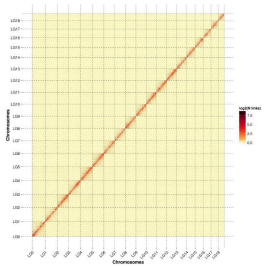
Apr 21, 2019

# The pipeline of Hi-C assembly of the Scapharca broughtonii genome



GigaScience

In 2 collections

DOI

dx.doi.org/10.17504/protocols.io.z8cf9sw

Chang-Ming CM Bai[1]

[1]Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences

GigaScience Press

Chang-Ming CM Bai

Yellow Sea Fisheries Research Institute, Chinese Academy of ...

OPEN ACCESS

**DOI: dx.doi.org/10.17504/protocols.io.z8cf9sw**

**External link: https://doi.org/10.1093/gigascience/giz067**

**Protocol Citation:** Chang-Ming CM Bai 2019. The pipeline of Hi-C assembly of the Scapharca broughtonii genome. **protocols.io https://dx.doi.org/10.17504/protocols.io.z8cf9sw**

**Manuscript citation:**
Chang-Ming Bai, Lu-Sheng Xin, Umberto Rosani, Biao Wu, Qing-Chen Wang, Xiao-Ke Duan, Zhi-Hong Liu, Chong-Ming Wang, Chromosomal-level assembly of the blood clam, *Scapharca* (*Anadara*) *broughtonii*, using long sequence reads and Hi-C, *GigaScience*, Volume 8, Issue 7, July 2019, giz067, https://doi.org/10.1093/gigascience/giz067



**Protocol status:** Working
**We use this protocol and it's working**

**Created:** April 21, 2019

**Last Modified:** April 21, 2019

**Protocol Integer ID:** 22500

## Abstract

This protocol include the detailed methods of Hi-C assembly of the *Scapharca broughtonii* genome

1 Run BWA (v0.7.10-r789) to align the Hi-C reads to the initially assembled S. broughtonii genome, and found the Hi-C reads mapped to the assembled genome.

Note

bwa index -a bwtsw fasta
bwa aln -M 3 -O 11 -E 4 -t 2 fq1
bwa aln -M 3 -O 11 -E 4 -t 2 fq2

2 Filter the mapped Hi-C reads obtalined in the step 1 using HiC-Pro (v. 2.10.0).

Note

mapped_2hic_fragments.py -v -S -s 100 -l 1000 -a -f -r -o

3 Extract valid interaction pair reads according the HiC-Pro results.

4 Break the initial assembly to 300 bp, and then run LACHESIS (v2e27abb) for assembling based on Hi-C data.

5 Run LACHESIS (v2e27abb) to assemble corrected contigs obtained in step 4 into chromosome and modified manually.

Note

(1) CLUSTER_MIN_RE_SITES = 22 ;
(2) CLUSTER_MAX_LINK_DENSITY=2 ;
(3) CLUSTER_NONINFORMATIVE_RATIO = 2 ;
(4) ORDER_MIN_N_RES_IN_TRUN=10 ;
(5) ORDER_MIN_N_RES_IN_SHREDS=10.