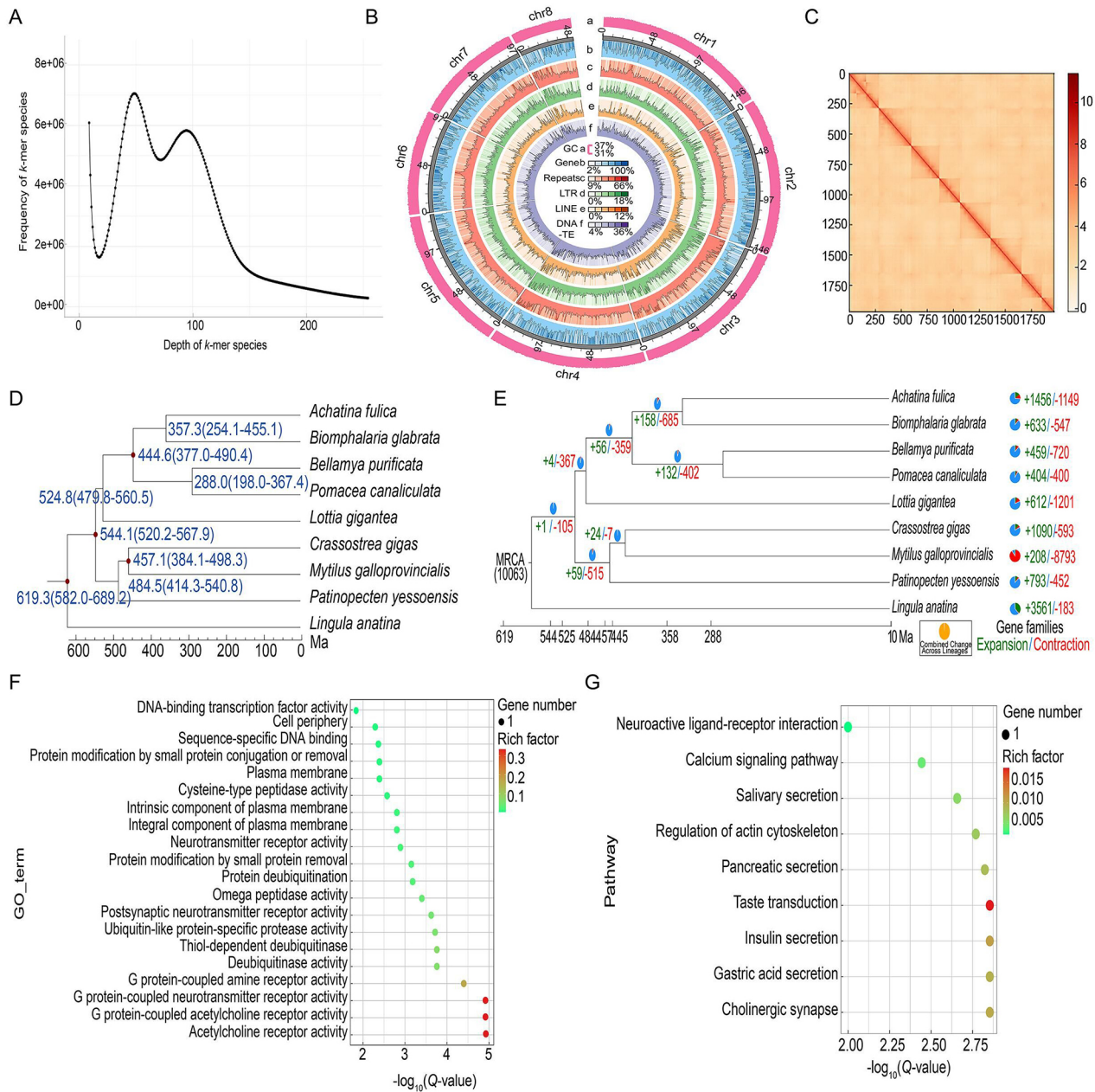


Figure 1 Characterization of chromosome-level genome of *Bellamyia purificata*

A: Frequency distribution of *k*-mer depth and *k*-mer species. B: Genome characteristics of *B. purificata*. From outer to inner circle, (a) GC content of genome, (b) gene distribution, (c) short read depth, (d) long tandem repeats (LTRs), (e) long read depth, and (f) DNA transposable element (TE). Bar height is proportional to number of items mapped to each genomic position. C: Genome-wide Hi-C heatmap of *B. purificata*. Blocks represent eight pseudochromosomes. Color bar represents contact density from white (low) to red (high). D: Estimates of species divergence times. E: Number of expanded and contracted gene families in *B. purificata*. F: GO enrichment analysis of positively selected genes. G: KEGG enrichment analysis of positively selected genes.

was consistent with that estimated by *k*-mer analysis. The total number of contigs was 265, with N50 reaching 45.14 Mb. The genome sizes of three other gastropod species (*Biomphalaria glabrata*, *Achatina fulica*, and *Lottia gigantea*) range from 359 Mb to 2.12 Gb (Adema et al., 2017; Guo et al., 2019; Simakov et al., 2013). In total, 100 786 genomic fragments (based on a step length of 1 kb) were randomly selected and mapped to the Non-Redundant (NT) database, with more than 80% of these fragments aligned to mollusk genomes. Based on Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis, 100% (954/954) of complete BUSCO genes were found in the assembly (Supplementary Table S2), including 94.9% complete and single-copy BUSCOs and 1.9% complete and duplicated BUSCOs. The Circos plot is shown in Figure 1B. Overall, the above results indicated that the genome assembly of *B. purificata* was of high quality.

The genome contigs were further anchored and oriented to chromosomes by Hi-C scaffolding. The Hi-C library generated 124.37 Gb of clean data, with 88.25% validly paired. Using LACHESIS software (v0.1.19), 99.28% of the assembled sequences were anchored to eight pseudochromosomes. The eight pseudochromosomes were clearly distinguished from the Hi-C heatmap and interactions within the pseudochromosomes were strong (Figure 1C), indicating high-quality anchoring. The final assembly yielded a high-quality genome of 984.33 Mb, with contig N50 of 37.21 Mb and scaffold N50 of 141.97 Mb (Supplementary Table S3).

A total of 482.49 Mb of repeat sequences were annotated, accounting for 47.93% of the total genome (Supplementary Table S4). This percentage is approximately the same as that of the genome survey. The major repetitive elements were DNA (20.97%), long terminal repeats (LTRs; 8.48%), and long interspersed nuclear elements (LINEs; 5.38%) (Supplementary Table S5). *De novo* and homology-based methods predicted 22 125 protein-coding genes. Average gene length, average CDS length, average exon length, average intron length, and average exon number per gene were 23 810 bp, 1 547 bp, 400.32 bp, 2 739 bp, and 8.46, respectively (Supplementary Table S6). BUSCO evaluation of the predicted genes revealed 923 orthologous genes, including 905 complete single-copy BUSCOs and 18 complete duplicated BUSCOs (Supplementary Table S2). A total of 18 958 genes, accounting for 85.69% of all predicted genes, were annotated using public databases (Supplementary Table S7). For non-coding RNA predictions, we successfully annotated 39 microRNAs (miRNAs), 202 transfer RNAs (tRNAs), 77 ribosomal RNAs (rRNAs), and 145 small non-coding RNAs (snRNAs), with average lengths of 82, 76, 333, and 142 bp, respectively (Supplementary Table S8).

Comparison of the *B. purificata* genome with that of eight other mollusk species (*Pomacea canaliculata*, Gastropoda; *Biomphalaria glabrata*, Gastropoda; *A. fulica*, Gastropoda; *Crassostrea gigas*, Bivalvia; *Lingula anatina*, Lingulata; *Lottia gigantea*, Gastropoda; *Mytilus galloprovincialis*, Bivalvia; *Patinopecten yessoensis*, Bivalvia) revealed 8 943 gene families and 990 single-copy genes. The *B. purificata* genome contained a total of 22 125 genes clustered into 17 659 gene families, including 1 857 unique families. Average gene number per family ranged from 1.385 (*B. purificata*) to 4.048

(*M. galloprovincialis*) for the nine species (Supplementary Table S9).

Based on the protein sequences of the single-copy genes, we constructed a phylogenetic tree, which showed that the divergence time between *B. purificata* and other shellfish species was about 465.0 (288.0–619.3) Ma (Figure 1D), consistent with previous studies (Sun et al., 2020). In addition, we identified 34 expanded gene families (339 genes) and 26 contracted gene families (72 genes) by comparing the genome of *B. purificata* with its most recent common ancestor (Figure 1E). Four protein-coding genes under positive selection were identified in *B. purificata* (FDR<0.05, Supplementary Table S10). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the positively selected genes showed enrichment in G protein-coupled protein receptor activity, acetylcholine receptor activity, and taste transduction function (Figure 1F, G).

In summary, we assembled the first chromosome-level genome of *B. purificata* by integrating PacBio long reads and Hi-C data. The assembled genome was 984.33 Mb, similar to the estimated size. Due to the intrinsic long length of PacBio sequencing, contig N50 of the assembled genome was 37.21 Mb and scaffold N50 was 141.97 Mb. This genome assembly and analysis provide important data for further study of Viviparidae species, laying a solid foundation for a range of breeding, conservation, and phylogenetic studies of *Bellamyia* in the future.

DATA AVAILABILITY

The whole-genome assembly of *B. purificata* was submitted to NCBI, GSA, and Science Data Bank under PRJNA818874, CRA007090, and CSTR:31253.11.sciencedb.01820, respectively.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

W.J. and X.J.C. conceived and managed the project. X.Y. M., G.H.L., G.C.X., P.X., and B.S. collected the sequencing samples. W.J. and X.J.C. wrote the manuscript. D.P.X. and H.B.W. contributed to revision of the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We are grateful for the constructive comments from three anonymous reviewers.

Wu Jin^{1,2,#}, Xiao-Juan Cao^{3,#}, Xue-Yan Ma^{1,2}, Guo-Hua Lv¹, Gang-Chun Xu^{1,2}, Pao Xu^{1,2}, Bing Sun³, Dong-Po Xu^{1,2,*}, Hai-Bo Wen^{1,2,*}

¹ Key Laboratory of Integrated Rice-Fish Farming Ecology,

Ministry of Agriculture and Rural Affairs, Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi, Jiangsu 214081, China

² Wuxi Fisheries College, Nanjing Agricultural University, Wuxi, Jiangsu 214128, China

³ College of Fisheries, Engineering Research Center of Green Development for Conventional Aquatic Biological Industry in the Yangtze River Economic Belt, Ministry of Education, Huazhong Agricultural University, Wuhan, Hubei 430070, China

*Authors contributed equally to this work

*Corresponding authors, E-mail: xudp@ffrc.cn; wenhb@ffrc.cn

REFERENCES

- Adema CM, Hillier LW, Jones CS, Loker ES, Knight M, Minx P, et al. 2017. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nature Communications*, **8**: 15451.
- Guo YH, Zhang Y, Liu Q, Huang Y, Mao GY, Yue ZY, et al. 2019. A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *GigaScience*, **8**(10): giz124.
- Huang SQ, Jiang HJ, Zhang L, Gu QH, Wang WM, Wen YH, et al. 2021. Integrated proteomic and transcriptomic analysis reveals that polymorphic shell colors vary with melanin synthesis in *Bellamyia purificata* snail. *Journal of Proteomics*, **230**: 103950.
- Jin W, Cao JY, Ma C, Ma XY, Lv GH, Wen HB, et al. 2022. Genetic diversity and genetic differentiation analysis of *Bellamyia purificata* in eleven populations based on the microsatellite makers. *Freshwater Fisheries*, **52**(2): 16–21. (in Chinese)
- Li KY, Wen MZ, Yang HW, Li YM, Liu ZW. 2007. Mutualistic relationship between freshwater snails and aquatic macrophytes. *Acta Ecologica Sinica*, **27**(12): 5427–5432. (in Chinese)
- Lydeard C, Cummings KS. 2019. *Freshwater Mollusks of the World: a Distribution Atlas*. Baltimore: Johns Hopkins University Press.
- Qu MZ, Qu YF, Ren WW, Wang YX, Wu QH. 2010. The mechanism of controlling *Microcystis* bloom by *Bellamyia Aeruginosa*. *Journal of Fudan University (Natural Science)*, **49**(3): 301–308. (in Chinese)
- Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, et al. 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature*, **493**(7433): 526–531.
- Sun J, Chen C, Miyamoto N, Li RS, Sigwart JD, Xu T, et al. 2020. The scaly-foot snail genome and implications for the origins of biomineralised armour. *Nature Communications*, **11**(1): 1657.
- Zhao F, Xie CX, Zhang N, Lv YJ, Li RJ. 2014. Efficiency of water purification and the nitrogen and phosphorous release of the sediment by different densities of *Bellamyia purificata*. *Journal of Hydroecology*, **35**(2): 32–38. (in Chinese)