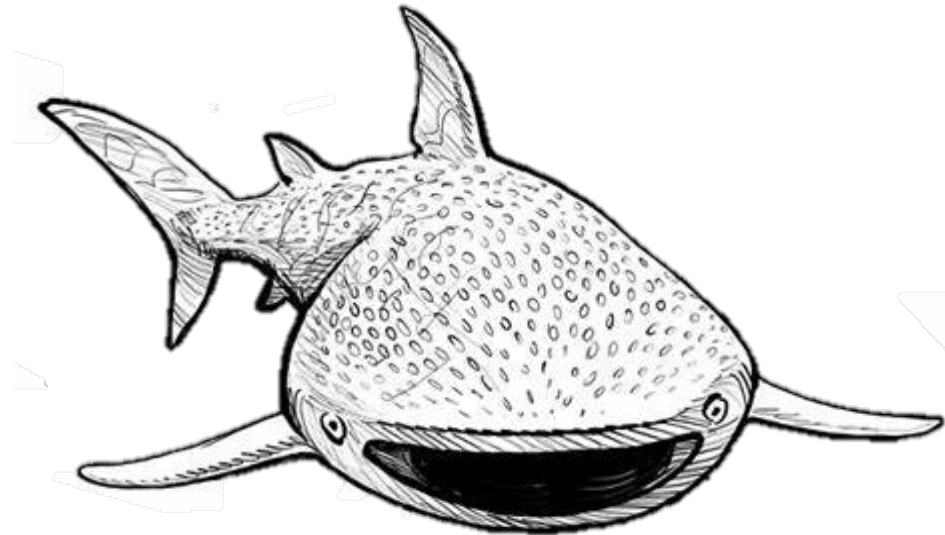


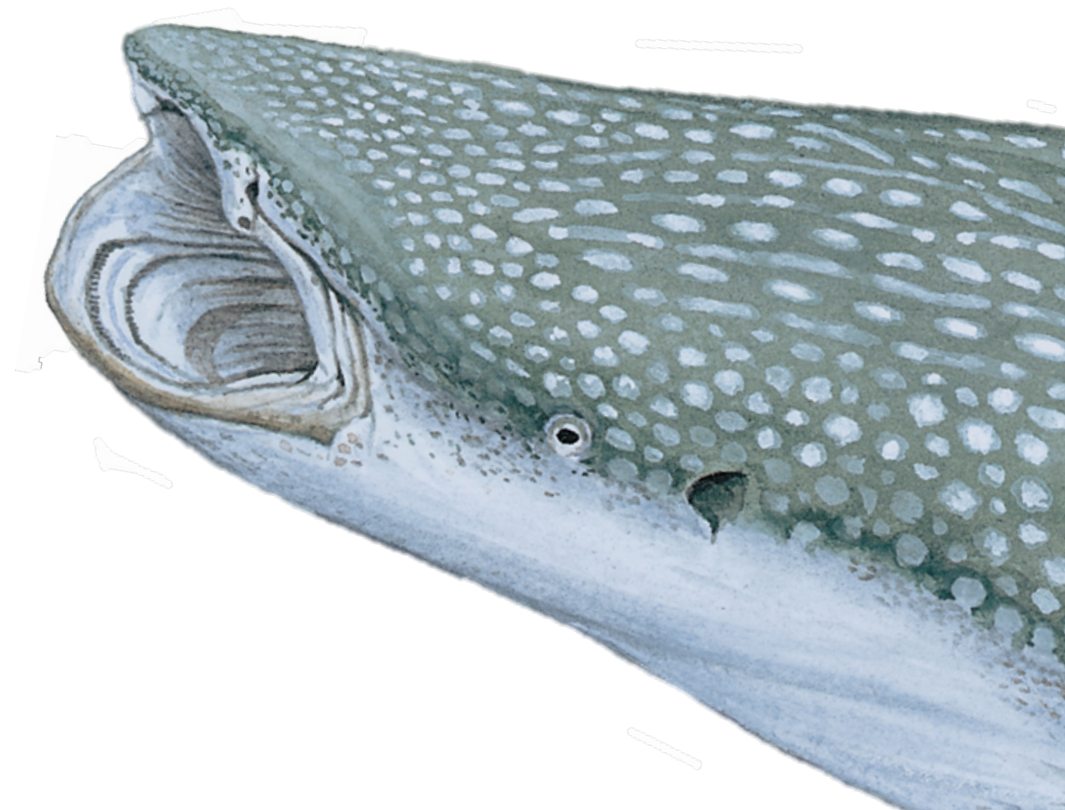
Sequencing and assembly of
the
genome of
whale shark
Rhincodon typus



Aisha Al-Failakawy
Intro. To Genomics 485
Fall 2020

outline

- General information(fun facts)
- Genome paper
- General genome info
- Sequencing strategy
- Sequencing methods
- Genome assembly
- Genome outcome



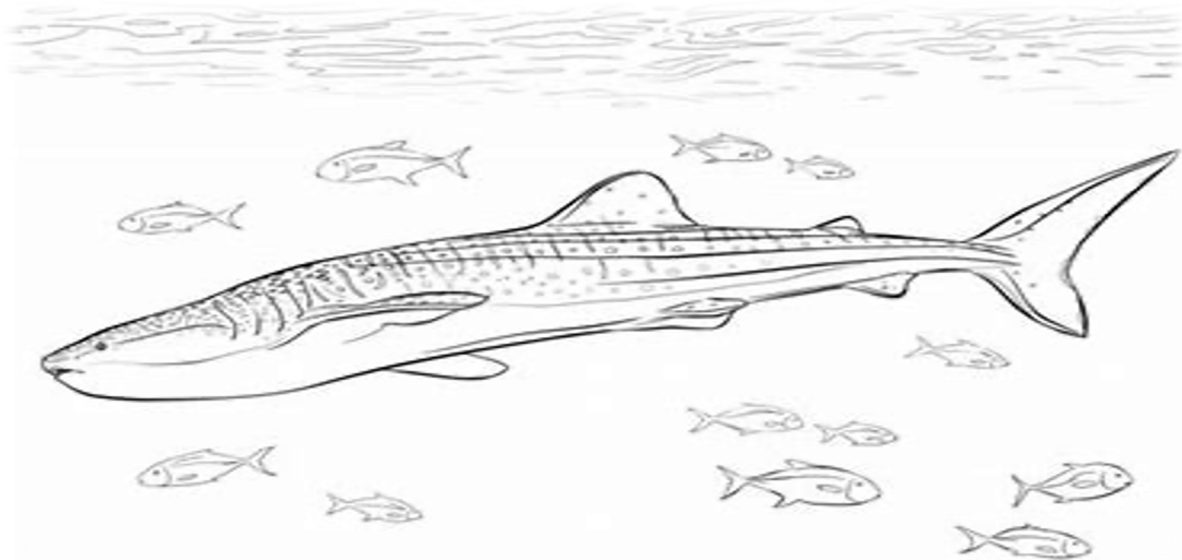
General information (fun facts)

- largest shark in the worlds
60 tons (weight)
18.8 m (length)



General information (fun facts)

- found in warm oceanic waters
- capable of diving to depths where waters approach freezing Temperatures.



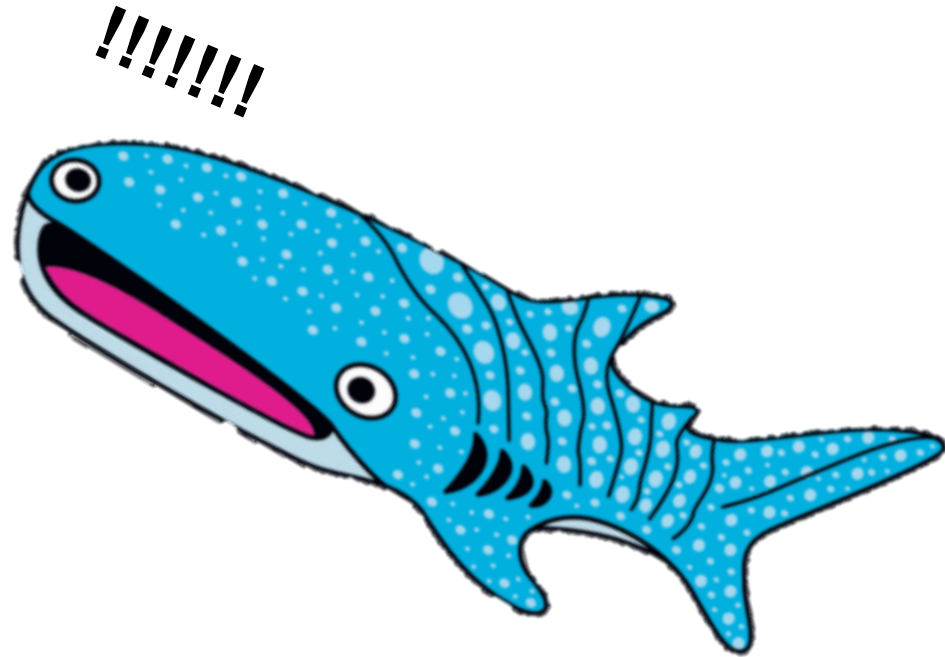
General information (fun facts)

- Peaceful with humans
- Easy to docile
- Swimmers are allowed to ride



General information (fun facts)

- Currently considered an endangered species



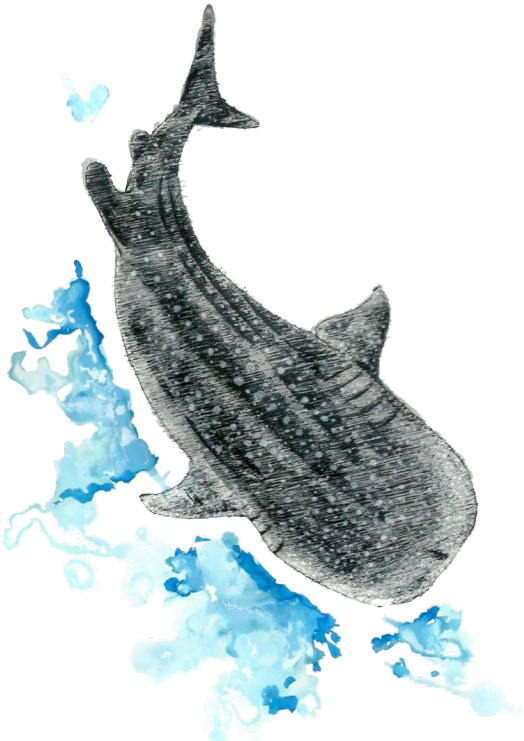
General information (fun facts)

- Contain over 300 row of tiny teeth
- 20 filter pads which it uses to filter feed



General information (fun facts)

- Males grow faster than females in the earlier stages of life
- Female larger in size



Genome paper

Read et al. *BMC Genomics* (2017) 18:532
DOI 10.1186/s12864-017-3926-9


BMC Genomics

RESEARCH ARTICLE

Open Access



Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* Smith 1828

Timothy D. Read^{1,2}, Robert A. Petit III^{1,2}, Sandeep J. Joseph^{1,2}, Md. Tauqeer Alam^{1,2}, M. Ryan Weil^{1,2}, Maida Ahmad^{1,2}, Ravila Bhimani^{1,2}, Jocelyn S. Vuong^{1,2}, Chad P. Haase^{1,2}, D. Harry Webb³, Milton Tan^{1,2*} 
and Alistair D. M. Dove³

General genome info

- There are few publications on the genetics and genomics of whale sharks.
- The genome sequence was derived from tissue samples from a male whale shark .
- The number of chromosomes has not yet been achieved
- Genome size : 3.44 Gb

Sequencing strategy

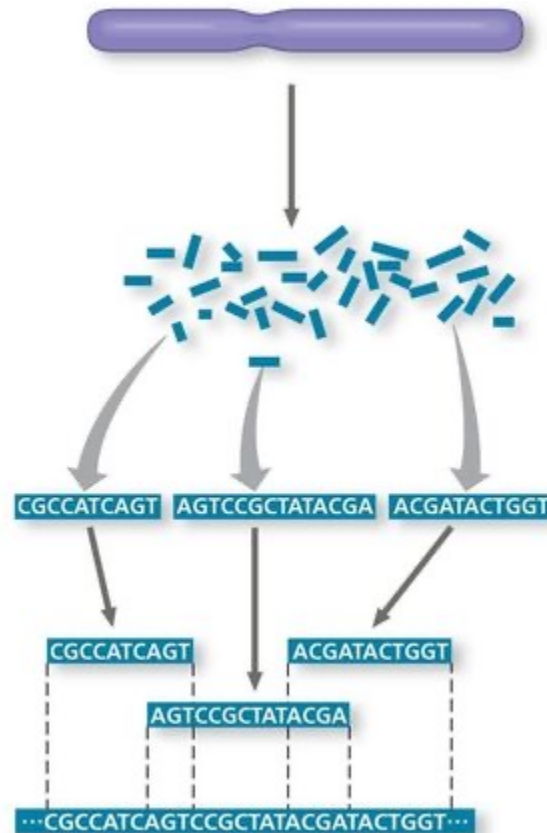
➤ Whole Genome Shotgun Sequencing

1 Cut the DNA into overlapping fragments short enough for sequencing.

2 Clone the fragments in plasmid or other vectors.

3 Sequence each fragment.

4 Order the sequences into one overall sequence with computer software.



Sequencing methods

➤ Illumina sequencing method

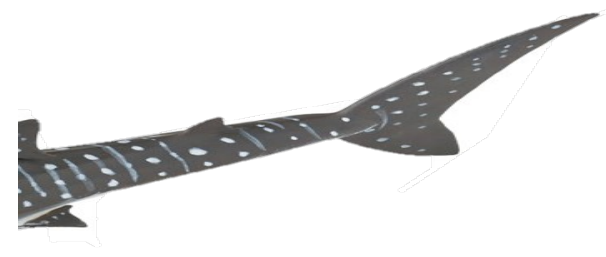
➤ Pyrosequencing method

Genome assembly

Attribute	Value	% of Total
Genome size (Gbp)	3.44	
DNA coding (bp)	10,400,226	0.41%
DNA G + C (bp)	1,059,229,091	41.3%
Number of scaffolds	997,976	
Scaffold N50 (bp)	5425	
Number of contigs	1,213,000	
Contig N50 (bp)	5304	
Protein coding genes	19,384	
Genes with function prediction	5380	27.8%
Genes assigned to KOGs	7038	36.3%
Genes with Pfam domains	6612	34.1%



Genome outcome



- The genome sequence of an organism is now perhaps the single most important gateway to understanding its biology.
- The current DNA sequence can be mined for new genotyping tools for population genomics and the protein set can be compared intensively against known functions.
- The long term goals include understanding the genetic nature of the large body size of the whale shark, its metabolic adaptations to its planktonic diet, and the evolution of its immune system in a comparative.
- We pursued a strategy of primarily using cost-effective Illumina short read sequencing to produce a preliminary *R. typus* genomic dataset. . This allowed us to maximize coverage of the genome with high quality data and give estimates of the genome size and extent of bacterial contamination of the source DNA and to provide what we believe is a quite complete, if fragmented, draft of the genome.
- Future work will enhance the whale shark genome assembly using long reads using the Pacific Biosciences technology.