

# Developing a platform of environmental omics for the green-lipped mussel *Perna viridis*

Jack C. H. Ip<sup>1</sup>, Priscilla T. Y. Leung<sup>1</sup>, J. W. Qiu<sup>2,3</sup>, Chris K. C. Wong<sup>2,3</sup>, Leo L. Chan<sup>3,4</sup>, Paul K. S. Lam<sup>3,4</sup>, and Kenneth M. Y. Leung<sup>1,3</sup>,

<sup>1</sup>The Swire Institute of Marine Science and School of Biological Sciences, the University of Hong Kong, Hong Kong, China.

<sup>2</sup>Department of Biology, Hong Kong Baptist University, Hong Kong, China

<sup>3</sup>State Key Laboratory in Marine Pollution (SKLMP), City University of Hong Kong, Hong Kong, China.

<sup>4</sup>Department of Biology and Chemistry, City University of Hong Kong, Hong Kong, China.

E-mail contact: [jackip@connect.hku.hk](mailto:jackip@connect.hku.hk)

---

## 1. Introduction

The green-lipped mussel *Perna viridis* is an important sentinel marine species commonly employed in pollution monitoring and ecotoxicological studies in Asia-Pacific region, and has been considered as a subtropical equivalent biomonitor of the temperate *Mytilus* species. However, the genomic information of *P. viridis* is still largely unexplored when compared with *Mytilus* species. This study aims to establish a transcriptomic profile of *P. viridis* using the next generation sequencing technology and provide a good representative set of genomic information for elucidation of toxic mechanisms upon pollutant-mediated stresses and identification of a suite of suitable biomarkers for pollution monitoring.

## 2. Materials and methods

### 2.1. Sample preparation

Adult *P. viridis* (40-50 mm shell length) were collected from different locations in the marine environment of Hong Kong and from those after 24-hour exposures to various physical challenges (e.g., different temperature, salinity, dissolved oxygen and pH levels) and chemical stressors (e.g., metals, endocrine disruptors, organic pollutant and engineered nano-particles) under laboratory conditions. Four individuals (with two males and females) were chosen for obtaining the three target tissues (i.e., hepatopancreas, gill and adductor muscle). The tissue samples were stabilized in RNA<sup>later</sup><sup>TM</sup> (QIAGEN, Hilden, Germany), and stored in -20°C before further analysis.

### 2.2. RNA preparation and sequencing

RNA of each sample was extracted with the RNeasy Mini Kit and digested with DNase I following manufacturer's instructions (QIAGEN, Hilden, Germany). For each sex and each tissue type, a total RNA sample was extracted from pooled tissues from the field and laboratory treated mussels. Thus, six cDNA libraries (i.e., 2 sexes x 3 tissue types) were constructed, and thereby having six individual lanes for Solexa Illumina sequencing. The sequencing was performed at the Centre of Genomic Sciences, the University of Hong Kong by using TruSeq Stranded mRNA Sample Prep Kit (Illumina Inc., San Diego, USA) and Illumina Genome Analyzer Ix (GAIIx).

### 2.3. *De novo* assembly and annotation

Only high quality reads were used to perform *de novo* assembly using Trinity version trinityrnaseq\_r2012-10-25 [2] with the default parameters. The assembled transcripts were further clustered using Cd-hit [3] to remove any redundancy using a similarity score of  $\geq 95\%$ . The annotations of the transcripts were performed based on sequence similarity search against the NCBI nr database and the three molluscan EST databases (*Aplysia californica*, GenBank Accession: AASC000000000; 255,605 EST entries; *Crassostrea gigas*, GenBank Accession: AFTI000000000; with 206,647 EST entries and *Lottia gigantea*, GenBank Accession: AMQO000000000; with 252,091 EST entries) constructed from NCBI's non-redundant database using BLASTx with an e-value threshold of  $1e^{-5}$  and HSP cut-off length of 33. The presence of conserved domain of the transcript was analyzed by InterProScan [4]. Functional classification was carried out based on Gene Ontology (GO) using BLAST2GO [1]. KEGG (Kyoto encyclopedia of genes and genomes) pathways were assigned to the transcripts using the online KEGG Automatic Annotation Server (KAAS, Kyoto encyclopedia of genes and genomes, <http://www.genome.jp/kegg/kaas/>).

## 2.4. Tissue- and gender-expression patterns

The filtered reads of the transcripts from individual tissue and sex were mapped to the assembled *de novo* transcriptome, and their expressed values of the counts were calculated as FPKM (Fragments Per Kilo bases per Million reads) with a cut-off value  $\geq 0.05$ . The abundance estimation of each gene between pair of samples was performed using RSEM analysis.

## 3. Results and discussion

### 3.1. Sequencing and *de novo* assembly

A global *de novo* transcriptome of *P. viridis* was assembled from a pool of the reads of the six cDNA libraries using Trinity software. A total of 544,272,542 high quality reads were obtained after filtering, and subjected to *de novo* assembly to yield 233,257 contigs. After clustering with Cd-hit, the resulting assembled transcriptome consisted a condense number of 150,111 unigenes and 192,879 transcripts. The lengths of transcript ranged from 200 to 43,014 bp, with an average length of 1,264 bp and a N50 length of 2,868 bp.

### 3.2. Functional annotation

The 192,879 non-redundant assembled transcripts were blasted against the NCBI nr database and the three molluscan EST databases, and resulted in 44,713 transcripts with at least a blast hit, and having a top match with the sequences from the Pacific oyster, *Crassostrea gigas* (27,651 transcripts). A total 93,668 transcripts were resulted from putative function for protein domain based on the analysis using InterProScan. There were 21,262 transcripts associated with at least one well-defined Gene Ontology (GO) term which was further categorized into 31 functional groups (GO term of level 2) based on three main categories of functions, i.e., biological process, cellular component and molecular function.

### 3.3. Tissue-specific pattern in associated with stress responses

Based on GO term assignment and KEGG annotation, we found that 214 genes were involved in “response to stress” under the level 2 of classification for “biological process”, and 244 genes were associated with “xenobiotics biodegradation and metabolism” category under the metabolism pathway. It is believed that genes from these two groups will be highly relevant to the stress-associated responses of *Perna viridis* upon the challenges of chemical pollutants or other environmental stresses. As revealed by multivariate analysis, expression patterns of the identified genes from these two groups were strongly tissue-specific but the differences between genders were little.

## 4. Conclusion

A *de novo* transcriptome of *Perna viridis* has been successfully constructed based on the mRNA of the three target tissues, i.e., adductor muscle, gill and hepatopancreas. A strong tissue-specific expression patterns were revealed in particular for genes associated with stresses and detoxification processes, while the differences of the gene expression between genders were not obvious. This study not only provides an informative transcriptomic platform for studying toxic mechanisms of various chemical contaminants towards *P. viridis*, but also highlights the importance of applying a tissue-specific approach in future investigations.

## 5. References

- [1] Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674-3676.
- [2] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis, X, Fan L, Raychowdhury R, Zeng QD et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7): 644-652
- [3] Li WZ, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658-1659.
- [4] Zdobnov EM, Apweiler R. 2001. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17(9):847-848.

*Acknowledgements* - This work is supported by the State Key Laboratory in Marine Pollution through a Seed Collaborative Research Fund (2011) to KMY Leung, and partially funded by the University of Hong Kong via a postgraduate studentship (to JCH Ip). The authors thank the helpers who assisted in field sampling, and Helen Leung and Cecily Law for technical support.