Chromosome-level Reference Genome of the Siamese Fighting Fish *Betta Splendens*, A Model Species for the Study of Aggression

Lulwa Alnassar
Introduction To Genomics
Fall 2020

Outline:

- General information 'fun facts'

- Genome paper

- General genome information

- Sequencing strategy

- Sequencing methods

- Genome assembly

- Genome outcome

General information 'fun facts':

- Aggressiveness

General information 'fun facts':

- Colorful.
- Large and flowy fins

Genome paper:

OXFORD (GIGA)^n SCIENCE

DATA NOTE

Chromosome-level reference genome of the Siamese fighting fish *Betta splendens*, a model species for the study of aggression

Guangyi Fan[1,2,3,4,†], Judy Chan[1,†], Kailong Ma[3,4,†], Binrui Yang[1], He Zhang[2,3,4], Xianwei Yang[2,3,4], Chengcheng Shi[2,3,4], Henry Chun-Hin Law[1], Zhitao Ren[1], Qiwu Xu[2,3,4], Qun Liu[2,3,4], Jiahao Wang[2,3,4], Wenbin Chen[3,4], Libin Shao[2,3,4], David Gonçalves[6], Andreia Ramos[6], Sara D. Cardoso [7], Min Guo[1], Jing Cai[1], Xun Xu [2,3,4], Jian Wang[3,5], Huanming Yang[3,5], Xin Liu [2,3,4,*] and Yitao Wang[1,*]

[1]State Key Laboratory of Quality Research of Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Avenida da Universidade, Taipa, Macau, China, [2]BGI-Qingdao, BGI-Shenzhen, Sino-German Ecopark, No.2877, Tuanjie Road, Huangdao District, Qingdao, Shandong Province, 266555, China, [3]BGI-Shenzhen, Building 11, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China, [4]China National GeneBank, BGI-Shenzhen, Jinsha Road, Dapeng New District, Shenzhen 518120, China, [5]James D. Watson Institute of Genome Sciences, Hangzhou 310058, China, [6]Institute of Science and Environment, University of Saint Joseph, Rua de Londres 16, Macao SAR, China and [7]Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal
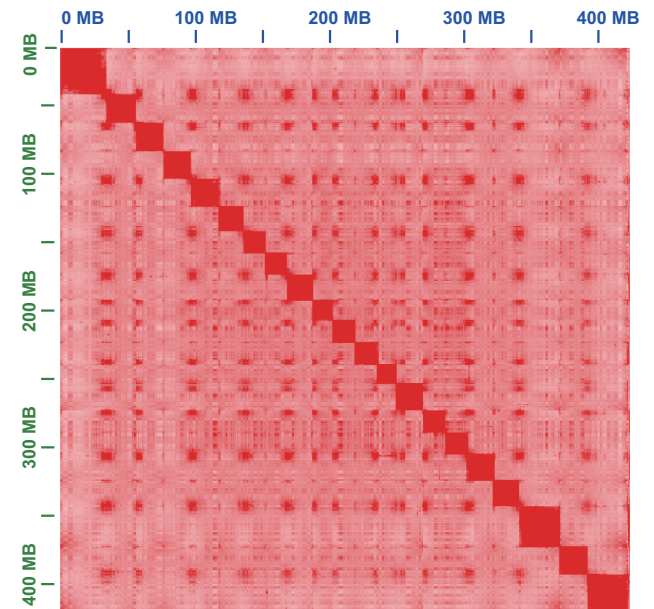
General genome information:

- Easy to keep and reproduce in captivity
- Genome size: 465.24-Mb genome
- Source: 5 male fishes.

Sequencing strategy:

- Whole-genome shotgun.
- DNA and RNA extraction.
- 21 chromosomes using Juicer.

Sequencing methods:

- HiSeq 2000 sequencing platform 'illumina'.
- Illumina sequencing method by synthesis.
- 6 DNA libraries.
- 3 short insert size libraries and 3 mate-pair libraries and 5 RNA sequencing.

Genome assembly:

Table 1: Statistics of the assembly using SOAPdenovo and Hi-C data

Type	Scaffold original	Contig original	Scaffold (Hi-C)	Contig (Hi-C)
Total number	92,886	138,929	91,819	139,323
Total length (bp)	465,240,853	421,527,246	465,132,837	421,527,246
Average length (bp)	5,008.73	3,034.12	5,066	3,026
N50 (bp)	949,032	19,014	19,754,490	18,890
N90 (bp)	59,769	3,504	13,781,534	3,470

Genome assembly:

- 52.34 Gb of clean reads using SOAPnuke.
- N50 scaffold size: 949.03 Kb.
- N50 contig size: 19.01 Kb.
- 75.24 Gb sequenced of data using 'BGISEQ-500 sequencing platform' and obtained 34.5 Gb valid reads (~45.8%).

Genome outcome:

- 15.12% transposable elements in the genome
- Mean coverage ratio of 75.3%.
- Giant: highest frequency of aggressive
- Half-moon: lowest frequency of aggressive

Genome outcome:

- SNV'S: 3.4M
- Indels: 27.305

Thank you for listening..
Any questions?