Exploring cavefish genome

Raghad Al-Refaii

Intro to genomics 485

Fall 2020

Outline

- General information

- Genome paper

- General Genome info

- Sequencing strategy

- Sequencing methods

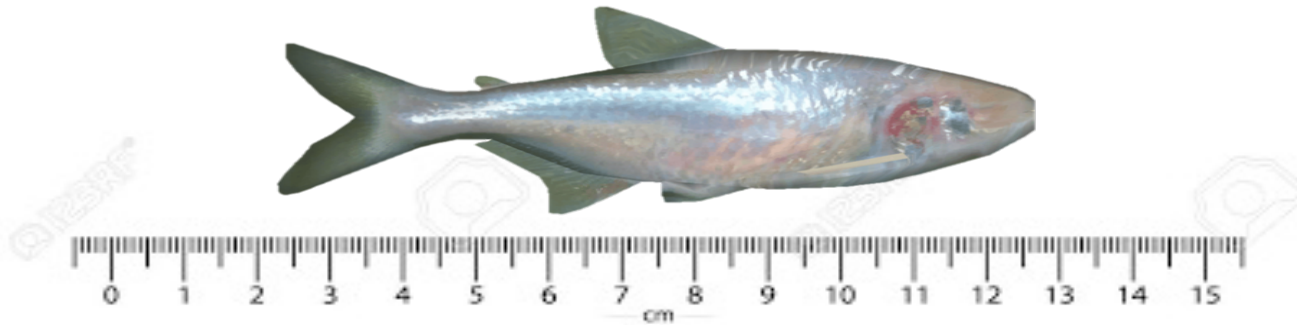- Genome assembly

- Genome outcome

General information

- It is endemic to China's massive southwestern karst area.

General information

- Quite small being between 2 and 13 cm.

General information

- Variation in visuality.

	S. grahami (Sg)	S. rhinocerous (Sr)	S. anshuiensis (Sa)
Species (photo)			
Eyes	Normal	Reduced	Lost

Genome paper

The *Sinocyclocheilus* cavefish genome provides insights into cave adaptation

Junxing Yang[1*†], Xiaoli Chen[2†], Jie Bai[2,3,4†], Dongming Fang[2,6†], Ying Qiu[2,3,5†], Wansheng Jiang[1†], Hui Yuan[2], Chao Bian[2,3], Jiang Lu[2,7], Shiyang He[2,7], Xiaofu Pan[1], Yaolei Zhang[2,8], Xiaoai Wang[1], Xinxin You[2,3], Yongsi Wang[2], Ying Sun[2,5], Danqing Mao[2], Yong Liu[2], Guangyi Fan[2], He Zhang[2], Xiaoyong Chen[1], Xinhui Zhang[2,3], Lanping Zheng[1], Jintu Wang[2], Le Cheng[5,9], Jieming Chen[2,3], Zhiqiang Ruan[2,3], Jia Li[2,3,7], Hui Yu[2,3,7], Chao Peng[2,3], Xingyu Ma[10,11], Junmin Xu[10,11], You He[12], Zhengfeng Xu[13], Pao Xu[14], Jian Wang[2,15], Huanming Yang[2,15], Jun Wang[2,16], Tony Whitten[4*], Xun Xu[2*] and Qiong Shi[2,3,10,11*]

General genome information

|	
Number of chromosomes	25 chromosomes
Source of genome	Muscle tissues
Genome size	Approximately 1.7 Gb
Gene number	Approximately 40.000 genes
Sequencing strategy	whole-genome shotgun
Sequencing method	Illumina
Contig N50	17-29Kb
Scaffold N50	0.9-1.3Mb

Sequencing strategy

A series of sequencing libraries (250 bp to 20 kb) were constructed and applied in a whole-genome shotgun sequencing strategy, and a total of 313.3, 174.0 and 188.2

Gb of raw data were obtained for the Sg, Sr and Sa

fishes, respectively.
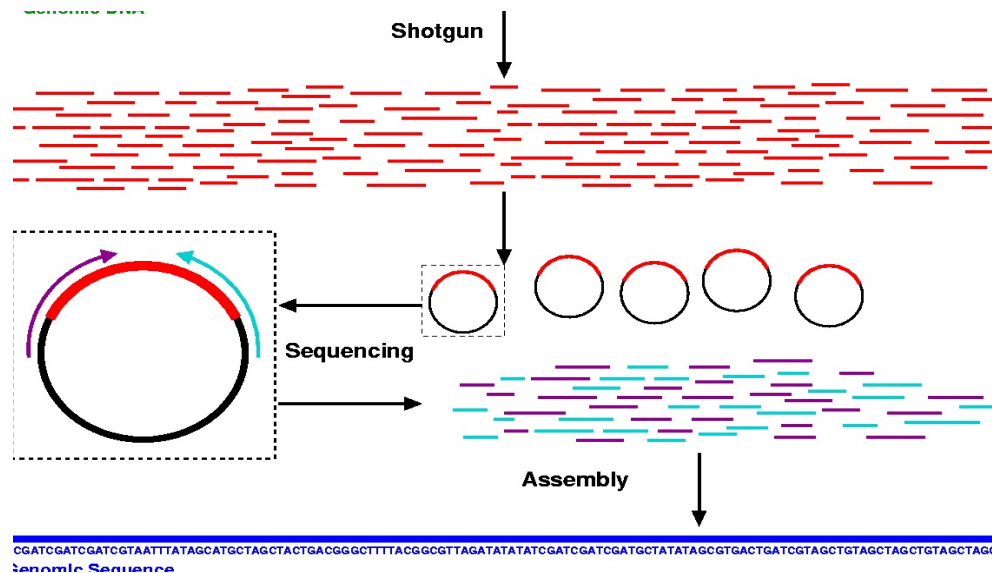
Sequencing methods

- DNA libraries were generated using Illumina
- The reads are between (250bp to 20kp)
- Paired-end sequencing was performed on Illumina for each library.
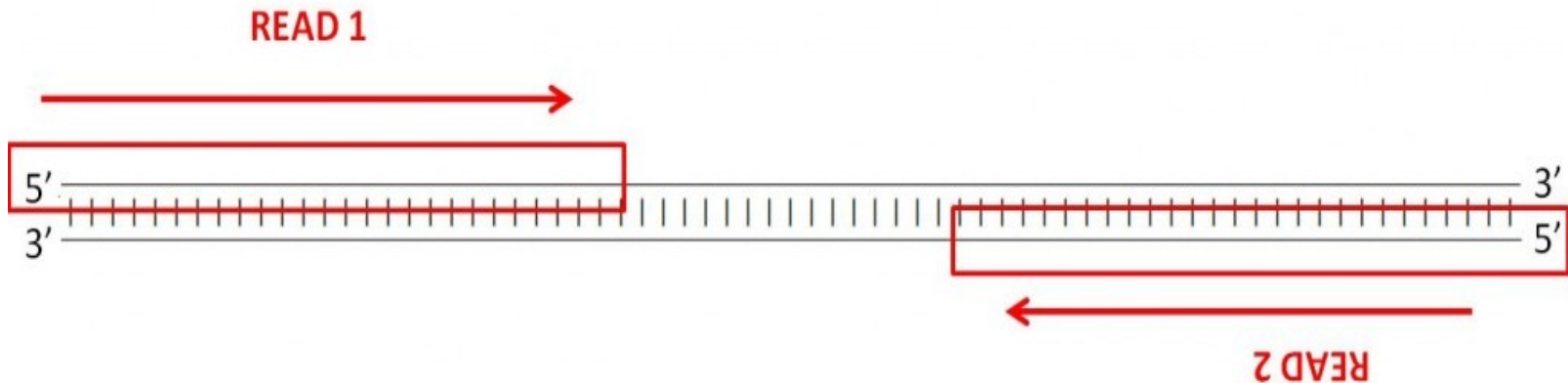
Genome assembly

- The obtained reads were assembled into contigs and Scaffolds with SOAPdenovo assembly.

- The coverage quality of de novo sequence data depends on the size and continuity of the contig.
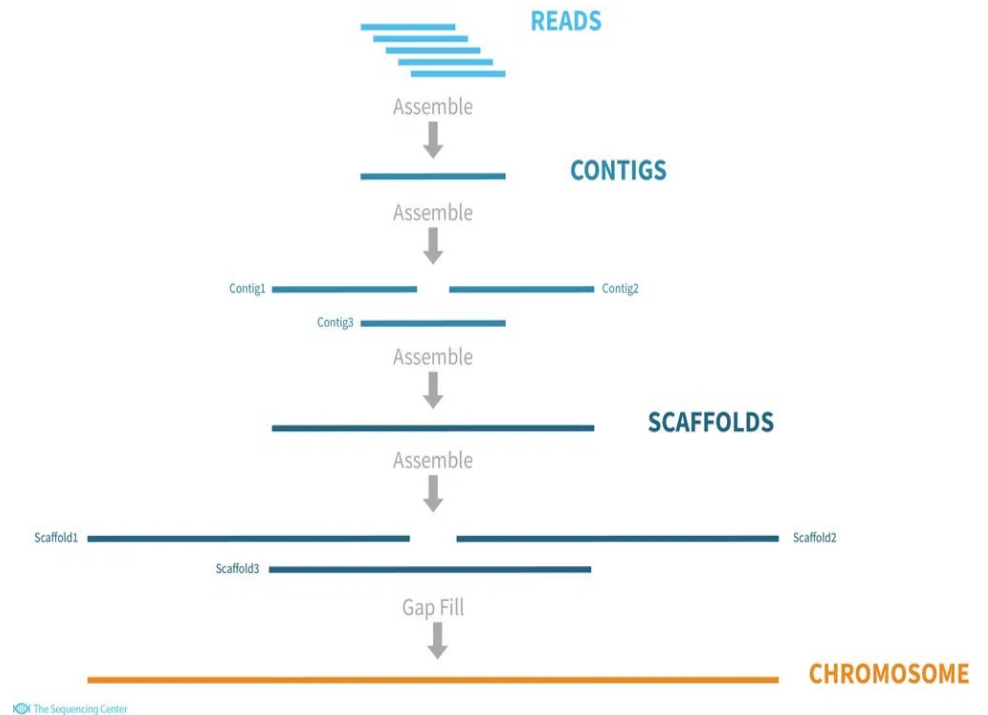
Genome Outcome

	S. grahami (Sg)	S. rhinocerous (Sr)	S. Anshuiensis (Sa)
N50 contig	29,335 bp	17,658 bp	16,708 bp
N50 scaffold	1.16 Mb	896.4 kb	1.25 Mb
sequence coverage/depth	48.57	42.3	47.68

Questions

- How is the eye loss of S. rhinocerous linked to their evolution and adaptation to caves?

- What is meant by karst area?