




## TECHNICAL NOTE

# A chromosomal-level genome assembly for the giant African snail *Achatina fulica*

Yunhai Guo<sup>1,†</sup>, Yi Zhang<sup>1,†</sup>, Qin Liu<sup>1</sup>, Yun Huang<sup>1</sup>, Guangyao Mao<sup>1</sup>, Zhiyuan Yue<sup>1</sup>, Eniola M. Abe <sup>1</sup>, Jian Li<sup>2</sup>, Zhongdao Wu<sup>3</sup>, Shizhu Li<sup>1</sup>, Xiaonong Zhou<sup>1</sup>, Wei Hu <sup>1,2,\*</sup> and Ning Xiao <sup>1,\*</sup>

<sup>1</sup>National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention; Key Laboratory of Parasite and Vector Biology, Ministry of Health; WHO Collaborating Centre for Tropical Diseases; Chinese Centre for Tropical Diseases Research, Shanghai 200025, P. R. China; <sup>2</sup> State Key Laboratory of Genetic Engineering, Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Science, Fudan University, Shanghai 200438, China and <sup>3</sup>Department of Parasitology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China

\*Correspondence address. Ning Xiao, National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, 207 Rui Jin Er Road, Shanghai 200025, China. E-mail: [xiaoning@nipd.chinacdc.cn](mailto:xiaoning@nipd.chinacdc.cn)  <http://orcid.org/0000-0002-1361-7013>; Wei Hu, School of Life Sciences, Fudan University, 2005 Songhu Road, Shanghai 200438, China. E-mail: [huw@fudan.edu.cn](mailto:huw@fudan.edu.cn)  <http://orcid.org/0000-0002-4432-5400>

<sup>†</sup>These authors contributed equally to this work.

## Abstract

**Background:** *Achatina fulica*, the giant African snail, is the largest terrestrial mollusk species. Owing to its voracious appetite, wide environmental adaptability, high growth rate, and reproductive capacity, it has become an invasive species across the world, mainly in Southeast Asia, Japan, the western Pacific islands, and China. This pest can damage agricultural crops and is an intermediate host of many parasites that can threaten human health. However, genomic information of *A. fulica* remains limited, hindering genetic and genomic studies for invasion control and management of the species.

**Findings:** Using a k-mer-based method, we estimated the *A. fulica* genome size to be 2.12 Gb, with a high repeat content up to 71%. Roughly 101.6 Gb genomic long-read data of *A. fulica* were generated from the Pacific Biosciences sequencing platform and assembled to produce a first *A. fulica* genome of 1.85 Gb with a contig N50 length of 726 kb. Using contact information from the Hi-C sequencing data, we successfully anchored 99.32% contig sequences into 31 chromosomes, leading to the final contig and scaffold N50 length of 721 kb and 59.6 Mb, respectively. The continuity, completeness, and accuracy were evaluated by genome comparison with other mollusk genomes, BUSCO assessment, and genomic read mapping. A total of 23,726 protein-coding genes were predicted from the assembled genome, among which 96.34% of the genes were functionally annotated. The phylogenetic analysis using whole-genome protein-coding genes revealed that *A. fulica* separated from a common ancestor with *Biomphalaria glabrata* ~182 million years ago. **Conclusion:** To our knowledge, the *A. fulica* genome is the first terrestrial mollusk genome published to date. The chromosome sequence of *A. fulica* will

Received: 8 January 2019; Revised: 9 May 2019; Accepted: 27 September 2019

© The Author(s) 2019. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

provide the research community with a valuable resource for population genetics and environmental adaptation studies for the species, as well as investigations of the chromosome-level of evolution within mollusks.

**Keywords:** giant African snail; *Achatina fulica*; Pacific Biosciences; Hi-C; chromosome assembly

## Data Description

### Introduction

The giant African snail, *Achatina fulica* (NCBI:txid6530), is a Gastropod species (Fig. 1). It is the largest terrestrial mollusk, with a voracious appetite, strong environmental adaptability, and high growth and reproduction rate [1–3]. Originating in East Africa, *A. fulica* over the past century has gradually invaded Southeast Asia, Japan, and the western Pacific islands [4–6] with the direct and indirect help of humans [7–9]. In mainland China, the first *A. fulica* invasion event was reported in 1931 [10]. At present, the snail has been found in the wild in Guangdong, Hainan, Guangxi, southern parts of Yunnan Province and Fujian Province, and a county of Guizhou Province [11]. *A. fulica* was included among the first 16 alien invasive species designated in China [12] in 2003 and was also listed by the International Union for Conservation of Nature as among the 100 most threatening alien invasive species [13]. This snail has been recognized as an agricultural and garden pest causing significant damage in both tropical and subtropical regions [9, 13, 14]. In addition, *A. fulica* is also the intermediate host of the parasitic nematode *Angiostrongylus cantonensis*. Human infection with angiostrongyliasis, which occurs mainly through consumption of snails carrying *A. cantonensis* larvae, causes eosinophilic meningoencephalitis [4, 11, 15–19]. As a consequence, *A. fulica* is attracting more and more attention in the fields of agricultural crop protection and human disease control.

To date, a variety of mollusk genomes have been analyzed and published, including those of 2 freshwater gastropod snails, *Pomacea canaliculata* [20] and *Biomphalaria glabrata* [21]. However, no genome has been reported for terrestrial mollusks. *A. fulica* is considered to be a destructive terrestrial gastropod that poses a significant hazard to agriculture, the environment, biodiversity, and human health. A chromosome-level genome of *A. fulica* could provide crucial resources in population genetics and evolution studies based on genomic sequencing data aiming to elucidate its invasion and adaptation history. Furthermore, the genome could also be used to probe gene expression during important biological processes, such as gene expression patterns in various developmental stages and the interaction of *Angiostrongylus* and *A. fulica*. In the present study we applied Illumina, PacBio, and Hi-C techniques to construct the chromosome of *A. fulica*. The genome is the first terrestrial mollusk genome, providing an important reference for the molecular mechanisms underlying its broad environmental adaptability and the development of a control strategy for its worldwide invasion.

### Sample and sequencing

An adult snail (Fig. 1), which was collected in Pingxiang city, Guangxi Autonomous Region, was used for reference genome construction. The snail was dissected and abdominal foot (17.4 g) and liver pancreas (40.4 g) tissues were collected and quickly frozen in liquid nitrogen overnight before transfer to storage at  $-80^{\circ}\text{C}$ . DNA was extracted using the traditional phenol/chloroform extraction method and was quality checked us-



Figure 1: *A. fulica* individual used for genome sequencing and assembly.

ing agarose gel electrophoresis, meeting the requirement for library construction for the Illumina X Ten (Illumina Inc., San Diego, CA, USA) and for the PacBio Sequel (Pacific Biosciences, Menlo Park, CA, USA) sequencing platforms.

RNA was extracted from the pallium, liver, foot, spleen, stomach, gut, and heart using TRIzol® Reagent (Life Technologies, Gaithersburg, USA). The RNA quality was checked using the Nanodrop ND-1000 spectrophotometer (LabTech, USA) and 2100 Bioanalyzer (Agilent Technologies, USA) with RNA integrity number  $>8$  (Supplemental Table S1 and Supplemental Fig. S1). The RNA from each sample was equally mixed for the RNA sequencing on the PacBio Sequel platform. First, messenger RNA molecules were reverse-transcribed to complementary DNA (cDNA) using Clontech SMARTer cDNA synthesis kit. After cDNA amplification and purification, 2 SMRTbell libraries of 0–4 and 4–10 kb were generated using the size selection in the BluePippin Size Selection System (Pacific Biosciences) and protocols suggested by the manufacturer. The final libraries were sequenced in the PacBio SEQUEL platform (Pacific Biosciences), resulting in 12,439,996 subreads totaling  $\sim 22.5$  Gb PacBio long reads with average length  $>1,801$  bps. Subsequently, a total of 782,613 circular consensus sequences were generated on the basis of the subreads, and 553,889 full-length non-chimeric sequences (FLNC) representing 23,726 gene loci were ultimately obtained. All aforementioned data processing was performed using SMRT Link v5.0 [22]. Moreover,  $\sim 70.37\%$  of the multi-exon FLNCs were really full-length sequences embracing all the exons of the gene locus predicted from the whole-genome sequences.

Using the DNA from the abdominal foot, a library with insertion length of 300 bp was constructed and sequenced using the Illumina sequencing platform according to the manufacturer's protocol. Approximately 202.23 Gb short reads were obtained using Illumina X Ten sequencing technology (Table 1), which was used for the following genome survey analysis, and for final base-level genome sequence correction. Meanwhile, four 20-kb libraries were constructed for PacBio Sequel sequencing. Using 16 sequencing SMRT cells, 104.6-Gb long reads were generated (Table 1). The mean and N50 lengths of the polymerases for sequencing cells ranged from 6.4 to 10.4 kb and from 12.3 to 20.3 kb, respectively. Those long genomic DNA reads were then used for reference genome construction.

## Genome feature estimation from k-mer method

With sequencing data from the Illumina platform, several genome characteristics could be evaluated for *A. fulica*. To ensure the quality of the analysis, ambiguous bases and low-quality reads were trimmed and filtered using the HTQC package (version 1.92.3) [23]. The following quality controls were performed under the framework of HTQC. First, the quality of bases at 2 read ends were checked. Bases in sliding 5-bp windows were deleted if the average quality of the window was below phred quality score of 20. Second, reads were filtered if the average phred quality score was <20 or the read length was <75 bp. Third, the mate reads were also removed if the corresponding reads were filtered.

The quality-controlled reads were used for genome character estimation. We calculated the number of each 17-mer from the sequencing data using the Jellyfish software (Jellyfish, [RRID:SCR.005491](#); version 2.0) [24], and the distribution was analyzed with GCE software (GCE, [RRID:SCR.017332](#); version 3) [25] and is shown in Supplemental Fig. S2. We estimated the genome size at 2.12 Gb with heterozygosity of 0.47% and repeat content of 71% in the genome. Previous studies have revealed that repeat content varies in mollusks and is correlated with genome size [26]. The large genome size and high proportion of repeat content of *A. fulica* provided additional supporting data for the statistical analysis. Moreover, 10,000 pairs of short reads were extracted randomly and were compared to the nt database and no obvious external contamination was found.

## Genome assembly by third-generation long reads

After removing adaptor sequences in polymerases, 101.6-Gb subreads were generated for the following whole-genome assembly. The average and N50 length of subreads reached 5.25 and 8.80 kb, respectively. To optimize the genome assembly using the PacBio sequencing data, we applied 2 packages in the assembly process, Canu v1.8 (Canu, [RRID:SCR.015880](#)) [27] and FALCON v0.2.2 (Falcon, [RRID:SCR.016089](#)) [28]. The Canu package was first applied for the assembly using default parameters. As a result, a 1.93-Gb genome was constructed with 10,417 contigs and a contig N50 length of 662.40 kb. FALCON was also employed using the `length_cutoff` and `pr_length_cutoff` parameters of 10 and 8 kb, respectively. We obtained 1.85 Gb genome with 8,585 contigs, with a contig N50 of 726.63 kb. We adopted the FALCON assembly as the reference genome for *A. fulica* (Table 2). Compared with the estimated genome size, the assembled version was relatively smaller, which may have resulted from the following 2 possible scenarios: the high repeat content of the genome and the probably larger size estimated from the k-mer analysis. The genome sequences were subsequently polished, PacBio long reads using Arrow [29] and Illumina short reads using Pilon [30] to correct base errors. The corrected genome was further applied for the following chromosome assembly construction using Hi-C data.

## In situ Hi-C library construction and chromosome assembly using Hi-C data

Liver pancreas tissue of *A. fulica* was used for library construction for Hi-C analysis, and the library was constructed using the identical method as previous studies [31]. Finally, the library was sequenced with 150 paired-end mode on the Illumina HiSeq X Ten platform (San Diego, CA, United States). From the Illumina sequencing platform, 1,313.87 million paired-end reads were ob-

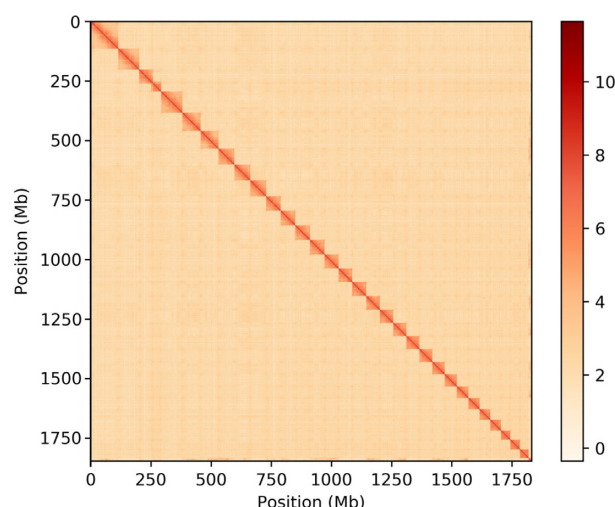


Figure 2: Contact matrix generated from the Hi-C data analysis showing sequence interactions in chromosomes. Color bar indicates the logarithm of the contact density.

tained for the Hi-C library (Table 1). The reads were mapped to the above *A. fulica* genome with Bowtie2 [32], with 2 ends of paired reads being mapped to the genome separately. To increase the interactive Hi-C read ratio, an iterative mapping strategy was performed as in previous studies, and only read pairs with both ends uniquely mapped were used for the following analysis. From the alignment status of the 2 ends, self-ligation, non-ligation, and other sorts of invalid reads, including StartNearRsite, PCR amplification, random break, LargeSmallFragments, and ExtremeFragments, were filtered out by Hi-Clib [33]. Through the recognition of restriction sites in sequences, contact counts among contigs were calculated and normalized.

According to previous karyotype analyses, *A. fulica* has 31 chromosomes [34]. By clustering the contigs using the contig contact frequency matrix, we were able to correct some minor errors in the FALCON assembly results. Contigs with errors were broken into shorter contigs. We obtained 8,701 contigs, slightly more than the 8,585 contigs in the FALCON assembly. We successfully clustered these contigs into 31 groups in Lachesis [35] using the agglomerative hierarchical clustering method (Fig. 2). Lachesis was further applied to order and orient the clustered contigs according to the contact matrix. As a result, 7,106 contigs were reliably anchored, ordered, and oriented on chromosomes, accounting for 99.32% of the total genome bases. The first near chromosomal-level assembly of *A. fulica* was obtained with 8,211 contigs, a contig N50 of 721.0 kb, and a scaffold N50 of 59.59 Mb (Tables 2 and 3).

## Genome quality evaluation

We assessed the quality of the genome of *A. fulica* after the assembly process. The quality evaluation was carried out in 3 aspects: continuity, completeness, and the mapping rate of next-generation sequencing data.

First of all, we compared the sequence number and contig N50 length of *A. fulica* with publicly available mollusk genomes and found that our assembly has a high quality on contig and scaffold N50 among mollusk genomes (Table 3). Traditional chromosomal genome assembly requires physical maps and genetic maps, which is enormously time- and labor-consuming.

**Table 1:** Sequencing data generated for *A. fulica* genome assembly and annotation

Library type	Platform	Library size (bp)	Data size (Gb)	Application
Short reads	HiSeq X Ten	350	202.24	Genome survey and genomic base correction
Long reads	PacBio SEQUEL	20,000	101.63	Genome assembly
Hi-C	HiSeq X Ten	300–500	199.73	Chromosome construction

**Table 2:** Statistics for genome assembly of *A. fulica*

Sample ID	Length (bp)		Number	
	Contig**	Scaffold	Contig**	Scaffold
Total	1,852,282,574	1,855,883,074	8,211	1,010
Max	5,947,392	116,558,012		
N50	721,038	59,589,303	697	13
N60	538,883	58,013,356	995	16
N70	399,612	53,672,006	1,396	20
N80	268,901	50,673,968	1,957	23
N90	141,756	44,109,545	2,888	27

\*\*This indicates the ultimate contigs because they were probably modified during the Hi-C step.

**Table 3:** Summary of the genome of *A. fulica* and other published mollusk genomes

Species	Size* (Mb)	Contig N50 (kb)	Scaffold N50 (kb)
<i>Achatina fulica</i> (this study)**	2,120	721	59,590
<i>Pomacea canaliculata</i> [20]**	570	995	38,000
<i>Crassostrea gigas</i> [36]	545	7.5	401
<i>Pinctada fucata</i> [37]	1,150	1.6	14.5
<i>Pinctada fucata new</i> [38]	1,150	21	324
<i>Pinctada fucata V2</i> [39]	1,150	21	167
<i>Biomphalaria glabrata</i> [21]	931	7.3	48
<i>Ruditapes philippinarum</i> [40]	1,370	3.3	32.7
<i>Patinopecten yessoensis</i> [41]**	1,430	38	41,000
<i>Radix auricularia</i> [42]	1,600	0.324	578
<i>Octopus bimaculoides</i> [43]	2,800	5.4	470
<i>Mytilus galloprovincialis</i> [26]	1,600	2.6	2.9
<i>Lottia gigantea</i> [44]	420	96	1,870
<i>Patella vulgata</i> [45]	1,460	3.1	3.1
<i>Aplysia californica</i>	1,760	9.6	917
<i>Conus tribblei</i> [46]	2,760	0.85	215
<i>Limnoperna fortunei</i> [47]	1,600	10	312
<i>Bathymodiolus platifrons</i> [48]	1,640	13.2	343
<i>Modiolus philippinarum</i> [48]	2,380	19.7	100.2
<i>Chlamys farreri</i> [49]	1,200	1.2	1.5
<i>Lingula anatina</i> [50]	463	55	294
<i>Argopecten purpuratus</i> [51]	885	80.1	1,020

\*Estimated size of the genome.

\*\*Genomes assembled into near chromosomal level.

With Hi-C data analysis, we successfully assembled the *A. fulica* genome into near chromosome-level with just 1 individual.

Second, the assembled genome was subjected to BUSCO (version 3.0, metazoa odb9) [52] to assess the completeness of the genome. Approximately 91.7% of the BUSCO genes were identified in the *A. fulica* genome, and >84.7% of the BUSCO genes were single-copy completed in our genome, illuminating a high level of completeness of the genome.

Third, next-generation sequencing short reads were aligned to the genome using the BWA package (BWA, RRID:SCR.010910; version 0.7.17) [53], and ~98.7% of paired reads were aligned to the genome, of which 98.24% were reads that were pair aligned.

### Repeat element and gene annotation

Tandem Repeat Finder 4.09 (TRF) [54] was used for repetitive element identification in the *A. fulica* genome. A *de novo* method applying RepeatModeler (RepeatModeler, RRID:SCR.015027) was used to detect transposable elements (TEs). The resulting *de novo* data, combined with the known repeat library from Repbase [55], were used to identify TEs in the *A. fulica* genome by means of RepeatMasker4-0-8 (RepeatMasker, RRID:SCR.012954) [56] software. All repetitive elements were masked in the genome before protein-coding gene prediction.



**Table 4:** Statistics for genome annotation of *A. fulica*

Database	Number (%)
InterPro	16,252 (68.50)
GO	12,101 (51.00)
KEGG all	21,325 (89.88)
KEGG	10,161 (42.83)
Orthology	
Swiss-Prot	17,050 (71.86)
TrEMBL	22,403 (94.42)
NR	22,553 (95.06)
Total	23,726 (100)

Protein-coding genes in the *A. fulica* genome were annotated using the *de novo* program Augustus 0.2.1 (Augustus, [RRID:SCR.008417](#)) [57]. Protein sequences of closely related species including *Aplysia californica*, *B. glabrata*, *Crassostrea gigas*, *Lottia gigantea*, and *Patinopecten yessoensis* were downloaded from the Ensembl database and aligned to the *A. fulica* genome with TBLASTN2.6.0 (TBLASTN, [RRID:SCR.011822](#)). Full-length transcripts obtained using Iso-Seq were mapped to the genome using Genewise (GeneWise, [RRID:SCR.015054](#)) [58]. Finally, gene models predicted from all above methods were combined by MAKERV2.31.10 (MAKER, [RRID:SCR.005309](#)) [59], resulting in 23,726 protein-coding genes. The gene number, gene length, coding DNA sequence (CDS) length, exon length, and intron length distribution were all comparable to those of the related mollusks (Fig. 3).

To functionally annotate protein-coding genes in the *A. fulica* genome, we searched all predicted gene sequences against the NCBI non-redundant nucleotide (NT) and protein (NR) and Swiss-Prot databases using the BLASTN (BLASTN, [RRID:SCR.001598](#)) [60] and BLASTX (BLASTX, [RRID:SCR.001653](#)) [61] utilities. Blast2GO (Blast2GO, [RRID:SCR.005828](#)) [62] was also used to assign gene ontology (GO) [63] and KEGG [64] pathways. A threshold of  $e$ -value of  $1e-5$  was used for all BLAST applications. Finally, 22,858 (96.34%) genes were functionally annotated (Table 4).

### Phylogenetic analysis of *A. fulica* with other mollusks

OrthoMCLv1.2 [65] was used to cluster gene families. First, proteins from *A. fulica* and closely related mollusks, including *A. californica*, *B. glabrata*, *C. gigas*, *Lingula anatina*, *L. gigantea*, *P. yessoensis*, *Octopus bimaculoides*, *Helobdella robusta*, *P. canaliculata*, and the outgroup, *Drosophila melanogaster*, were all-to-all blasted by the BLASTP (BLASTP, [RRID:SCR.001010](#)) [61] utility with an  $e$ -value threshold of  $1e-5$ . Only proteins from the longest transcript were used for genes with alternative isoforms. We identified 25,448 gene families for *A. fulica* and the related species; among them 675 single-copy ortholog families were detected.

Using single-copy orthologs, we could probe the phylogenetic relationships for *A. fulica* and other mollusks. To this end, protein sequences of single-copy genes were aligned using CLUSTALX2.0 (Clustal X, [RRID:SCR.017055](#)) [66]. Guided by the protein multi-sequence alignment, the alignment of the CDSs for those genes was generated and concatenated for the following analysis. The phylogenetic relationships were constructed using PhyML3.0 (PhyML, [RRID:SCR.014629](#)) [67] using the concatenated nucleotide alignment with the JTT+G+F model. The MCMCTree program in PAML4 [67] was used to estimate the species divergent time scales for the mollusks using the approximate likelihood method and calibrated according to the fossil

records. We found that *A. fulica* was most closely related to *B. glabrata*, and the 2 species diverged from their common ancestor ~243 million years ago (MYA) (Fig. 4).

## Conclusion

We reconstructed the first chromosome-level assembly for *A. fulica* using an integrated sequencing strategy combining PacBio, Illumina, and Hi-C technologies. Using the long reads from the PacBio Sequel platform and short reads from the Illumina X Ten platform, we successfully constructed a contig assembly for *A. fulica*. Leveraging contact information among contigs from Hi-C technology, we further improved the assembly to near chromosome-level quality (Table 3 and Fig. 2). We predicted 23,726 protein-coding genes in the *A. fulica* genome, and 22,858 of the genes were functionally annotated with putative functions. With 675 single-copy orthologs from *A. fulica* and other related mollusks, we constructed the phylogenetic relationship of these mollusks and found that *A. fulica* might have diverged from its common ancestor *B. glabrata* ~177.1–187.1 MYA. Given the increasing interest in mollusk genomic evolution and the biological importance of *A. fulica* as an invasive animal, our genomic and transcriptome data will provide valuable genetic resources for follow-on functional genomics investigations by the research community.

## Ethics Statement

This study was approved by the Animal Care and Use committee of the National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention.

## Availability of Supporting Data and Materials

The Illumina, PacBio, and Hi-C sequencing data are available from NCBI via the accession numbers SRR8369706, SRR8369311, and SRR8371669, respectively. The Illumina transcriptome sequencing data were deposited to NCBI via the accession numbers SRR8371872 and SRR8371873. The genome, annotation, and intermediate files have been uploaded to the GigaScience GigaDB Database [68].

## Additional Files

**Supplemental Table S1.** Summary of RNA quality of samples. The high-quality samples are highlighted in red for the PacBio library construction and sequencing.

**Supplemental Figure S1.** Bioanalyzer summary reports for samples used in the transcriptome sequencing.

**Supplemental Figure S2.** The distribution of  $k$ -mer species estimated for *A. fulica*. The total number of  $k$ -mer species is 178,847,565,204, with the peak value (depth) being 76.

## Abbreviations

BLAST: Basic Local Alignment Search Tool; BUSCO: Benchmarking Universal Single-Copy Orthologs; BWA: Burrows-Wheeler Aligner; cDNA: complementary DNA; CDS: coding DNA sequence; FLNC: full-length non-chimeric sequence; Gb: gigabase pairs; GCE: Genomic Character Estimator; GO: gene ontology; kb: kilobase pairs; KEGG: Kyoto Encyclopedia of Genes and Genomes; Mb: megabase pairs; MYA: million years ago; NCBI: National Center for Biotechnology Information; PacBio: Pacific

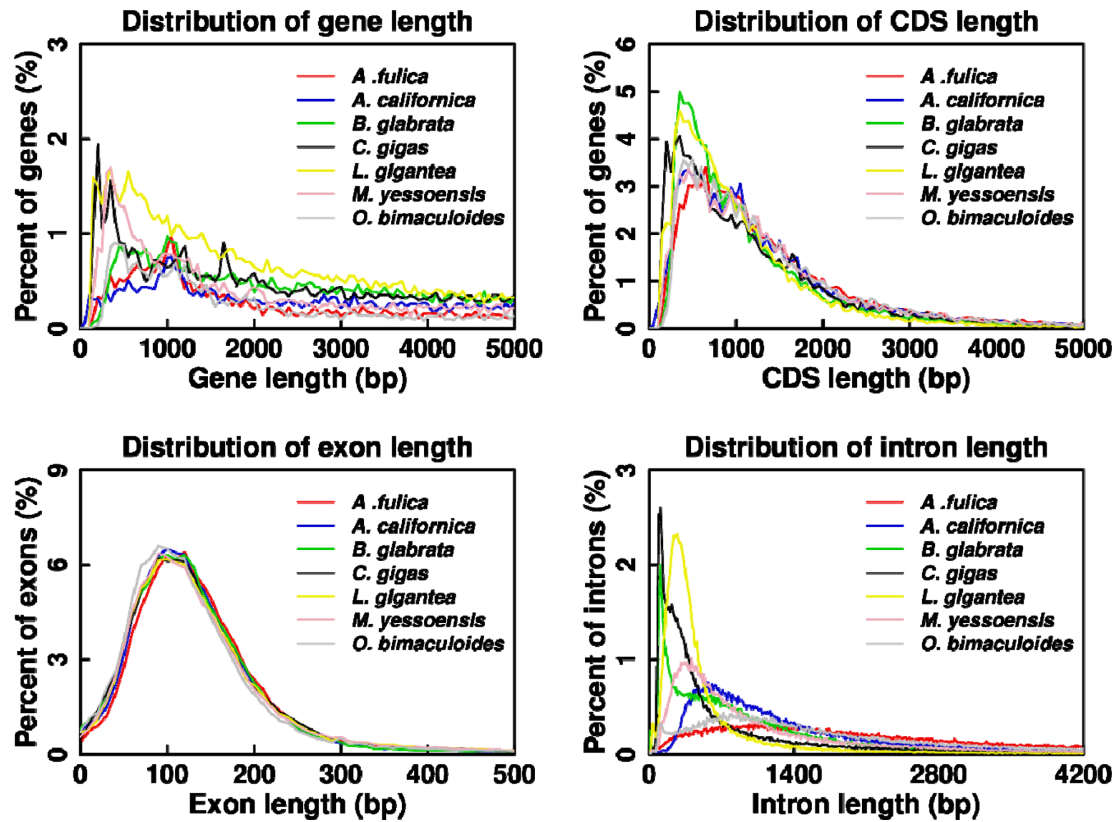


Figure 3: Length distribution comparison of genes (A), CDSs (B), exons (C), and introns (D) for *A. fulica* to those in the closely related mollusk species *A. californica*, *B. glabrata*, *C. gigas*, *L. gigantea*, *P. yessoensis*, and *O. bimaculoides*.

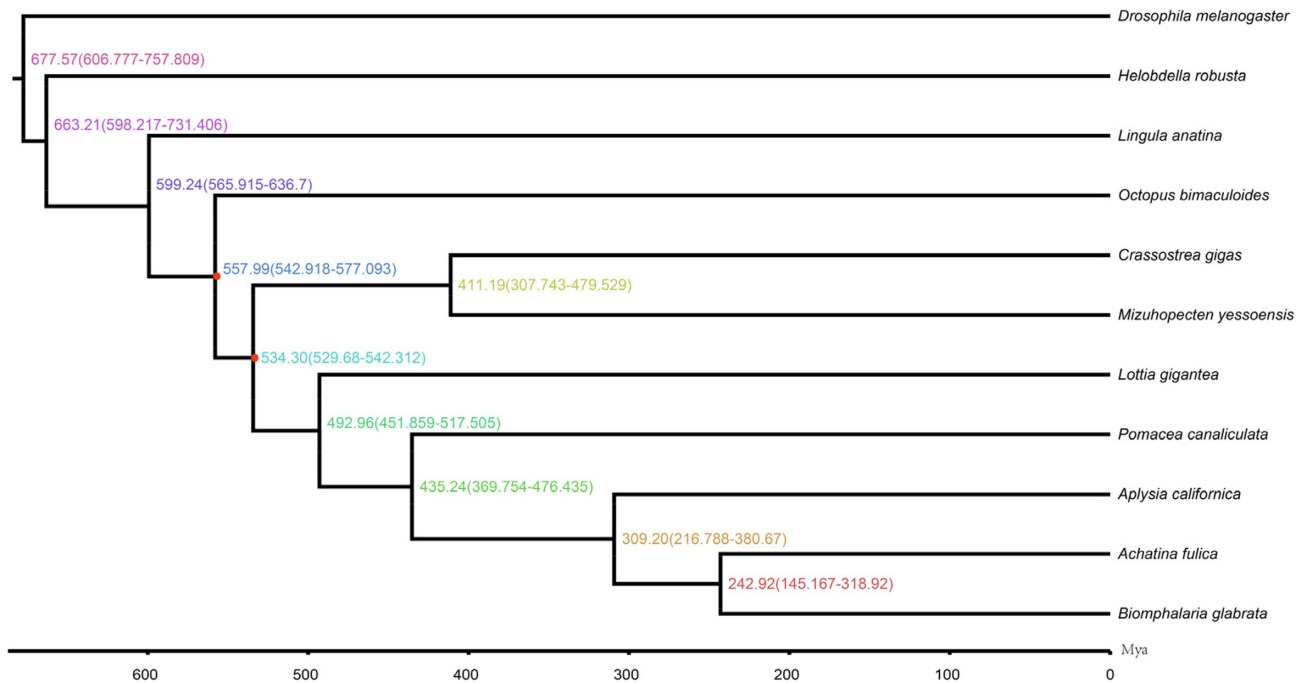


Figure 4: Phylogenetic relationship between *A. fulica* and related species. The divergence times (million years ago [MYA]) with 95% confidence intervals are labeled at branch sites. Red dots in the tree denote the fossil recalibration sites, with a maximum and minimum age of Bivalve/Gastropod divergence of 543 and 530 MYA and maximum age of Mollusk crown group divergence of 549 MYA.

Biosciences; SMRT: Single-Molecule Real-Time; TE: transposable element; TrEMBL: Translation of European Molecular Biology Laboratory.

## Competing Interests

The authors declare that they have no competing interests.

## Funding

This work was supported by the National Key Research and Development Program of China (Nos. 2016YFC1200500 and 2016YFC1202000).

## Authors' Contributions

Y.G., X.Z., W.H. and N.X. conceived the project; Y.G., Y.Z., Y.H., G.M., Z.Y., E.A. and J.L. collected the samples and extracted the genomic DNA. Y.G., Y.Z., Q.L., Z.W. and S.L. performed the genome assembly and data analysis. Y.G., X.Z., W.H., and N.X. wrote the manuscript. W.H. and N.X. read and approved the final version of the manuscript.

## Acknowledgements

The authors thank Frasergen Bioinformatics for providing technical support for this work.

## References

- Schreurs J. Investigations on the Biology, Ecology and Control of Giant African Snail 290 in West New Guinea. Manokwari Agricultural Research Station, 1963.
- Albuquerque FS, Peso-Aguiar MC, Assunção-Albuquerque MJ. Distribution, feeding behavior and control strategies of the exotic land snail *Achatina fulica* (Gastropoda:Pulmonata) in the Northeast of Brazil. *Braz J Biol* 2008; **68**:6.
- Thiengo SC, Fernandez MA, Torres EJ, et al. First record of a nematode *Metastrongyloidea* (*Aelurostrongylus abstrusus* larvae) in *Achatina* (*Lissachatina*) *fulica* (Mollusca, Achatinidae) in Brazil. *J Invertebr Pathol* 2008; **98**:6.
- Lv S, Zhang Y, Liu HX. Invasive snails and an emerging infectious disease: results from the first national survey on *Angiostrongylus cantonensis* in China. *PLoS Negl Trop Dis* 2009; **3**(2):e368.
- Cowie RH. Non-indigenous land and freshwater molluscs in the islands of the Pacific: conservation impacts and threats. In: Sherley G, ed. *Invasive Species in the Pacific: A Technical Review and Regional Strategy*. Pacific Regional Environmental Program; 2000.
- Cowie RH. Can snails ever be effective and safe biocontrol agents? *Int J Pest Manag* 2001; **47**:18.
- Cowie RH, Robinson DG. *Pathways of Introduction of Non-indigenous Land and Freshwater Snails and Slugs*. Washington, DC: Island Press; 2003.
- Kotangale JP. Giant African snail (*Achatina fulica* Bowdich). *J Environ Sci Eng* 2011; **53**:6.
- Raut SK, Barker GM. *Achatina fulica* Bowdich and Other Achatinidae as Pests in Tropical Agriculture. Wallingford, UK: CABI International; 2002.
- Jarreit VHC. The spread of the snail *Achatina fulica* to south China. *Hong Kong Nat* 1931; **2**:3.
- Shan L, Yi Z, Peter S. Emerging angiostrongyliasis in mainland China. *Emerg Infect Dis* 2008; **14**(1):4.
- [http://www.mee.gov.cn/gkml/zj/wj/200910/t20091022\\_172155.htm](http://www.mee.gov.cn/gkml/zj/wj/200910/t20091022_172155.htm).
- Lowe S, Browne SM, Boudjrlas S, et al. 100 of the World's Worst Invasive Alien Species: A Selection from the Global Invasive Species Database. Auckland, New Zealand: ISSG; 2000.
- Mead AR. *Pulmonates Volume 2B. Economic Malacology with Particular Reference to Achatina fulica*. London, UK: Academic Press; 1979.
- Alicata JE. The discovery of *Angiostrongylus cantonensis* as a cause of human eosinophilic meningitis. *Parasitol Today* 1991; **7**(6):151–3.
- Prociv P, Spratt DM, Carlisle MS. Neuro-angiostrongyliasis: unresolved issues. *Int J Parasitol* 2000; **30**(12–13): 1295–303.
- Deng ZH, Zhang QM, Huang SY, et al. First provincial survey of *Angiostrongylus cantonensis* in Guangdong Province, China. *Trop Med Int Health* 2012; **17**:4.
- Maldonado JA, Simoes RO, Oliveira AP, et al. First report of *Angiostrongylus cantonensis* (Nematoda: Metastrongylidae) in *Achatina fulica* (Mollusca: Gastropoda) from Southeast and South Brazil. *Mem Inst Oswaldo Cruz* 2010; **105**:4.
- Vitta A, Polseela R, Nateeworanart S, et al. Survey of *Angiostrongylus cantonensis* in rats and giant African land snails in Phitsanulok Province, Thailand. *Asian Pac J Trop Med* 2011; **4**:3.
- Liu C, Zhang Y, Ren Y, et al. The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. *Gigascience* 2018; **7**(9), doi:10.1093/gigascience/giy101.
- Adema CM, Hillier LW, Jones CS, et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat Commun* 2017; **8**:15451.
- Pacific Biosciences. [www.pacb.com](http://www.pacb.com).
- Neff KL, Argue DP, Ma AC, et al. Mojo Hand, a TALEN design tool for genome editing applications. *BMC Bioinformatics* 2013; **14**:1.
- Marcais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 2011; **27**(6):764–70.
- Liu B, Shi Y, Yuan J, et al. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *Quant Biol* 2013; **35**:62–7.
- Murgarella M, Puiu D, Novoa B, et al. A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. *PLoS One* 2016; **11**(3):e0151561.
- Koren S, Walenz BP, Berlin K, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 2017; **27**(5):722–36.
- Chin CS, Peluso P, Sedlazeck FJ, et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 2016; **13**(12):1050–4.
- Chin C-S, Alexander DH, Marks P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 2013; **10**(6):563.
- Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014; **9**(11):e112963.
- Gong G, Dan C, Xiao S, et al. Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis. *Gigascience* 2018; **7**(11), doi:10.1093/gigascience/giy120.

32. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;**10**(3):R25.
33. Burton JN, Adey A, Patwardhan RP, et al. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol* 2013;**31**(12):1119.
34. Sun T. Chromosomal studies in three land snails. *Sinozoologia* 1995;**12**:154–62.
35. Near TJ, Dornburg A, Eytan RI, et al. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc Natl Acad Sci U S A* 2013;**110**(31):12738.
36. Zhang G, Fang X, Guo X, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012;**490**(7418):49–54.
37. Takeuchi T, Kawashima T, Koyanagi R, et al. Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. *DNA Res* 2012;**19**(2):117–30.
38. Takeuchi T, Koyanagi R, Gyoja F, et al. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zool Lett* 2016;**2**:3.
39. Du X, Fan G, Jiao Y, et al. The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *Gigascience* 2017;**6**(8):1–12.
40. Mun S, Kim YJ, Markkandan K, et al. The whole-genome and transcriptome of the Manila clam (*Ruditapes philippinarum*). *Genome Biol Evol* 2017;**9**(6):1487–98.
41. Wang S, Zhang J, Jiao W, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol* 2017;**1**(5):120.
42. Schell T, Feldmeyer B, Schmidt H, et al. An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). *Genome Biol Evol* 2017;**9**(3):585–92.
43. Albertin CB, Simakov O, Mitros T, et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 2015;**524**(7564):220–4.
44. Simakov O, Marletaz F, Cho SJ, et al. Insights into bilaterian evolution from three spiralian genomes. *Nature* 2013;**493**(7433):526–31.
45. Kenny NJ, Namigai EK, Marletaz F, et al. Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *Patella vulgata* (Mollusca, Patellogastropoda) and *Spirobranchus (Pomatoceros) lamarcki* (Annelida, Serpulida). *Mar Genomics* 2015;**24**(Pt 2):139–46.
46. Barghi N, Concepcion GP, Olivera BM, et al. Structural features of conopeptide genes inferred from partial sequences of the *Conus tribblei* genome. *Mol Genet Genomics* 2016;**291**(1):411–22.
47. Uliano-Silva M, Dondero F, Dan Otto T, et al. A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei*. *Gigascience* 2018;**7**(2), doi:10.1093/gigascience/gix128.
48. Sun J, Zhang Y, Xu T, et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat Ecol Evol* 2017;**1**(5):121.
49. Jiao W, Fu X, Dou J, et al. High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. *DNA Res* 2014;**21**(1):85–101.
50. Luo YJ, Takeuchi T, Koyanagi R, et al. The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat Commun* 2015;**6**:8301.
51. Li C, Liu X, Liu B, et al. Draft genome of the Peruvian scallop. *Gigascience* 2018;**7**(4), doi:10.1093/gigascience/giy031.
52. Simão FA, Waterhouse RM, Ioannidis P, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31**(19):3210–2.
53. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;**25**(14):1754–60.
54. Benson G. Tandem Repeats Finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999;**27**(2):573–80.
55. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 2015;**6**:11.
56. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* 2004;**5**(1):4.10.1–4.4.
57. Stanke M, Keller O, Gunduz I, et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 2006;**34**(suppl 2):W435–W9.
58. Birney E, Clamp M, Durbin R. GeneWise and genomewise. *Genome Res* 2004;**14**(5):988–95.
59. Cantarel BL, Korf I, Robb SM, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 2008;**18**(1):188–96.
60. Gertz EM, Yu YK, Agarwala R, et al. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol* 2006;**4**:41.
61. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;**10**:421.
62. Conesa A, Götz S, García-Gómez JM, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005;**21**(18):3674–6.
63. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;**32**(suppl 1):D258–D61.
64. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;**28**(1):27–30.
65. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003;**13**(9):2178–89.
66. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* 2003;**1**:2.3.1–22.
67. Guindon S, Lethiec F, Duroux P, et al. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 2005;**33**(suppl 2):W557–W9.
68. Guo Y, Zhang Y, Liu Q, et al. Supporting data for “A chromosomal-level genome assembly for the giant African snail *Achatina fulica*.” *GigaScience Database* 2019. <http://dx.doi.org/10.5524/100647>.