Draft genome of the Hong Kong oyster, *Crassostrea hongkongensis*

Ziniu Yu

South China Sea Institute of Oceanology,

Chinese Academy of Sciences (CAS), Guangzhou, China

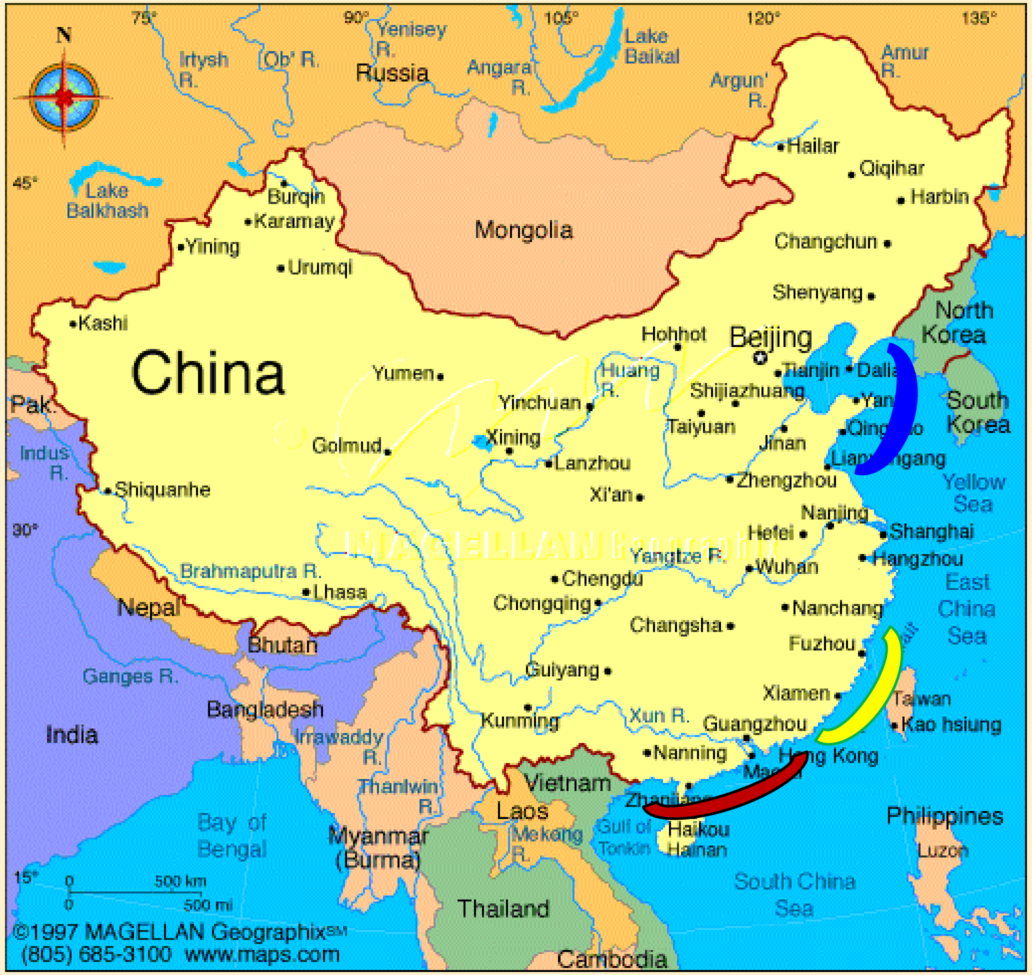
Draft genome of the Hong Kong oyster, *Crassostrea hongkongensis*

Hong Kong oyster *Crassostrea hongkongensis*

- The dominant oyster species along **coastal waters of South China Sea**, a ever-growing cultivation industry for several decades, with anuanl landing of ～1.4 million tons;

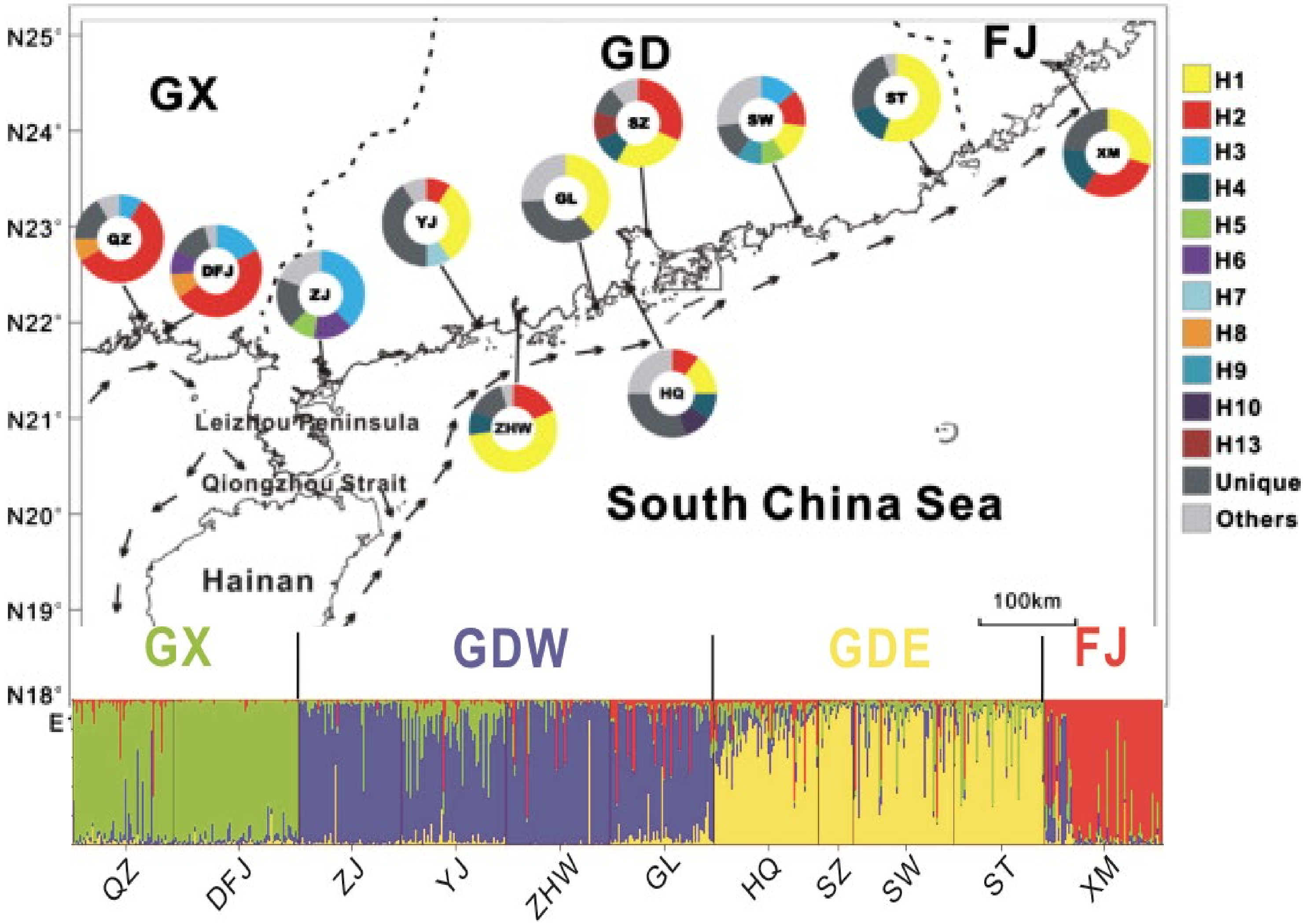
- A good species for comparative genomics with other species like *C. gigas*, in the aspect of speciation, production traits and adaptation to coast stresses;

- A species which requires for environmental bio-marker development for monitoring heavy metals, pollutants, antibiotics, microplastics and so on in estuarine region.



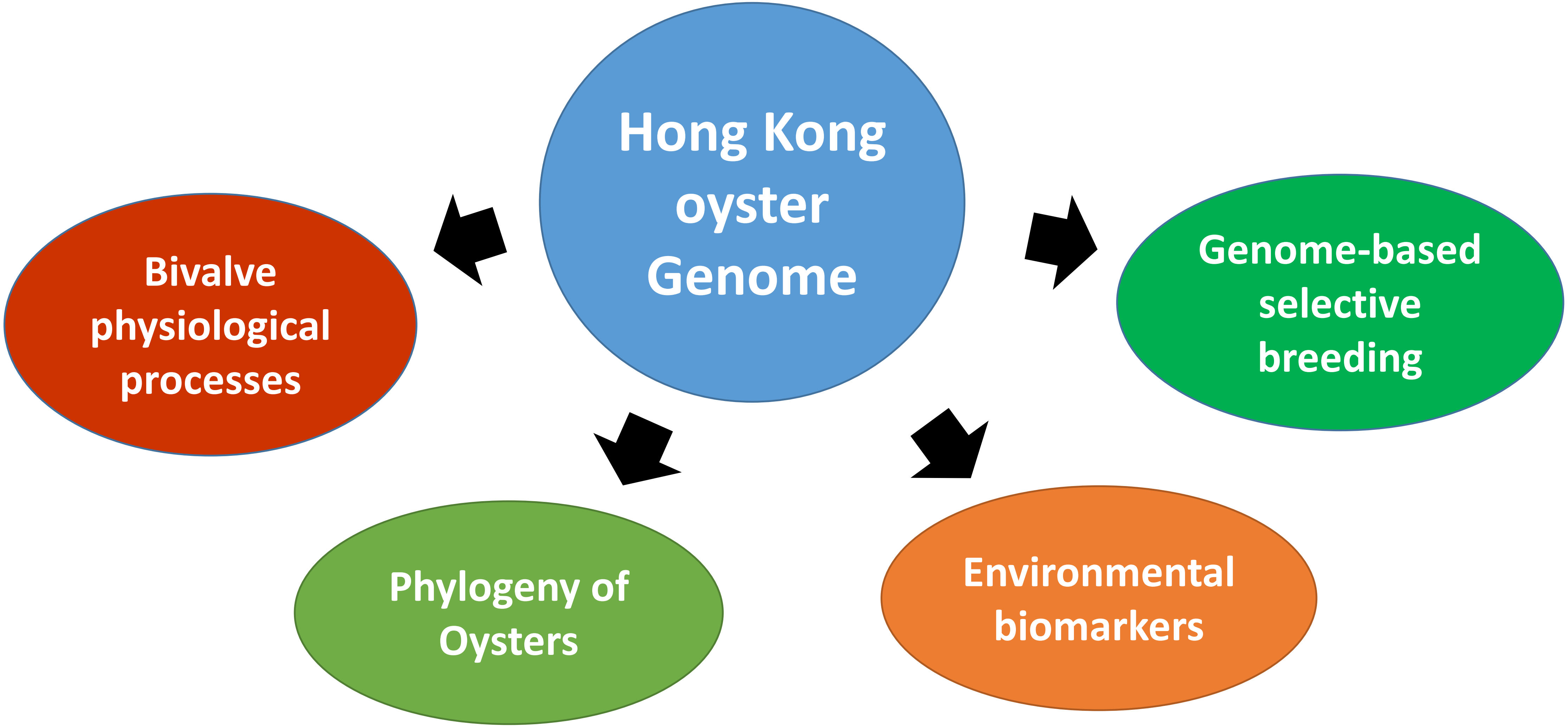

横琴蚝

Genetic variations of *Crassostrea hongkongensis*



Genetic variations in populations of *C. hongkongensis* using cox1 haplotype and microsatellite data

The purposes of genome project of Hong Kong oyster



Hong Kong oyster Genome

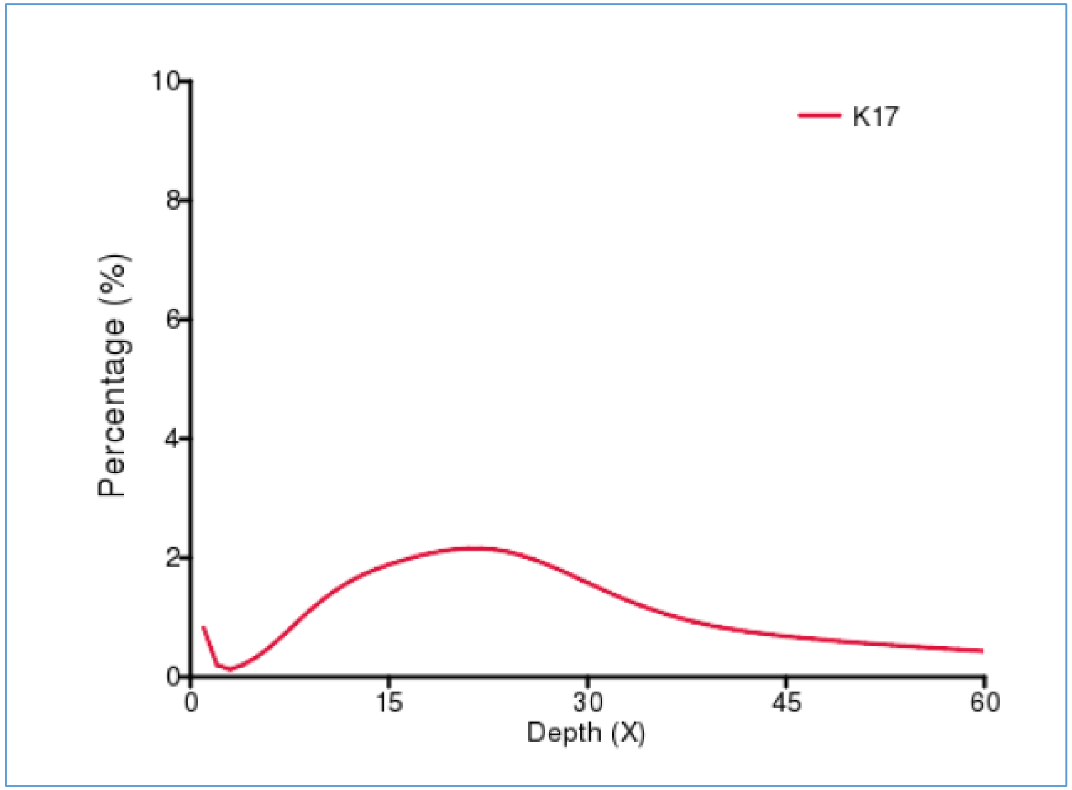
Bivalve physiological processes

Genome-based selective breeding

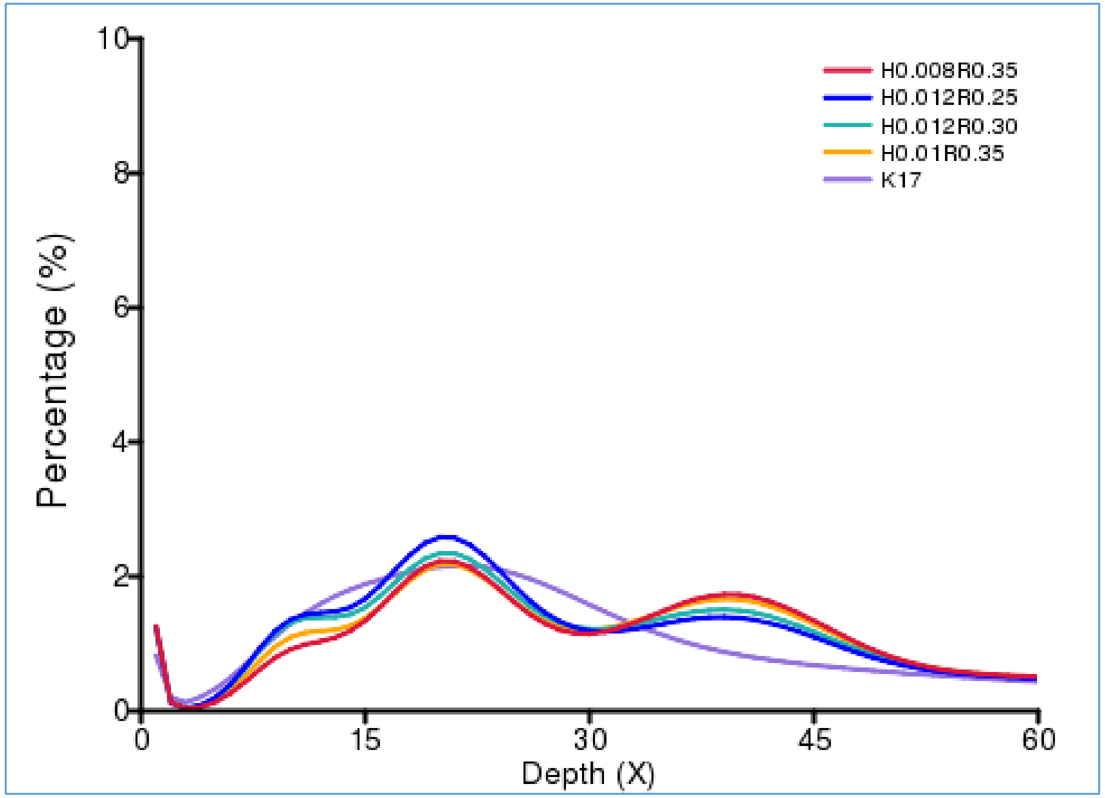
Phylogeny of Oysters

Environmental biomarkers

Genome survey of Hong Kong oyster



Distribution of 17-mer

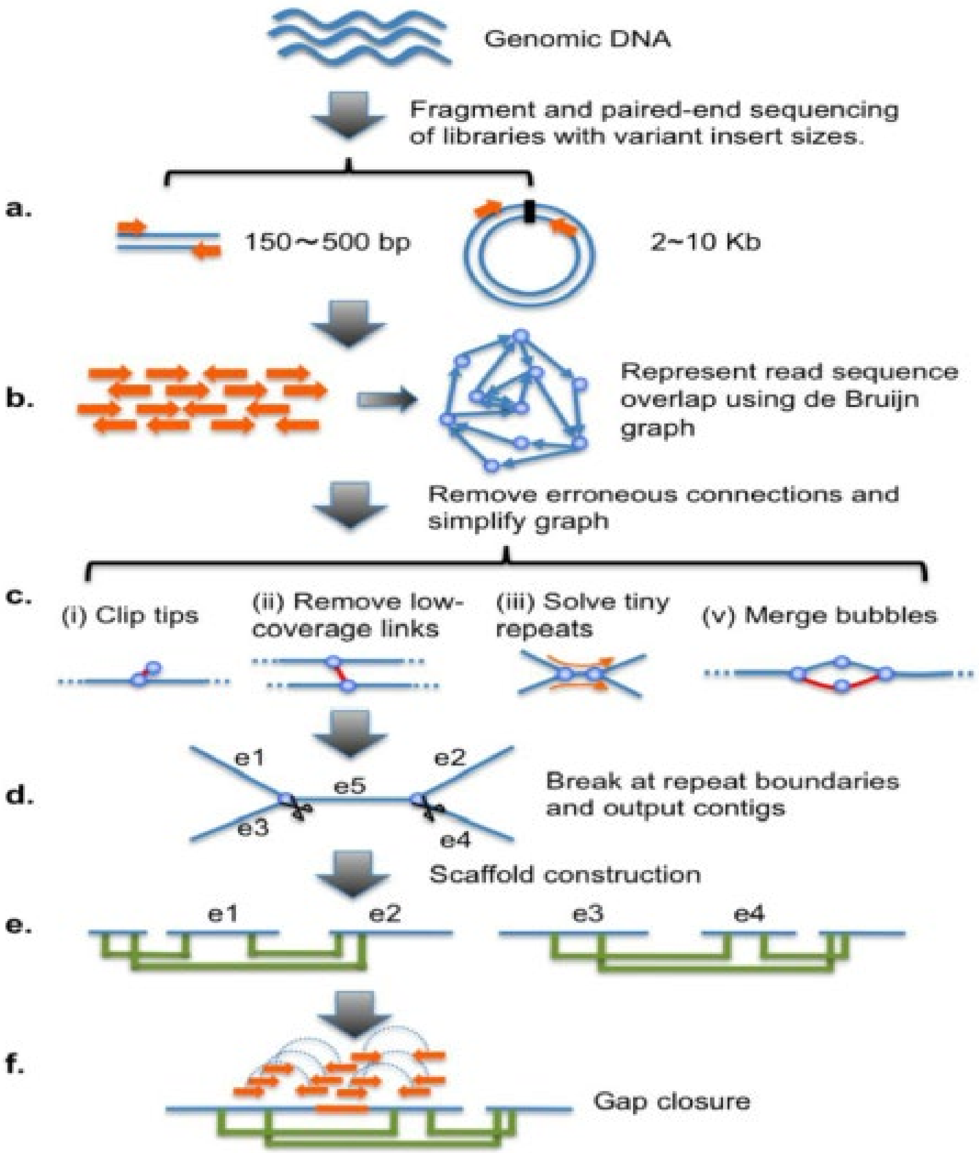


Heterozygosity estimation

The estimated genome size is 695M based on methods of K-mer, and heterozygosity is about 1.2%.

Strategy for genome sequencing and assembly

Flowchart for soap de novo



Next Generation Sequencing (NGS) Technology

Short fragment Library
200bp, 500bp

Long fragment Library
3Kb,4Kb,5Kb,8kb,10kb,15kb

libraries construction for Hong Kong oyster genome

The 147.25Gb raw data with coverage depth of 207.18X;

Average Q20 and Q3 reaching 92.85% and 87.25%, respectively.

Statistics of Genome assembly

Scaffold number	Scaffold length (bp)	Scaffold N50 (bp)	Scaffold N90 (bp)	Scaffold max (bp)	Gap total length (bp)
7,509	714,880,424	618,244	94,812	8,764,835	54,234,013
Contig number	Contig length (bp)	Contig N50 (bp)	Contig N90 (bp)	Contig max (bp)	GC content (%)
60,159	660,646,411	20,341	5,047	207,568	33.16

The whole length of genome was 714.88Mb, with N50 of Scaffold and Contig of 618.24Kb and 20.34Kb, respectively.

Evaluation of genome assembly

Range of length	Total number	Aligned number	Percentage (%)
All	22,826	22,602	99.02
>=500	21,540	21,199	98.42
>=1,000	13,572	13,328	98.20

Statistics of coverage rate of intragenic region

Evaluation of the precision of genome ---- single base error rate

Contig length (bp)	Correct base number (bp)	Error base number (bp)	Error base percentage (%)
660,646,411	660,642,356	4,055	0.0006%

Statistics of single base error rate of genome

Annotation of repetitive sequence

Type	Number	Length(bp)	Percentage (%)
ClassII/Helitron/Helitron	274,681	84,162,463	11.77
ClassII/Helitron/Helitron	274,681	84,162,463	11.77
ClassII/Crypton/Crypton	262,342	59,159,013	8.28
ClassI/LARD/?	137,873	34,941,118	4.89
ClassII/?/?	104,497	27,526,483	3.85
ClassI/PLE/Penelope	90,636	24,312,374	3.4
ClassII/MITE/?	92,846	24,047,069	3.36
ClassII/MITE/?	92,846	24,047,069	3.36
ClassII/TIR/?	47,437	23,421,755	3.28
ClassI/LTR/Gypsy	53,822	21,977,116	3.07
ClassII/TIR/Tc1-Mariner	72,301	18,318,946	2.56
ClassI/DIRS/DIRS	30,322	10,801,892	1.51
ClassI/LINE/RTE	30,303	8,980,258	1.26
Total:	1,513,619	415,712,330	**58.15**

Statistics showed that repetitive sequences occupied 58.15% of whole genome

Prediction of protein coding gene

Method	Software	Gene number
Ab initio	Augustus	32,267
	GeneID	49,203
	SNAP	66,338
	Genscan	22,014
	GlimmerHMM	69,398
Unigene	GMAP	17,294
	PASA	3,205
Homology-based	GeneWise	30,943
Integration	GLEAN	35,624

the statistics of gene prediction

Prediction of non-coding RNA and psuedogene

the statistics of non-coding RNA

RNA classification	Number	Family
miRNA	807	535
rRNA	83	2
tRNA	154	45

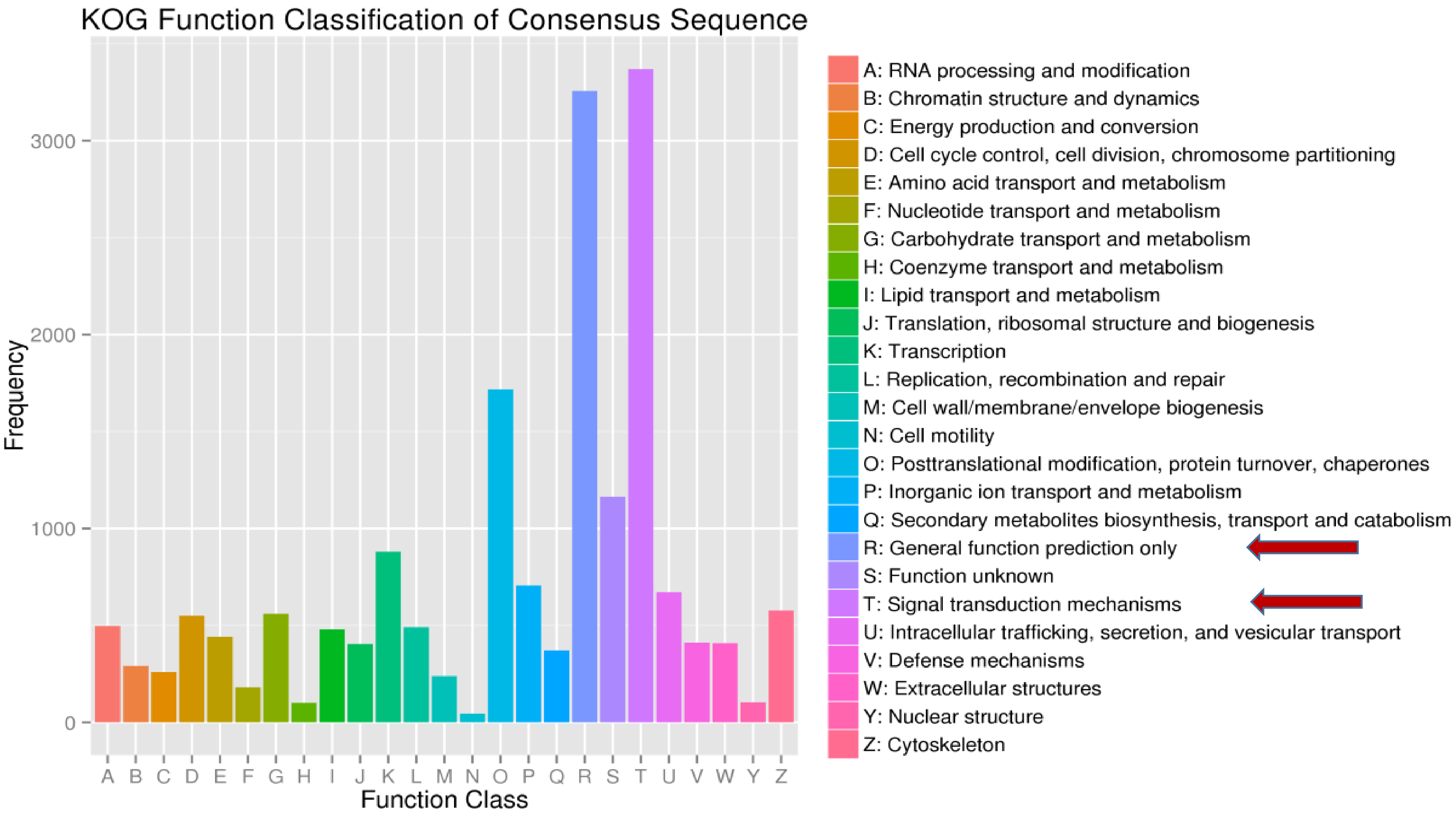
the statistics of pseudogene

Software	Number
GeneWise	2,607

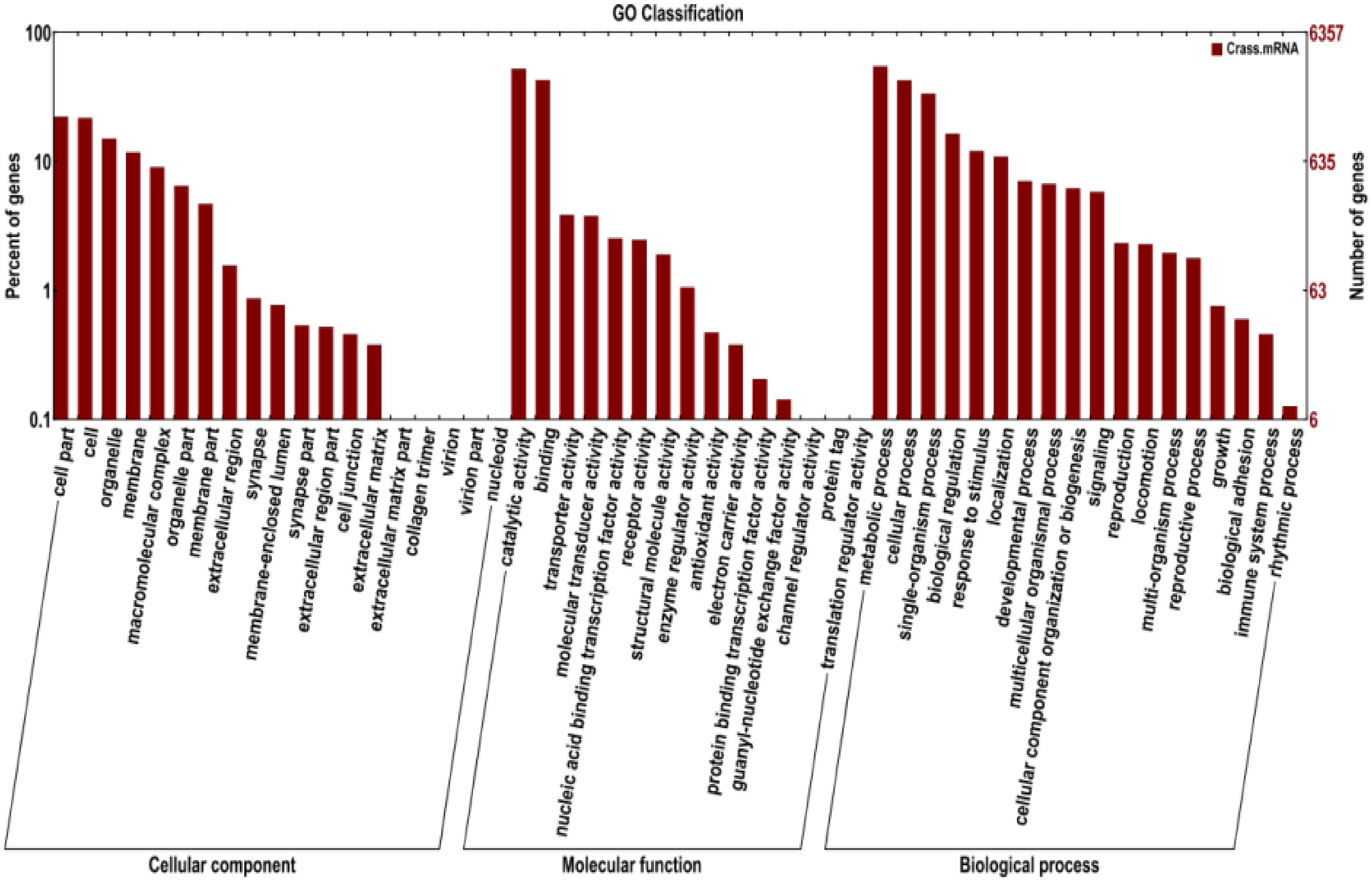
Statistics of gene function annotation

Annotation database	Annotated number	Percentage (%)
KOG	15,913	44.67%
GO	6,357	17.84%
KEGG	7,290	20.46%
TrEMBL	32,297	90.66%
NR	32,298	90.66%
All Annotated	32,362	90.84%

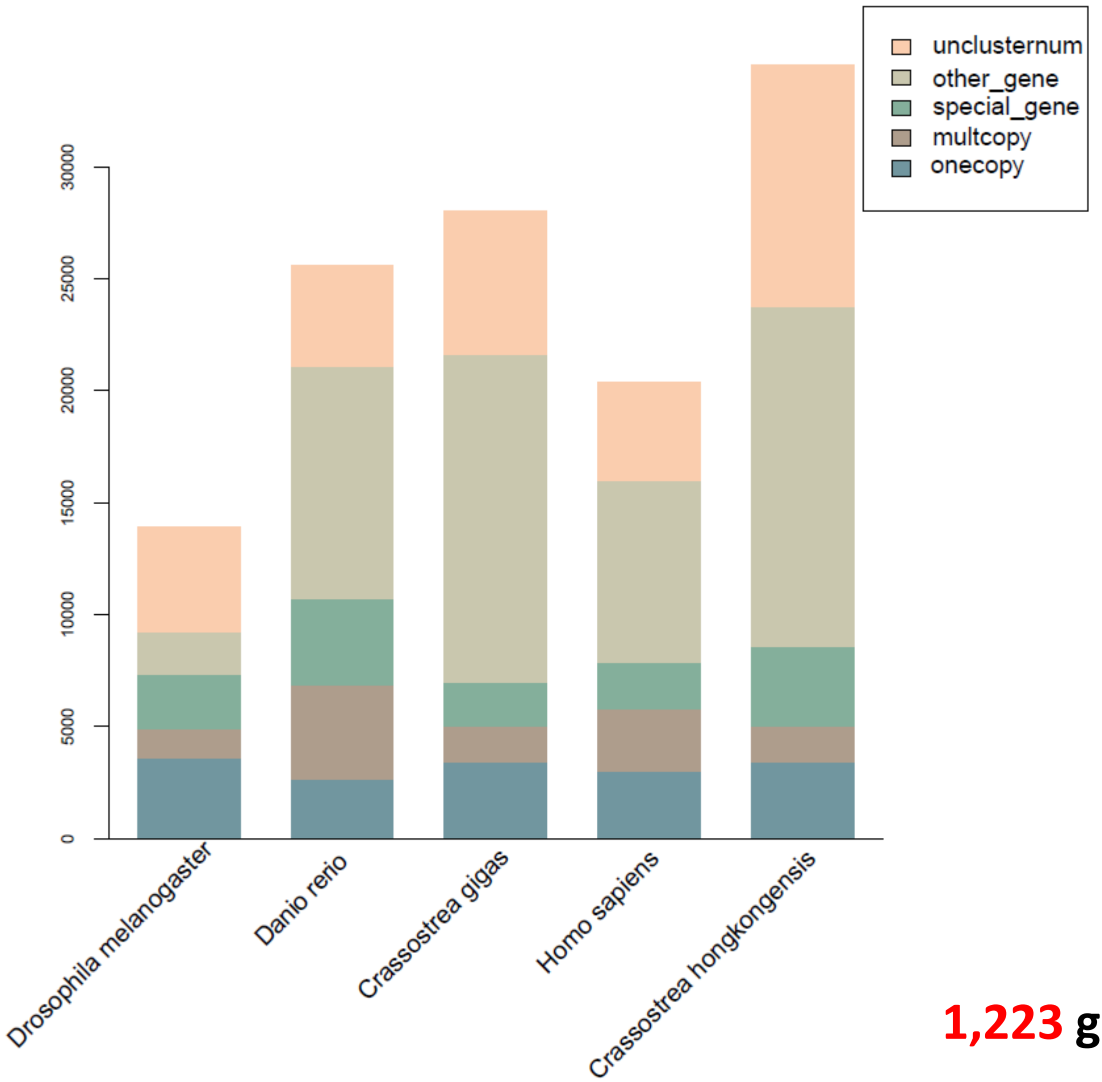
Statistical diagram of KOG function annotation

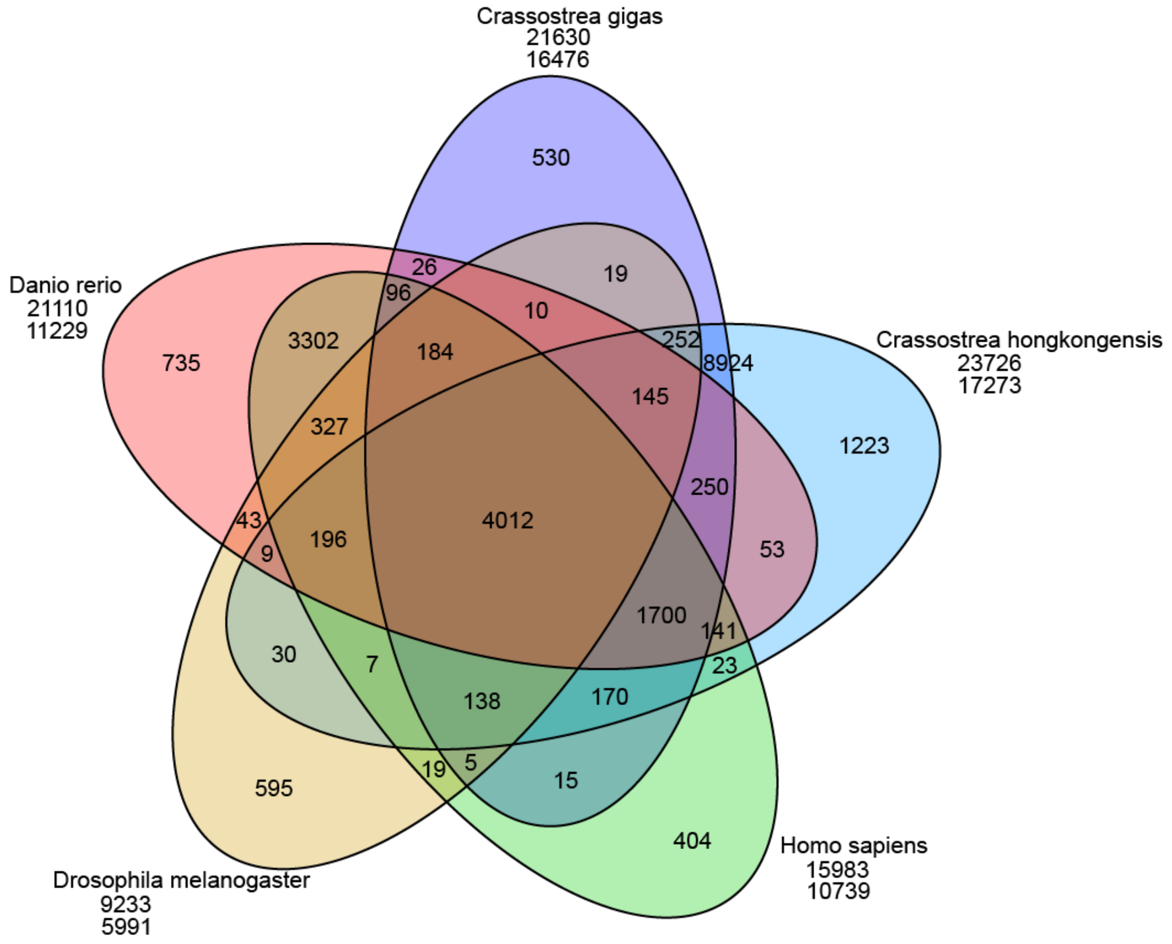


GO annotation



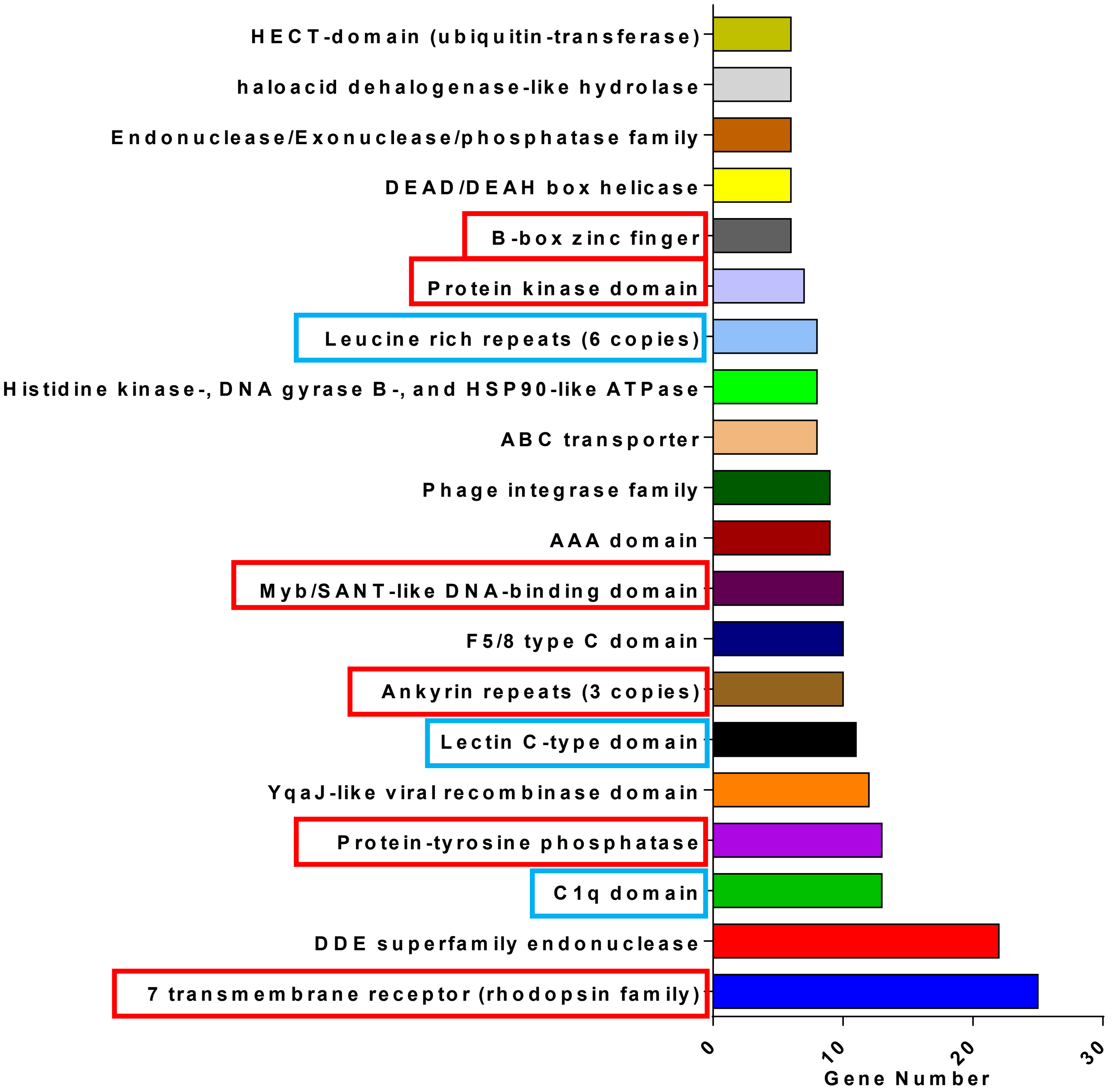
Comparative genomics Cluster of gene family





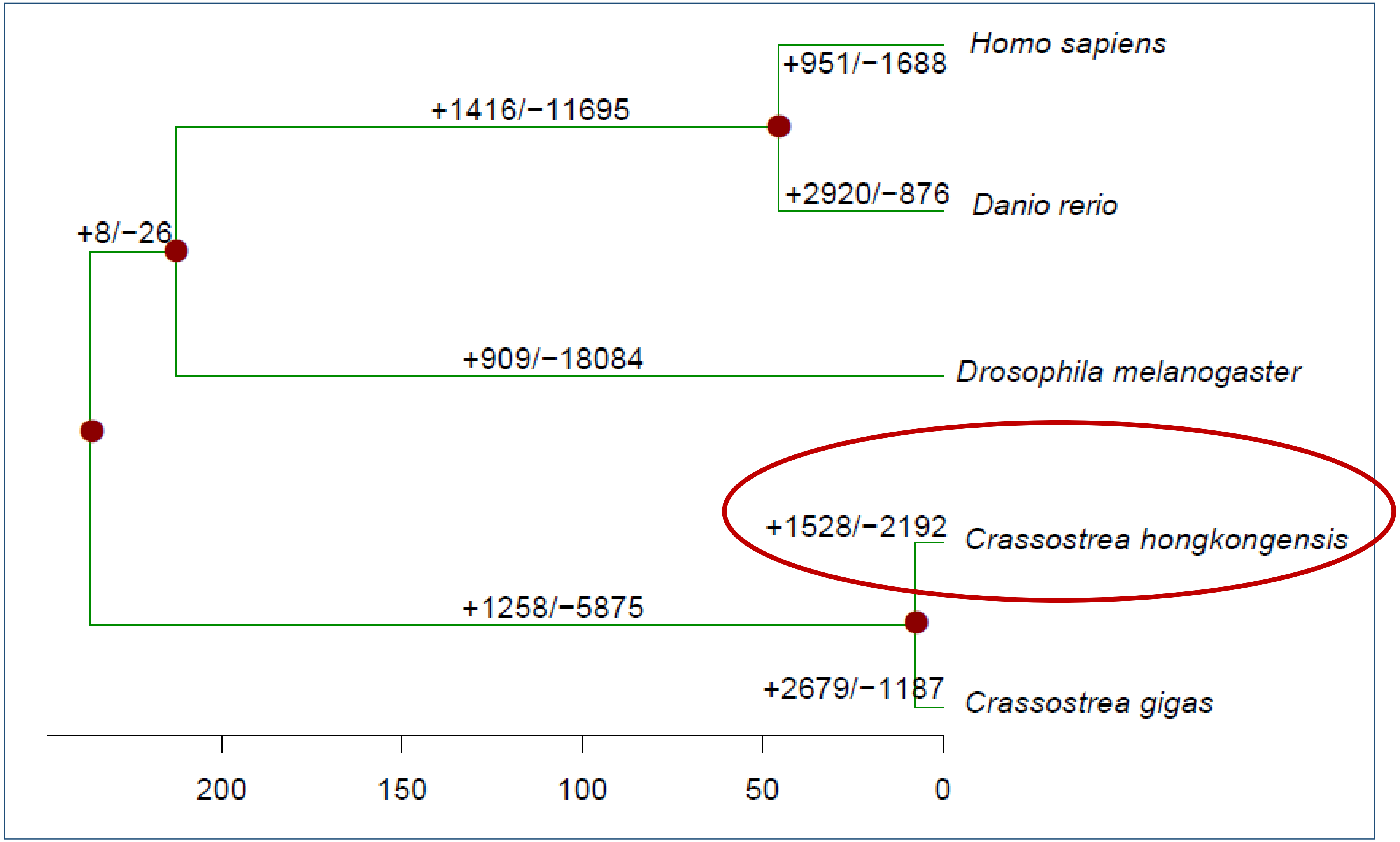
1,223 gene families unique to *C. hongkongensis* were identified

Top 20 specific gene family from 1223 unique ones of hongkong oyster

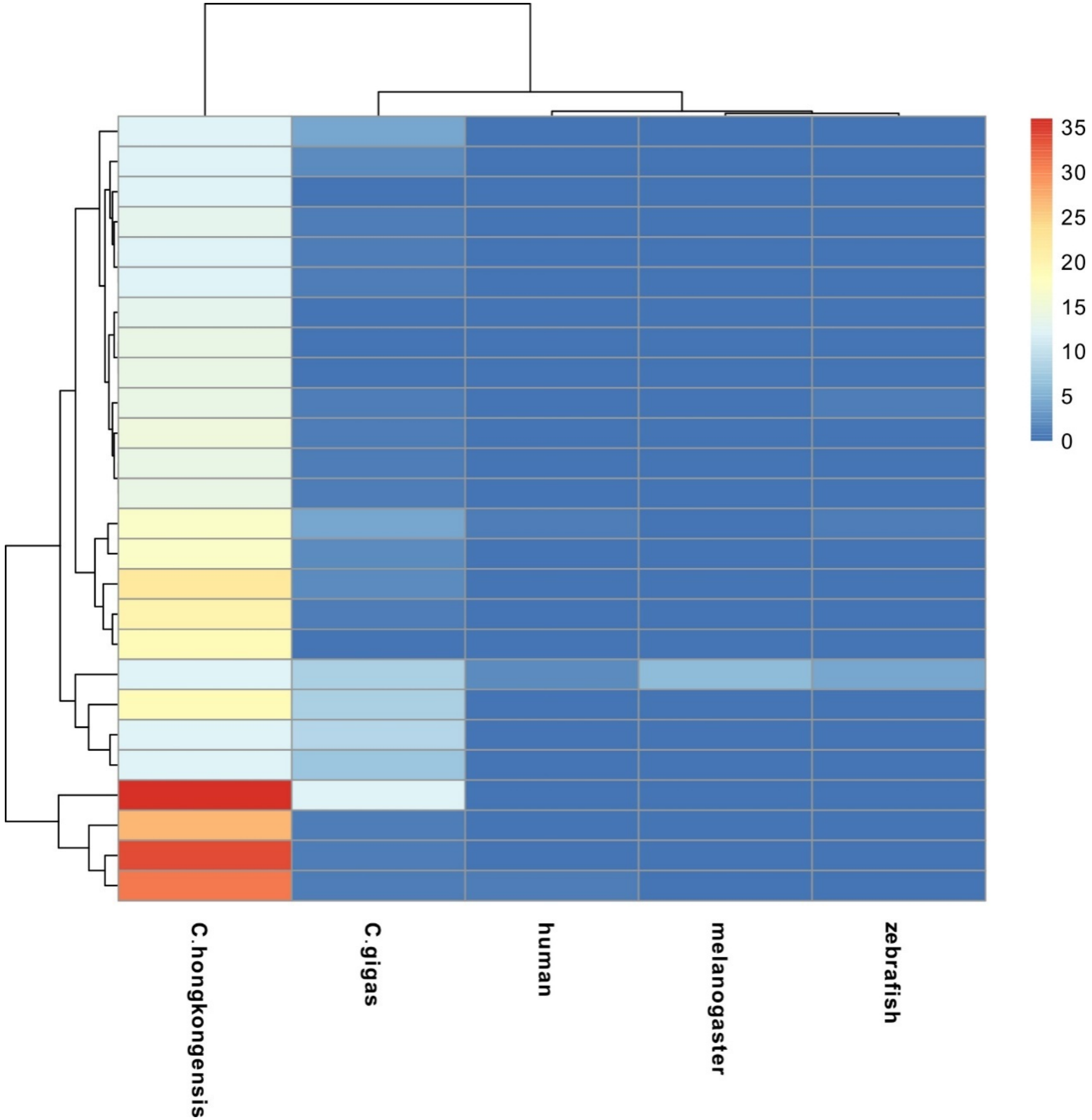


Substantial Hongkong oyster specific genes are involved in signaling transduction and immune response.

Gene family expansion and contraction



Specific family expansion in hong kong oyster



Salinity regulation:

Polycystin cation channel

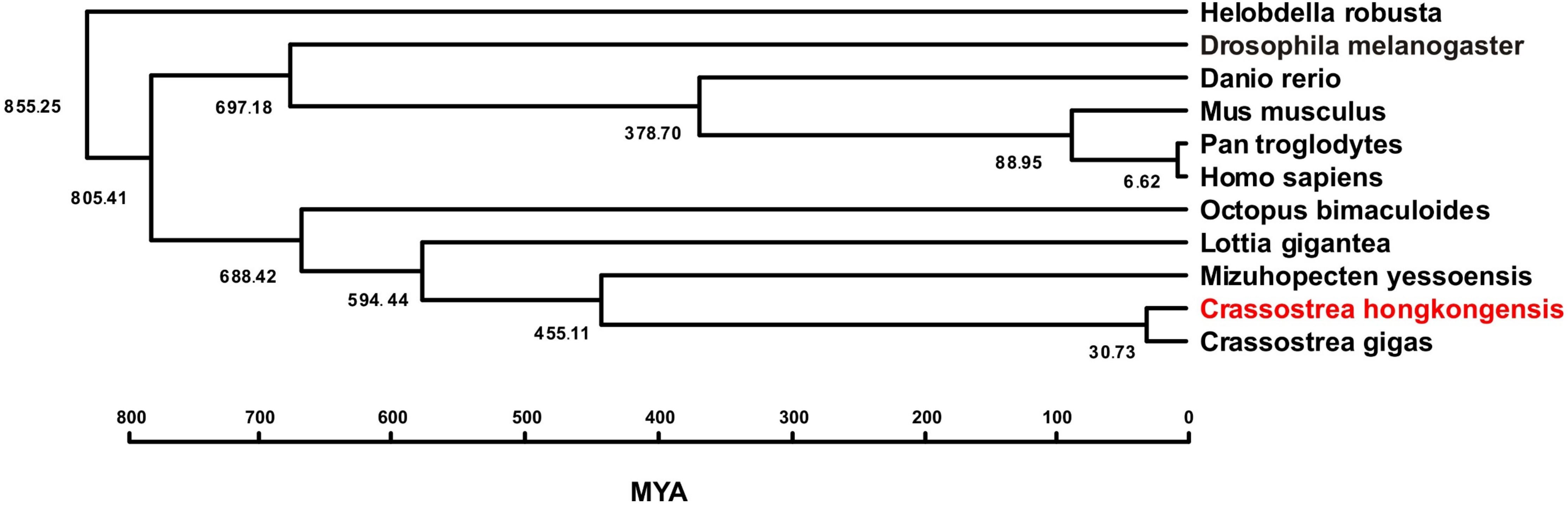
Signaling transduction:

Adenylate and Guanylate cyclase catalytic domain
HAT family dimerization domain

Growth and metabolism:

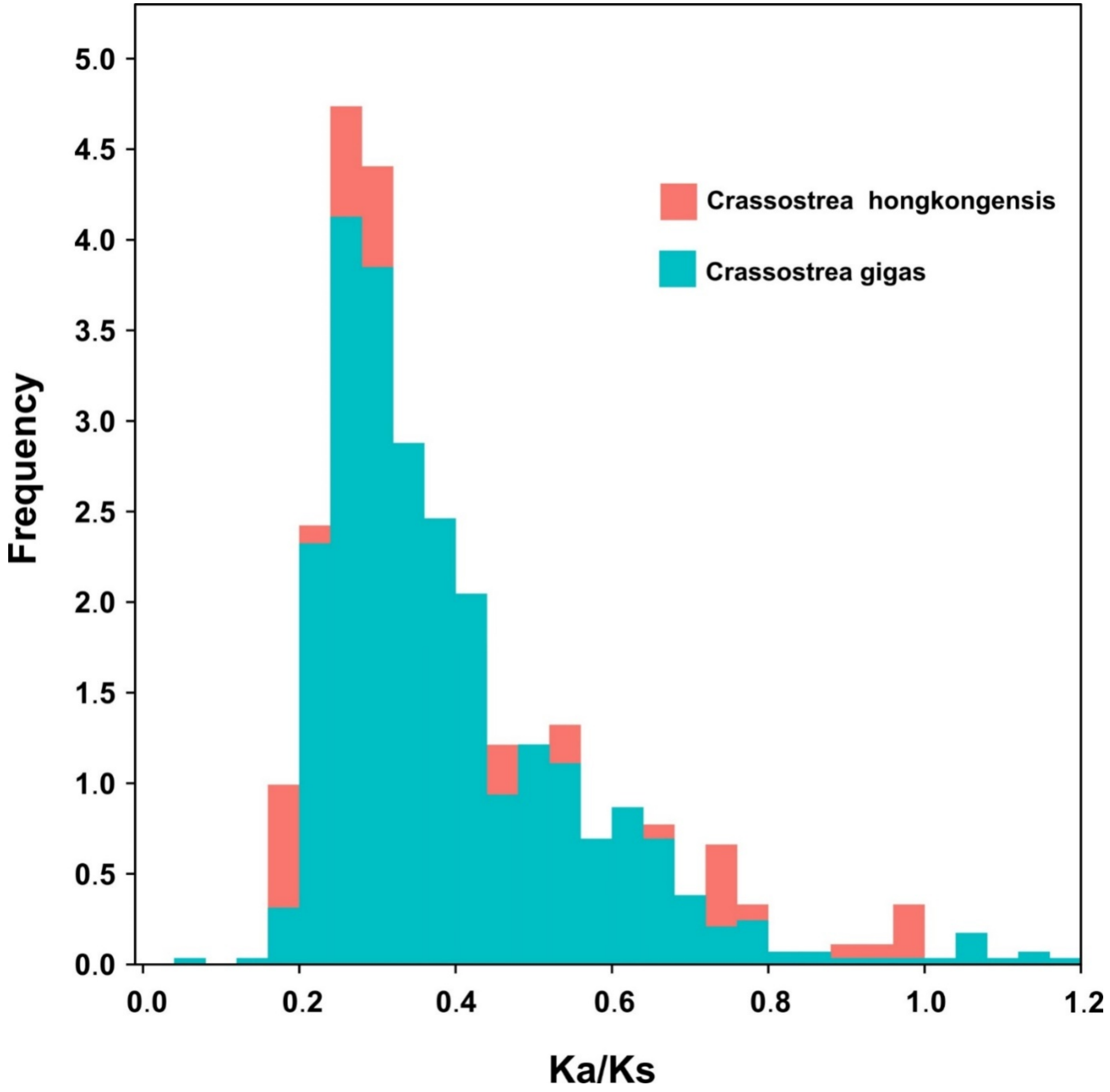
DNA polymerase B
apolipoprotein
Low temperature viability protein

Phylogenetic tree for divergence among species



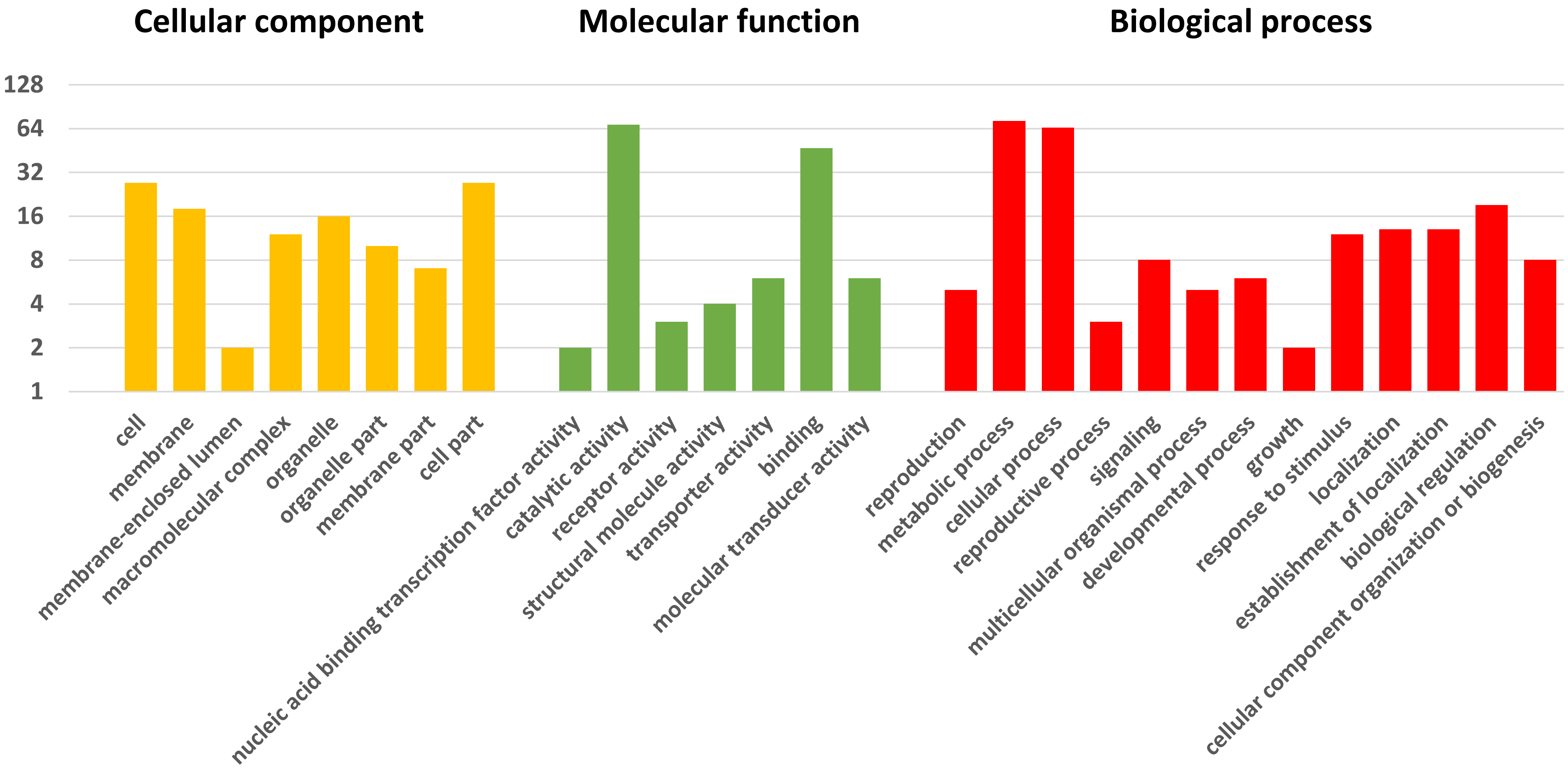
The divergence time of two oysters may occur at 30.73 MYA, which is close to inference based on mitochondrial evidence (26.12MYA).

Selective pressure of single copy gene

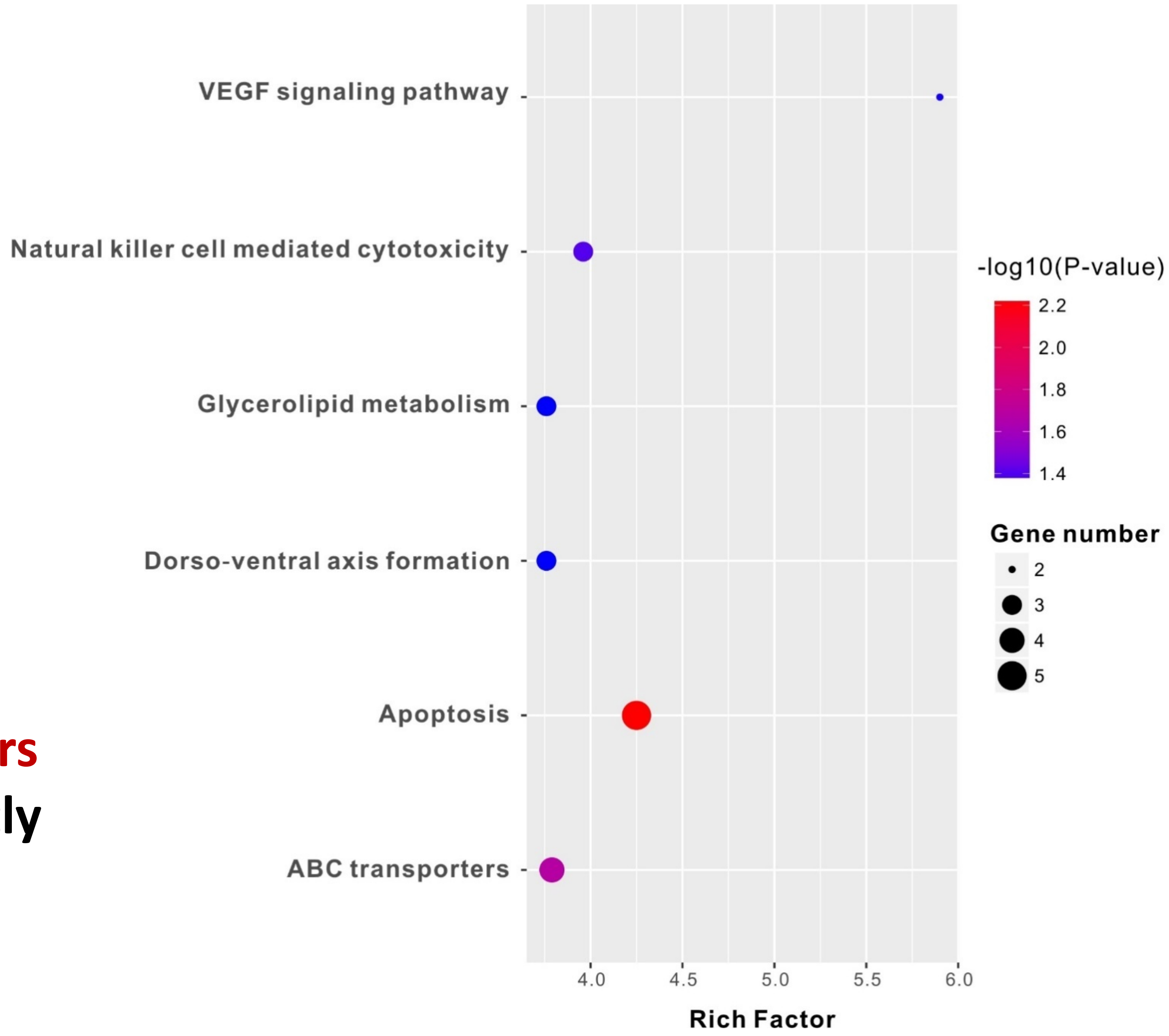


The mean of Ka/Ks of is 0.42, 583 of genes are rapidly evolving in hongkong oyster genome.

GO classification of Rapid evolving genes

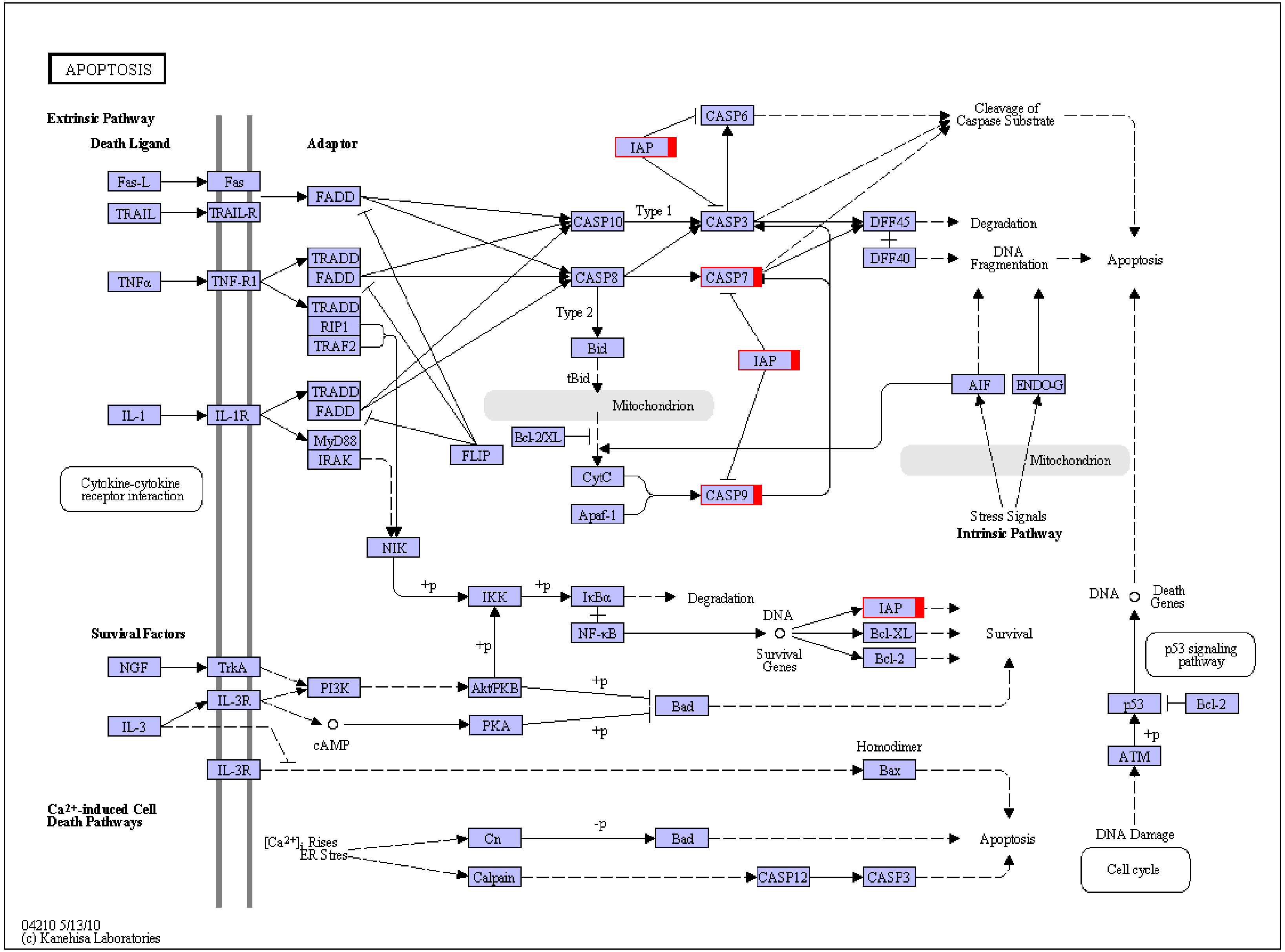


Pathway enrichment of Rapid evolving genes



The pathway ABC transporters and Apoptosis are significantly enriched

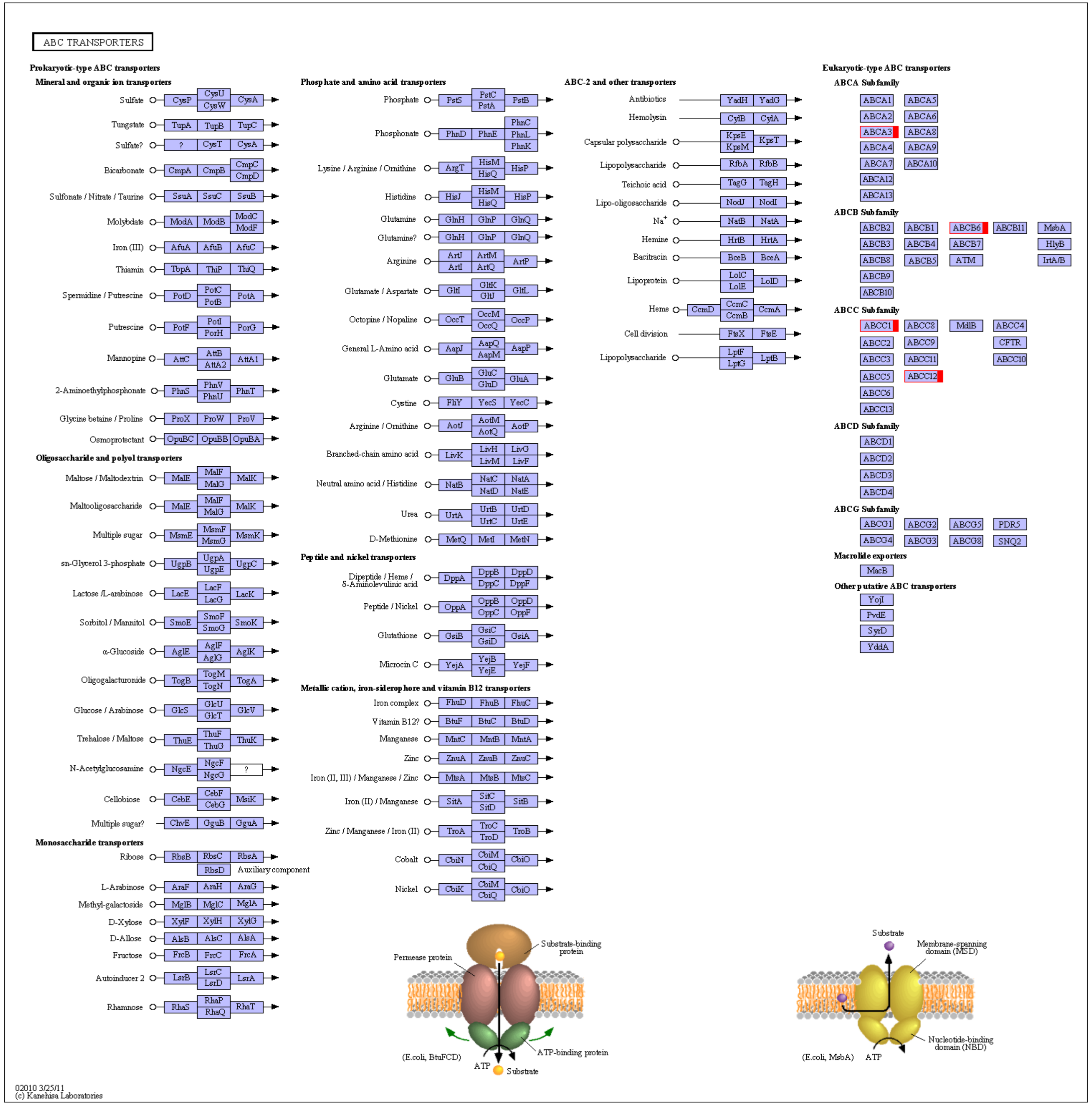
Apoptosis Pathway



Rapid evolution of apoptosis may benefit oyster to cope with more stressful environment.

ABC transporter Pathway

ATP-binding cassette transporters (ABC transporters) are largest and oldest families of transport system, Including amino acids, and other solutes, may contribute to euryhaline of hong kong oyster.

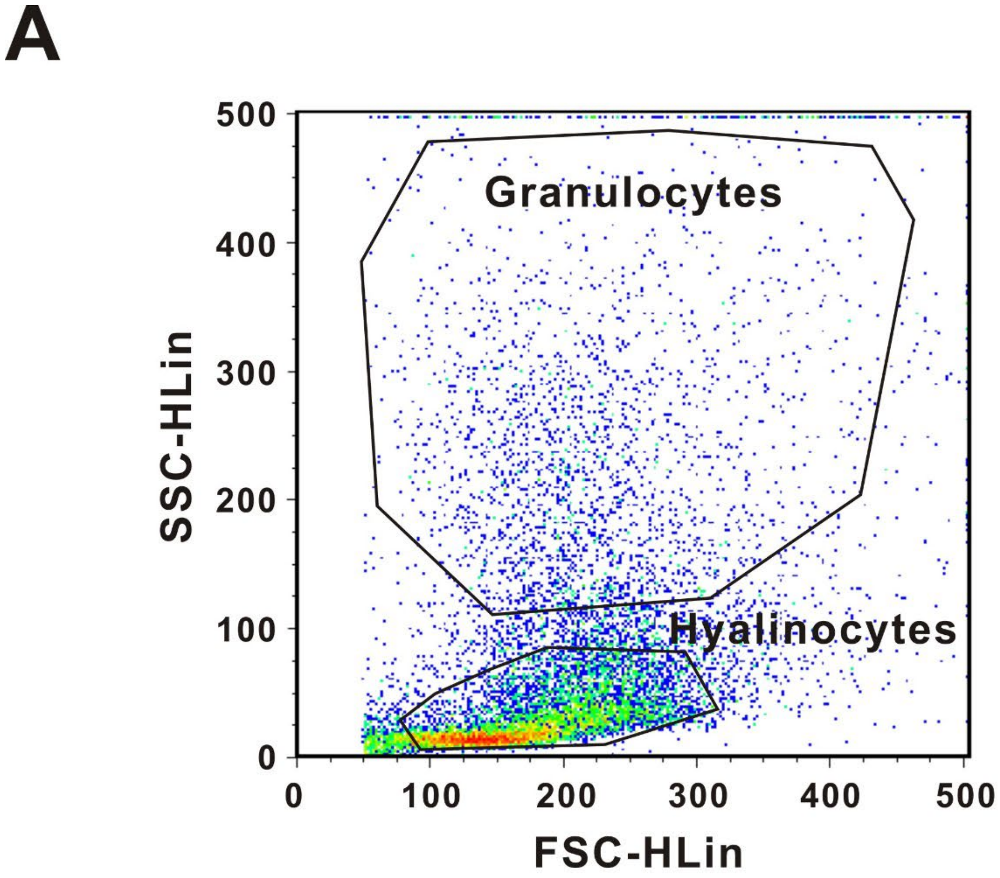


Conclusion

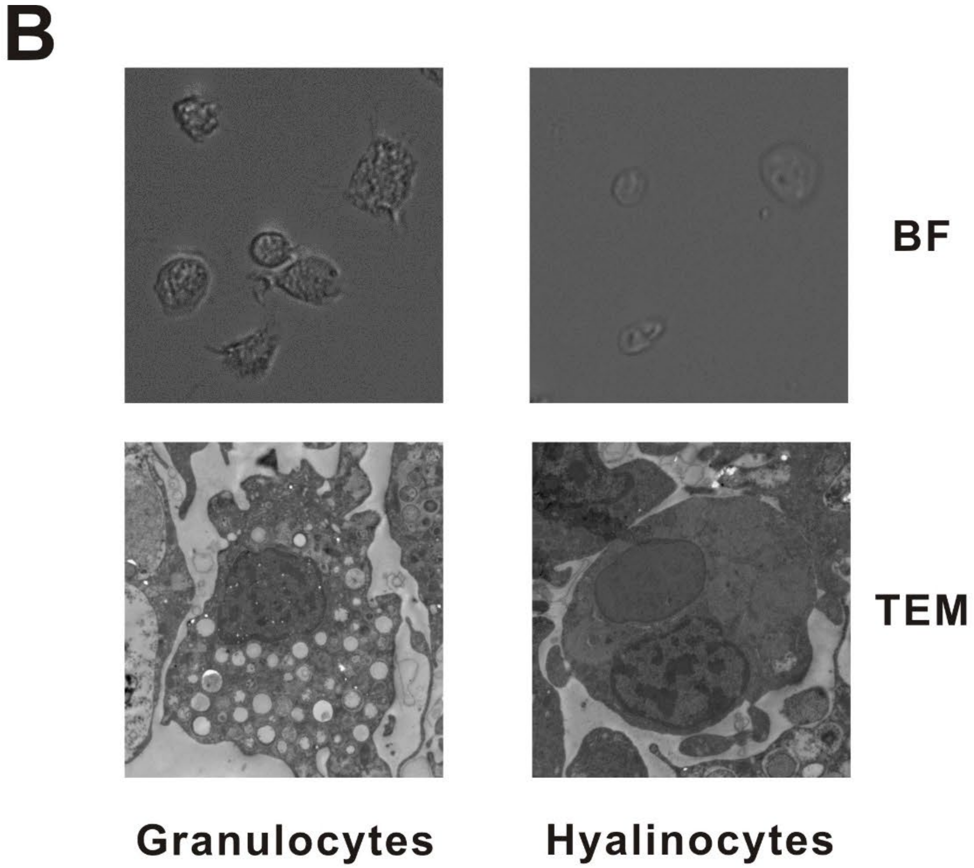
1. The estimated final genome assembly is**714.88 Mb**, covering about 98.20% of the estimated genome size;
2. A total number of **35,624 genes** were predicted; of which 90.84% were annotated on the basis of available genomic databases;
3. **1,223 gene families** were found to be specific to C. hongkongensis, including substantial genes involved in signaling transduction and immune response**;**
4. C. gigas and C. hongkongensis may diverge from **30.73 MYA**;
5. Rapid evolving genes of hong kong oyster are significantly enriched with **ABC transporters and Apoptosis**, which may benefit to shape its specific stress adaptation.

Transcriptome reveals molecular basis for oyster hemocytes differentiation

Two cell types of oyster hemocytes

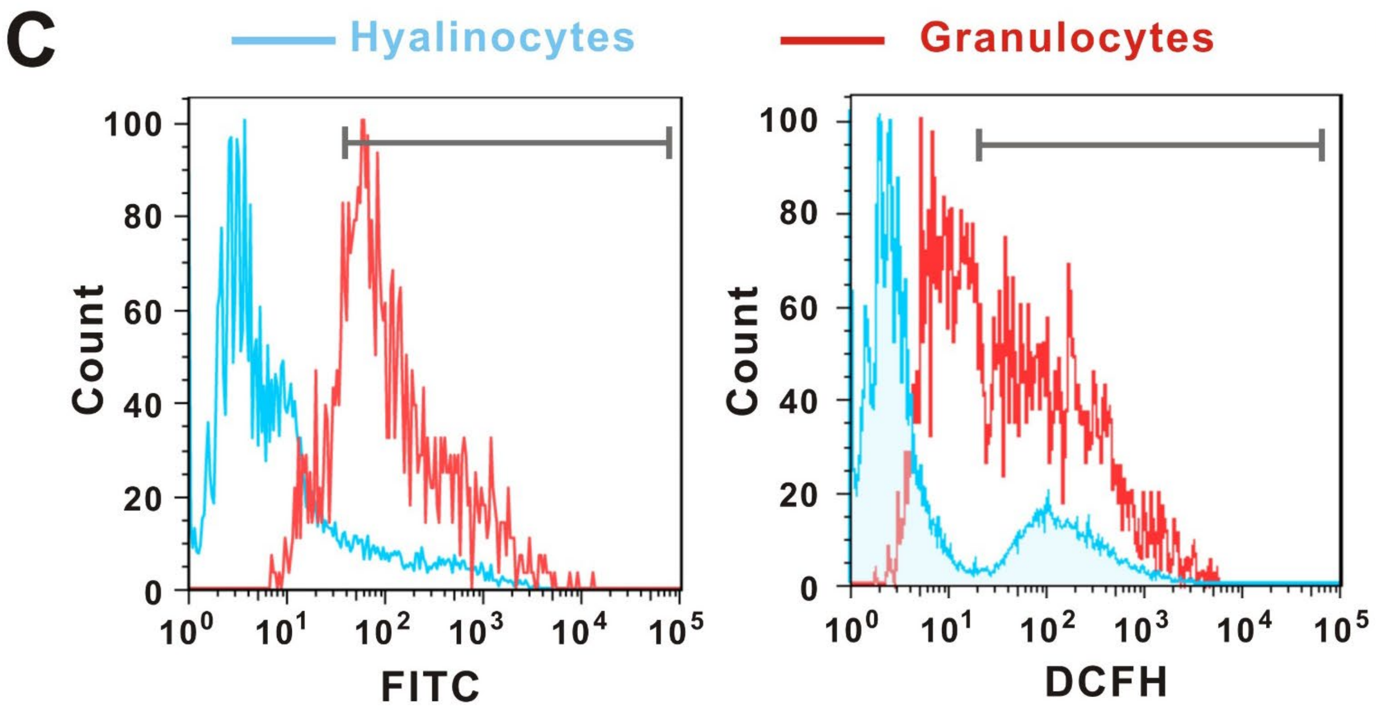


Cytotype of hemocytes by Flow cytometry



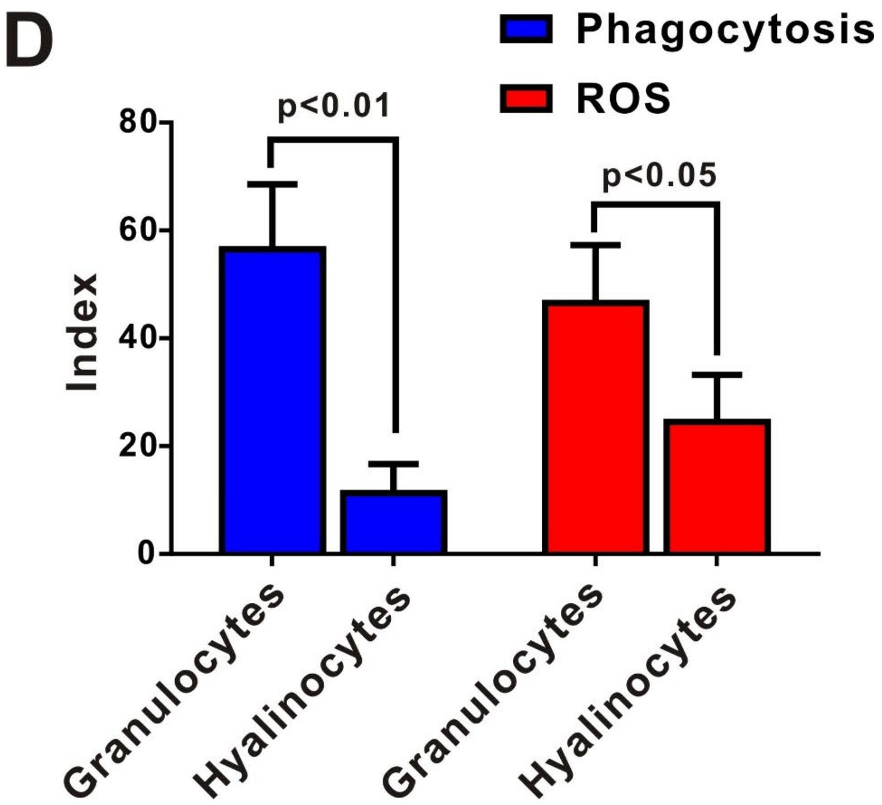
Morphological analysis

Granulocytes and hyalinocytes have differentiated in function



Phagocytosis

ROS



Different ability in phagocytosis and ROS production

Experimental design

Question:

What is the molecular basis for hemocytes functional differentiation?

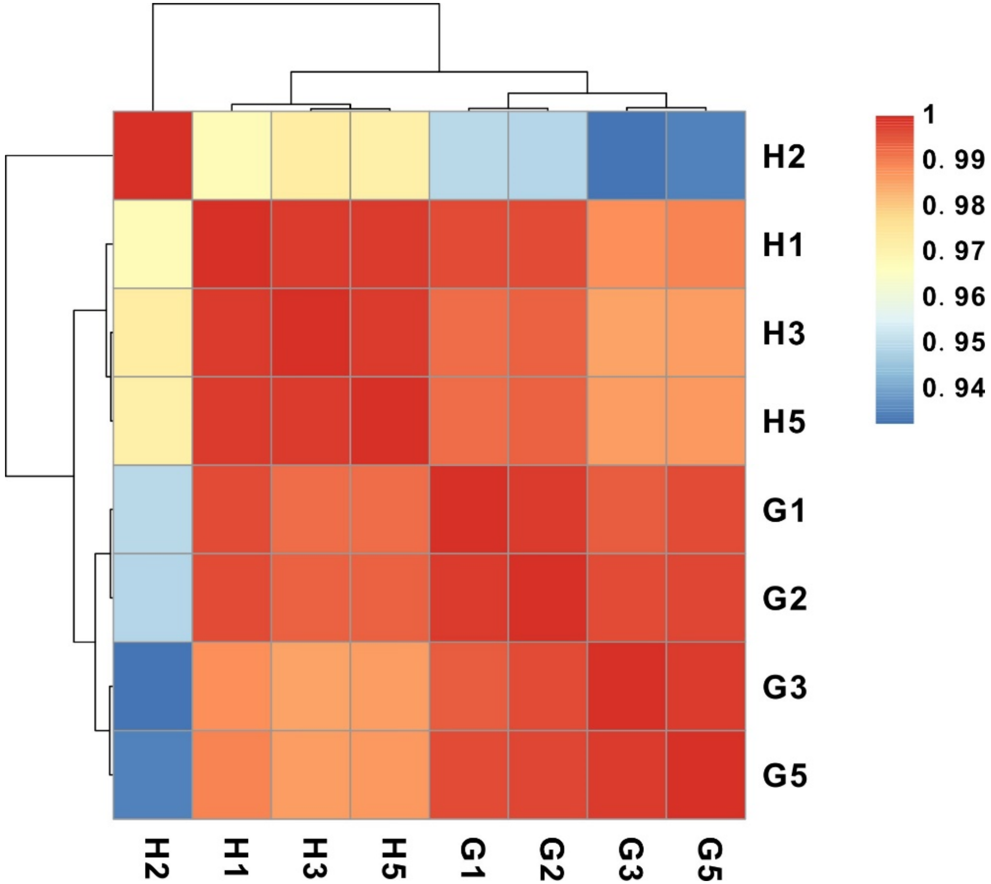
Cell sorting

RNA-seq with low input RNA

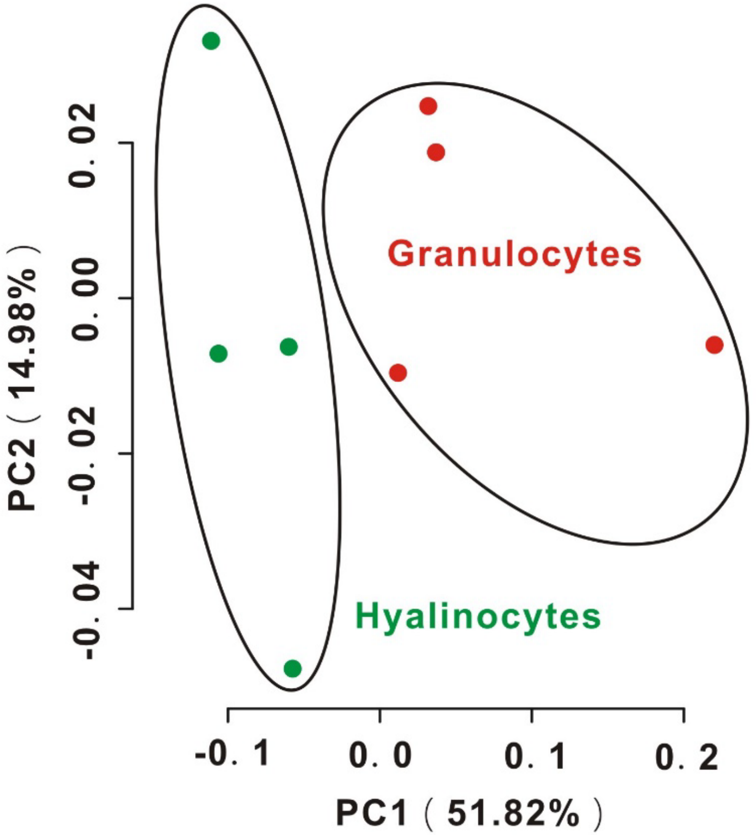
Expression profile

Functional identification

The expression profile analysis by RNA-seq



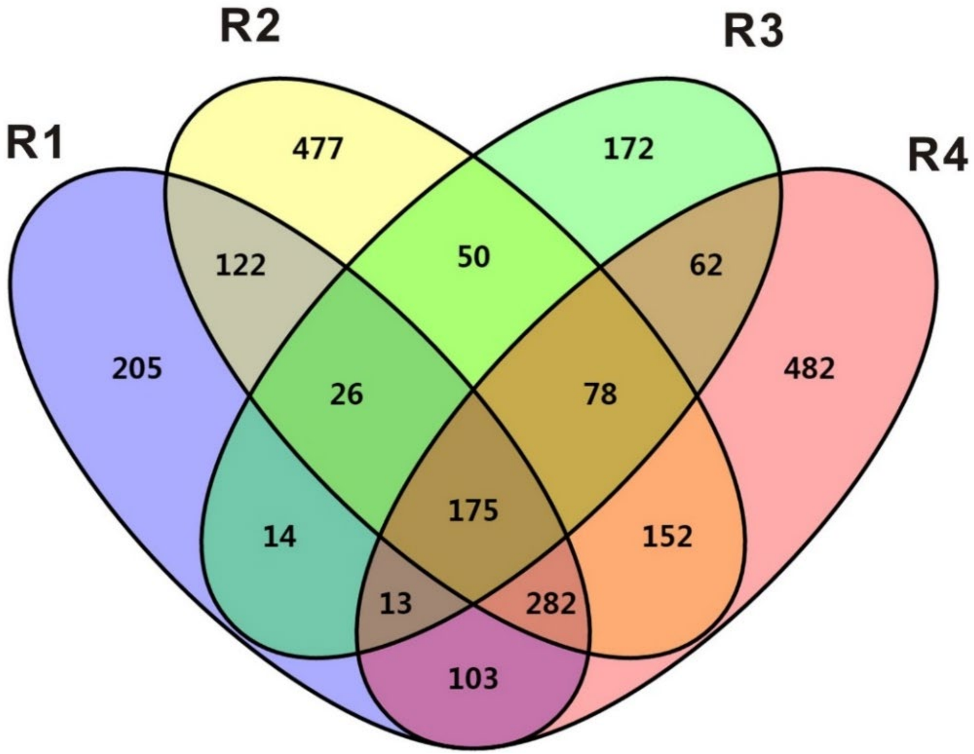
Correlation between biological replications



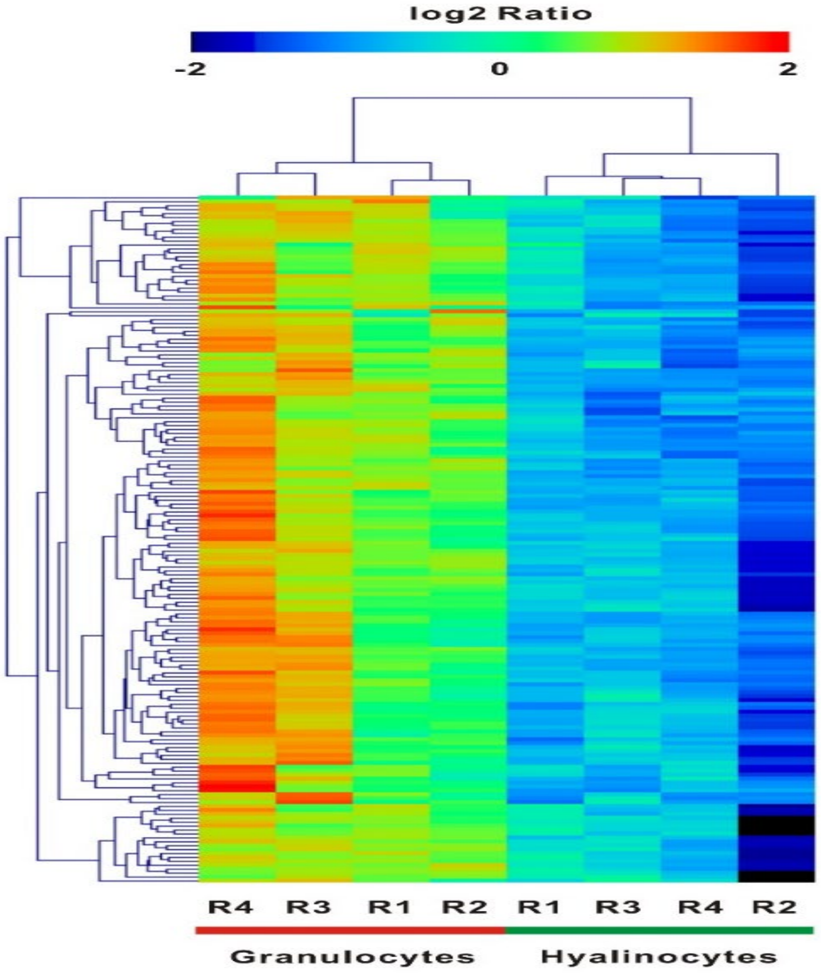
Principal component analysis

The core DEGs are dominantly expressed in the Granulocytes

DEGs =differently expressed genes

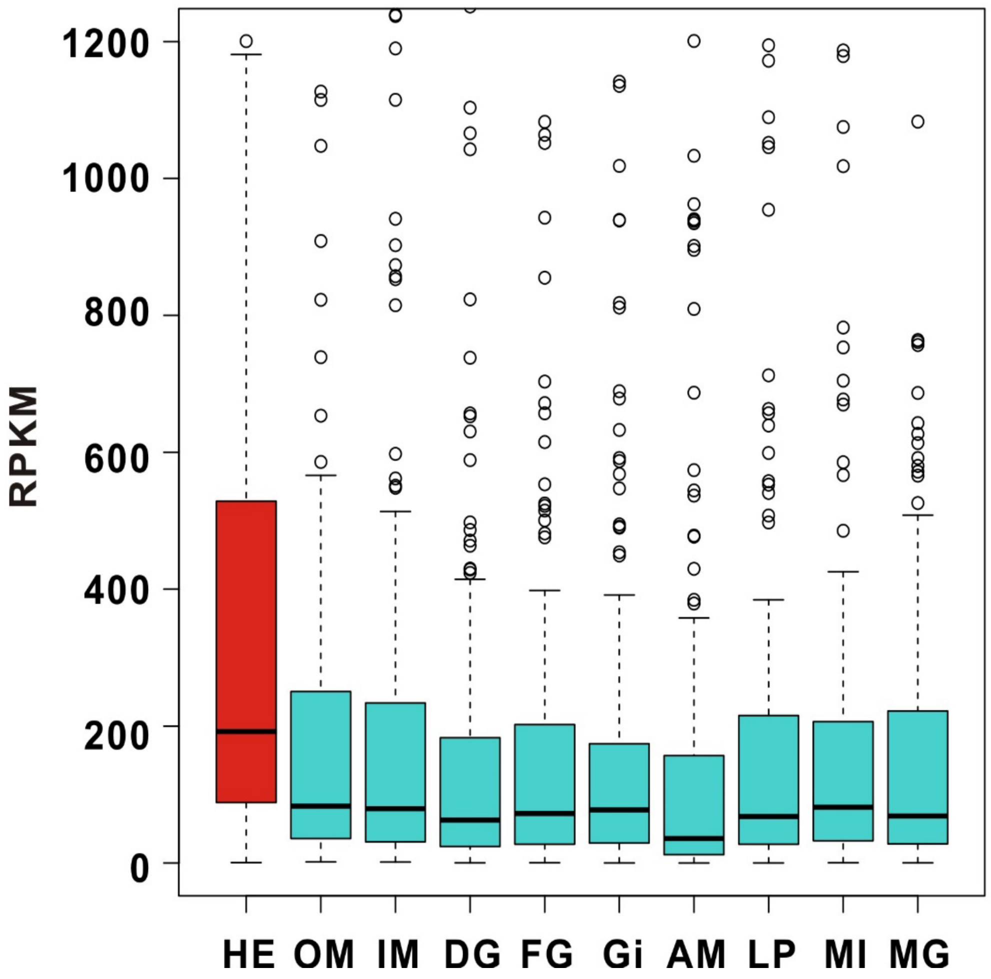


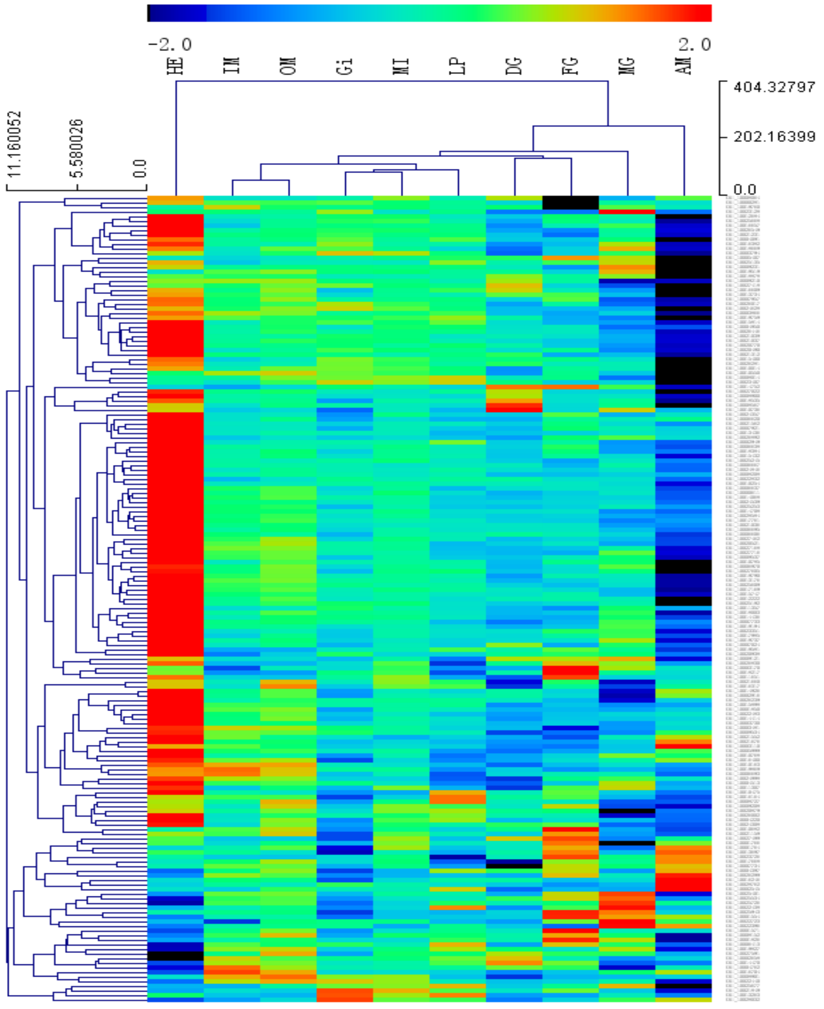
Veen Plot shows 175 of core DEGs



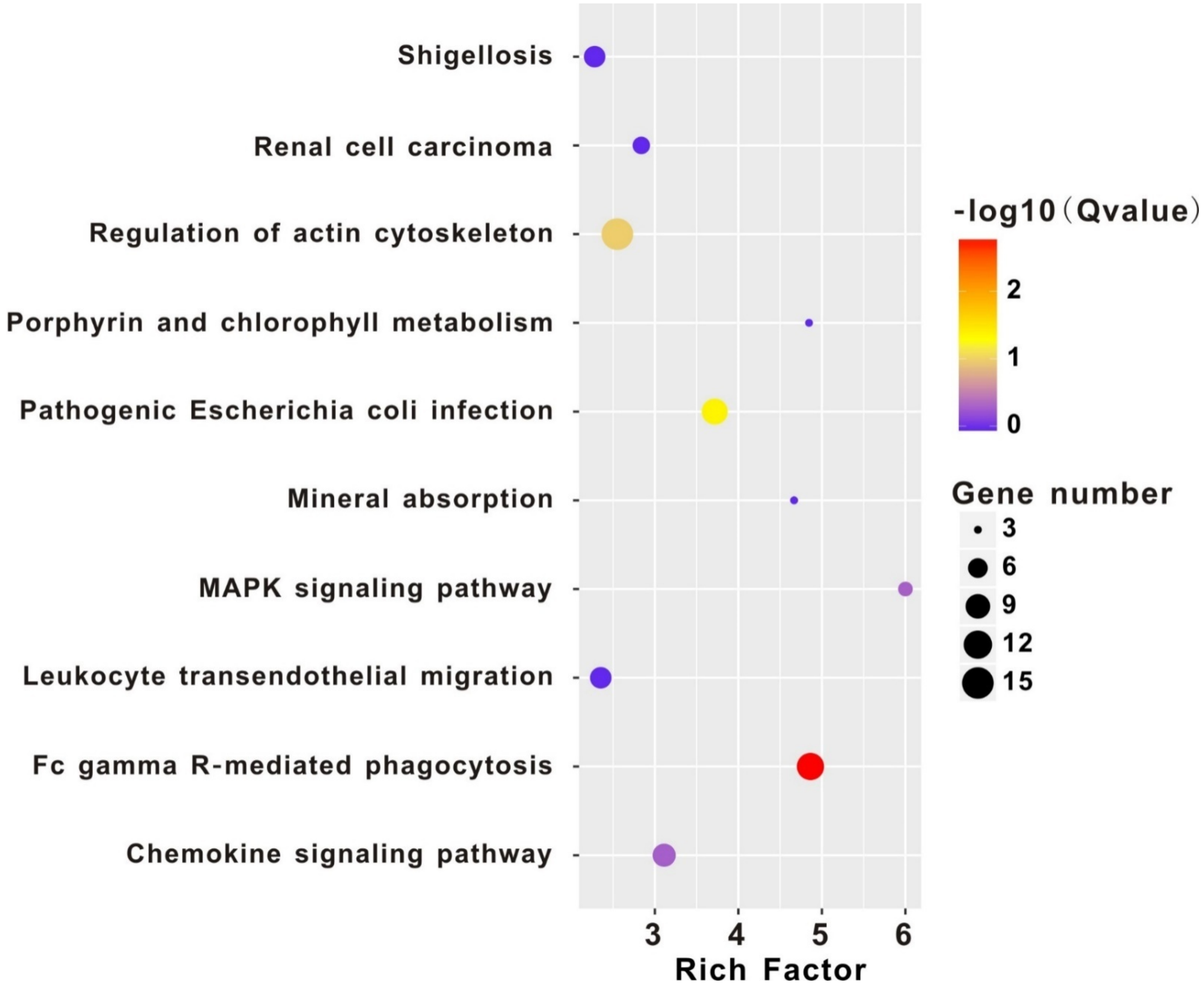
Heatmap of these core DEGs

The majority of core DEGs are specifically high-expressed in hemocytes

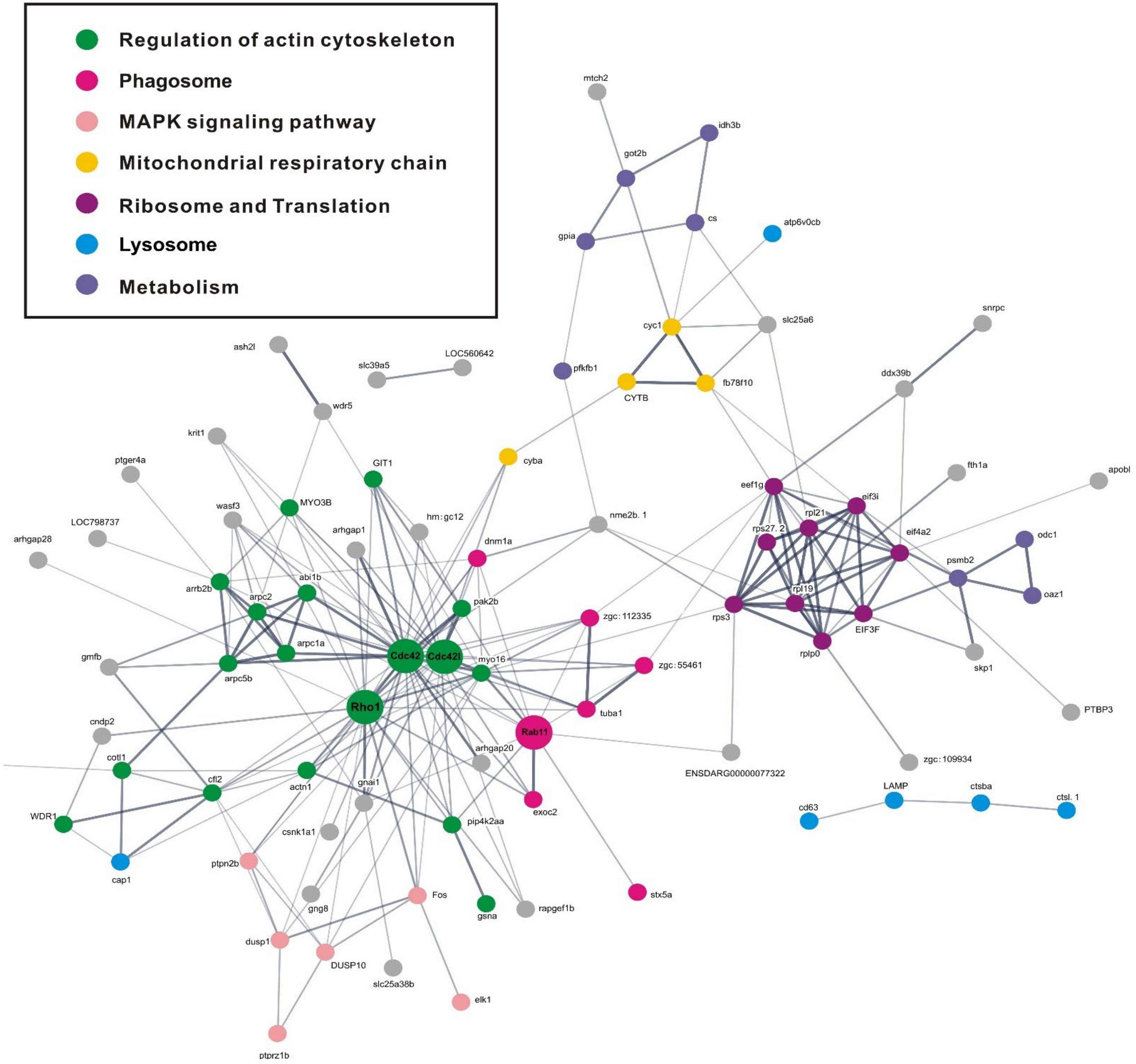




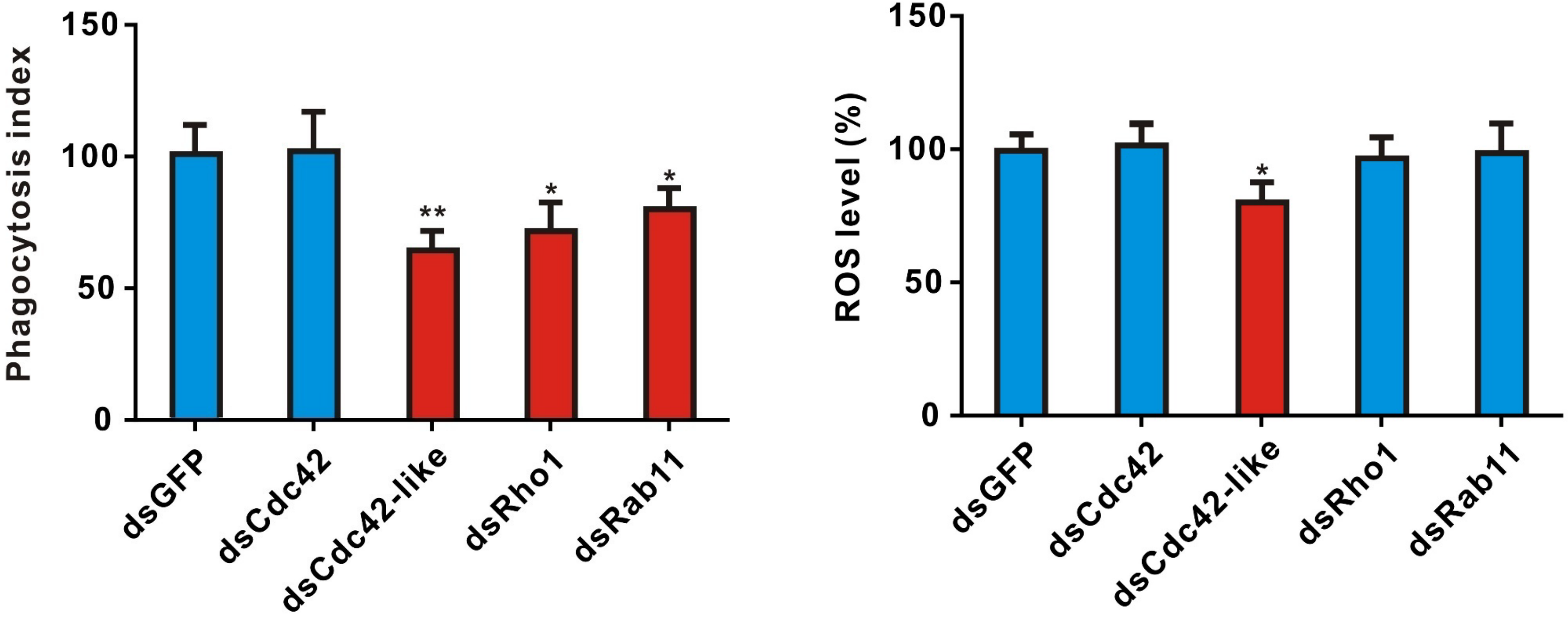
Pathway enrichment analysis with 175 DEGs



Protein-protein interaction analysis revealed the important connect between phagocytosis and actin cytoskeleton



Identification of hubgenes functions after RNAi



The effect of Knockdown of hubgenes on phagocytosis and ROS production

Conclusion

1. Granulocytes and hyalinocytes are two types of hemocytes and have differentiated in function;
1. 175 core DEGs are dominantly expressed in the Granulocytes, and also demonstrated in hemocytes-specific expression pattern;
2. Pathway enrichment analysis reveals these DEGs are mainly distributed in phagocytosis, regulation of actin cytoskeleton and MAPK signaling pathway , confirming that granulocytes are main immune cells in oysters;
1. Protein-protein interaction analysis demonstrated the several key hubgenes (cdc42,rho,rab11) regulate phagosome formation and actin cytoskeleton, as confirmed by RNAi.

Acknowledgment

Dr. Yang Zhang

Dr. Zhiming Xiang

Dr. Jun Li

Dr. Yuehuan Zhang

Dr. Fan Mao

NSFC

中国科学院

CHINESE ACADEMY OF SCIENCES

Thank you for attention!