



Apr 21, 2019

Working

## Key protocols for chromosome-level genome assembly of the *Scapharca (Anadara) broughtonii*.

Chang-Ming Bai<sup>1</sup>, Lu-Sheng Xin<sup>1</sup>, Umberto Rosani<sup>2</sup>, Biao Wu<sup>1</sup>, Qing-Chen Wang<sup>1</sup>, Xiao-Ke Duan<sup>3</sup>, Zhi-Hong Liu<sup>1</sup>, Chong-Ming Wang<sup>1</sup>

<sup>1</sup>Key Laboratory of Maricultural Organism Disease Control, Ministry of Agriculture; Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology; Qingdao Key Laboratory of Mariculture Epidemiology and Biosecurity; Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China, <sup>2</sup>Department of Biology, University of Padua, Padua 35121, Italy and Alfred Wegener Institute – Helmholtz Centre for Polar and Marine Research, Waddensea Station Sylt, Germany, <sup>3</sup>Biomarker Technologies Corporation, Beijing 101200, China

[dx.doi.org/10.17504/protocols.io.zimf4c6](https://doi.org/10.17504/protocols.io.zimf4c6)

Chang-Ming Bai

### ABSTRACT

**Background:** The blood clam, *Scapharca (Anadara) broughtonii*, is an economically and ecologically important marine bivalve of the Family Arcidae. The efforts that have been made to study their population genetics, breeding, cultivation and stock enrichment were somewhat hindered by the lack of a reference genome. Here, we reported the complete genome sequence of *S. broughtonii*, a first reference genome of the Family Arcidae.

**Funding:** A total of 75.79 Gb clean data was generated with the PacBio and Oxford Nanopore platforms, which represented approx. 86× coverage of the *S. broughtonii* genome. *De novo* assembly of these long reads resulted in an 884.5 Mb genome, with a contig N50 of 1.80 Mb and scaffold N50 of 45.00 Mb, respectively. Genome Hi-C scaffolding resulted in 19 chromosomes containing 99.35% of bases of the assembled genome. Genome annotation revealed that a considerable part of the genome (46.1%) is composed by repeated sequences, while 24,045 protein-coding genes were predicted and 84.7% of them were annotated.

**Conclusion:** We report here the chromosomal-level assembly of the *S. broughtonii* genome based on long read sequencing and Hi-C scaffolding. The genomic data could be served as reference genome for the Arcidae Family and will provide a valuable resource for the scientific community and aquaculture sector.

### Collection protocols

	The PacBio libraries preparation for the <i>Scapharca broughtonii</i> . by Chang-Ming Bai	<a href="#">RUN</a>
	The pipeline of assembly and annotation of the <i>Scapharca broughtonii</i> genome by Chang-Ming Bai	<a href="#">RUN</a>
	Extraction of <i>Scapharca broughtonii</i> genomic DNA by Chang-Ming Bai	<a href="#">RUN</a>
	The Illumina libraries preparation for the <i>Scapharca broughtonii</i> by Chang-Ming Bai	<a href="#">RUN</a>
	The pipeline of Hi-C assembly of the <i>Scapharca broughtonii</i> genome by Chang-Ming Bai	<a href="#">RUN</a>
	The Nanopore libraries preparation for the <i>Scapharca broughtonii</i> by Chang-Ming Bai	<a href="#">RUN</a>



This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited