Mar 24, 2021

Protocols for "Chromosome-level genome assembly of the humpback puffer, Tetraodon palembangensis"

Rui Zhang[1], Chang Li[1], Mengjun Yu[1], Xiaoyun Huang[1], Mengqi Zhang[1], Shanshan Liu[1], Shanshan Pan[1], Weizhen Xue[1], Congyan Wang[1], Chunyan Mao[1], He Zhang[2], Guangyi Fan[1]

[1]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China;

[2]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China.Department of Biology, Hong Kong Baptist University, Hong Kong, China

| 1 | Works for me | dx.doi.org/10.17504/protocols.io.bs8inhue |

BGI   GIGA   1 more workspace

zhangrui7

SUBMIT TO PLOS ONE

ABSTRACT

In this study we assembled the first chromosome-level genome assembly of the humpback puffer (*Tetraodon palembangensis*), a species of poisonous freshwater pufferfish mainly distributed in Southeast Asia. This utilized Oxford Nanopore and BGISEQ-500 sequencing, stLFR libraries and Hi-C assembly, and the protocols for nucleic acid extraction, library construction and sequencing are all made available here. The resulting genome size was 362 Mb with a ~1.78 Mb contig N50 and ~15.8 Mb scaffold N50. Based on the assembled genome, ~61.5Mb (18.11%) repeat sequences were identified, 19,925 genes were annotated, and 90.01% of these genes could be predicted with function. Finally, a phylogenetic tree of ten teleost fish species was successfully constructed using this and public data. These protocols are made available for others carrying vertebrate genome assembly, and the open data and published study for comparative genome studies.

CREATED

Mar 11, 2021

LAST MODIFIED

Mar 24, 2021

ABSTRACT

In this study we assembled the first chromosome-level genome assembly of the humpback puffer (*Tetraodon palembangensis*), a species of poisonous freshwater pufferfish mainly distributed in Southeast Asia. This utilized Oxford Nanopore and BGISEQ-500 sequencing, stLFR libraries and Hi-C assembly, and the protocols for nucleic acid extraction, library construction and sequencing are all made available here. The resulting genome size was 362 Mb with a ~1.78 Mb contig N50 and ~15.8 Mb scaffold N50. Based on the assembled genome, ~61.5Mb (18.11%) repeat sequences were identified, 19,925 genes were annotated, and 90.01% of these genes could be predicted with function. Finally, a phylogenetic tree of ten teleost fish species was successfully constructed using this and public data. These protocols are made available for others carrying vertebrate genome assembly, and the open data and published study for comparative genome studies.

FILES

Hi-C library preparation for the Lateolabrax maculatus genome
Version 1
by Chang Li

RNA isolation with TRIzol
Version 1
by Magdalena Julkowska, King Abdullah University of Science and Technology

BGISEQ-500 WGS library construction
Version 1
by Xinming Liang, Beijing Genomics Institute

The pipeline of Hi-C assembly
Version 1
by Kailong Ma

stLFR library construction
Version 1
by lichunhua

RNA and DNA extraction (Qiagen) of frozen tissue to nanopore sequencing
Version 1
by Eric Pederson, Uppsala University

BGISEQ-500 Sequencing
Version 1
by Xinming Liang, Beijing Genomics Institute

Oxford Nanopore sequencing and library construction
Version 1
by zhangrui7