

# Project update and summary

Nov 29, 2016



# Project Scope

- Project F15FTSUSAT0328 – *Panopea Generosa*
- Project F15FTSUSAT0327 – *Ostrea lurida*
- Genome size of species is unknown, and estimated at 500Gb
- Genome survey and preliminary assembly

# Work that has been done

- Reads filtering and cleaning up
  - Tool: SOAPnuke

		Panopea(M)	Ostrea(M)
Reads Number	49PE	719.57	898.069
	150PE	489.066	327.896
Bases Number	49PE	35258.9	44005.4
	150PE	73359.9	49184.3

- Kmer analysis and Genome size prediction
  - Tool: Kmerfreq\_v5.0
  - Genome Size

	Panopea	Ostrea
genome_size	2.97Gb	1.90Gb

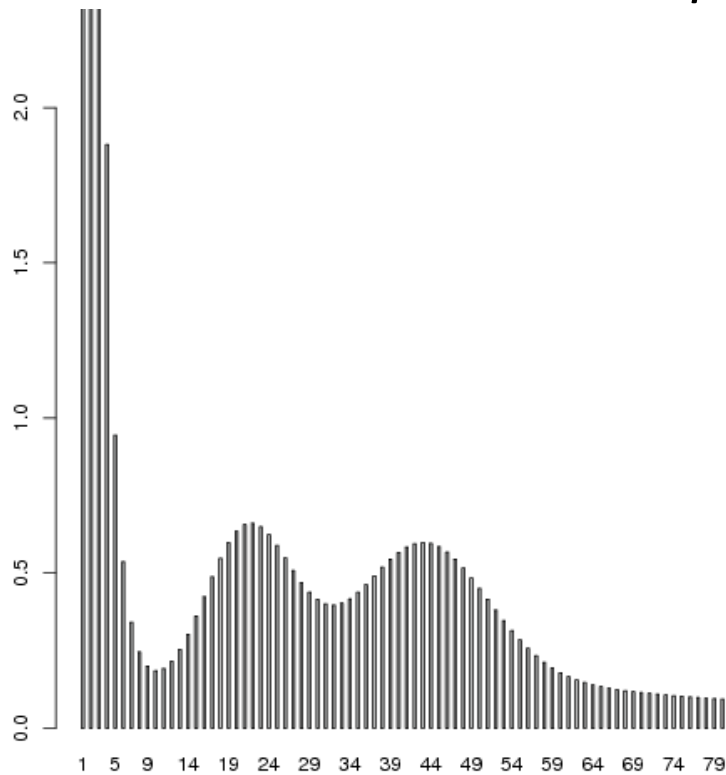
# Work that has been done

- Heterozygous rate prediction
  - Tool readsim\_v2.2
  - 48 parameter combinations have been tested, and compared to actual Kmer data.
  - *Ostrea*: False ratio-0.001; Repeat ratio-0.15;Hybrid ratio-0.005;
  - *Panopea*: False ratio-0.001;Repeat ratio-0.15;Hybrid ratio-0.020.
- Preliminary de-novo assembly – repeated 3-4 times.
  - Tool: Soapdenovo;
- GC-content and depth analysis

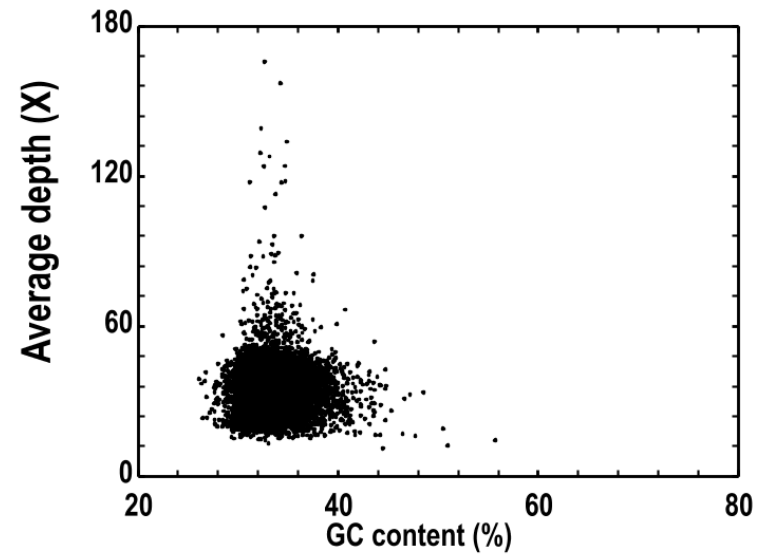
## *Panopea Generosa*

	scaffold	contig
Max length (bp)	154,899	44,528
N50 (bp)	4,432	1,357
Total length (bp)	1,302,267,261	1,083,430,729
Number>=100bp	129,6135	195,6322
Number>=2000bp	163,983	111,859
GC Rate	0.286	0.337

## *Panopea Generosa*



17mer depth distribution

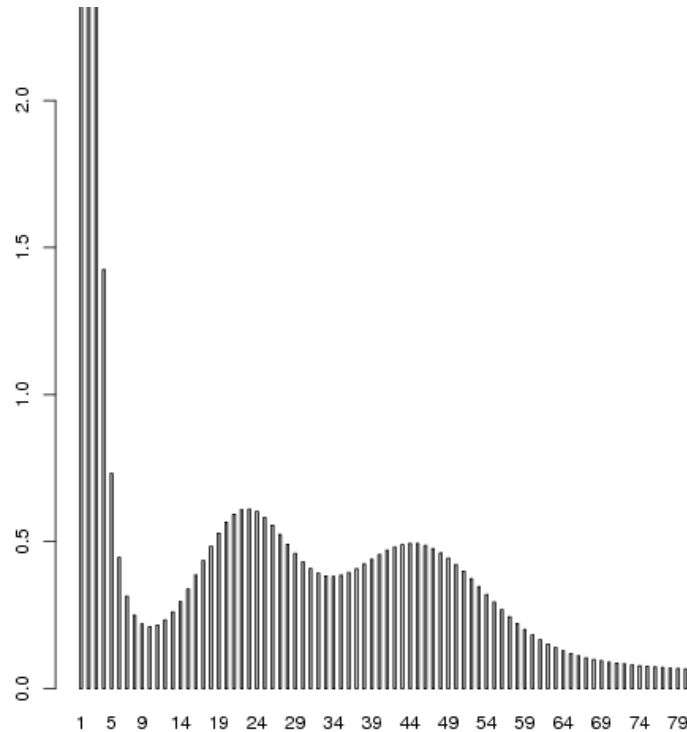


Distribution of GC depth  
Normal pattern, no contamination

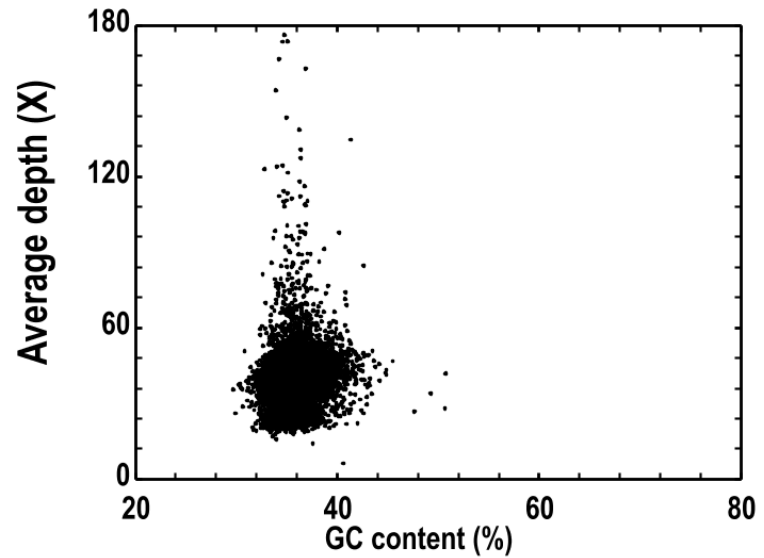
## *Ostrea lurida*

	scaffold	contig
Max Length (bp)	130,073	27,058
N50 (bp)	3,635	1,315
Total Length (bp)	776,306,190	627,311,244
Number>=100bp	765,755	1,135,869
Number>=2000bp	107,343	63,776
GC Rate	0.295	0.358

## *Ostrea lurida*



17mer depth distribution



Distribution of GC depth  
Normal pattern, no contamination



# Future work

- Major difficulties
- Higher genome complexity requires deeper sequencing coverage and more mate-pair libraries;
- Gap-closing has been conducted 2-3 times for each species.
- Repeat this procedure will slightly improve result, may also introduce inaccurate assembly.
- Gap-closing requires special computational notes with large memory 100-300GB
- TAT – 1-3 months per sample  
Time: 1-3 month/sample
- Option – 1) Close this project with 50% assembly fee (\$2500 per projects) waived. 2) We finish assembly work, and add gene structure prediction and functional annotation for free. 3) We make 1-2 run of gap closing work.

# Usage of the data generated

- Design PCR primer of interested mutations;
- Making genome comparison with other species;
- Use it as reference to do population study;