

RESEARCH ARTICLE

Open Access

# *De novo* transcriptome analysis of *Perna viridis* highlights tissue-specific patterns for environmental studies

Priscilla TY Leung<sup>1</sup>, Jack CH Ip<sup>1</sup>, Sarah ST Mak<sup>1</sup>, Jian Wen Qiu<sup>2,3</sup>, Paul KS Lam<sup>2,4</sup>, Chris KC Wong<sup>2,3</sup>, Leo L Chan<sup>2</sup> and Kenneth MY Leung<sup>1,2\*</sup>

## Abstract

**Background:** The tropical green-lipped mussel *Perna viridis* is a common biomonitor throughout the Indo-Pacific region that is used for environmental monitoring and ecotoxicological investigations. However, there is limited molecular data available regarding this species. We sought to establish a global transcriptome database from the tissues of adductor muscle, gills and the hepatopancreas of *P. viridis* in an effort to advance our understanding of the molecular aspects involved during specific toxicity responses in this sentinel species.

**Results:** Illumina sequencing results yielded 544,272,542 high-quality filtered reads. After *de novo* assembly using Trinity, 233,257 contigs were generated with an average length of 1,264 bp and an N50 length of 2,868 bp; 192,879 assembled transcripts and 150,111 assembled unigenes were obtained after clustering. A total of 93,668 assembled transcripts (66,692 assembled genes) with putative functions for protein domains were predicted based on InterProScan analysis. Based on similarity searches, 44,713 assembled transcripts and 25,319 assembled unigenes were annotated with at least one BLAST hit. A total of 21,262 assembled transcripts (11,947 assembled genes) were annotated with at least one well-defined Gene Ontology (GO) and 5,131 assembled transcripts (3,181 assembled unigenes) were assigned to 329 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The quantity of assembled unigenes and transcripts obtained from male and female mussels were similar but varied among the three studied tissues, with the highest numbers recorded in the gills, followed by the hepatopancreas, and then the adductor muscle. Multivariate analyses revealed strong tissue-specific patterns among the three different tissues, but not between sexes in terms of expression profiles for annotated genes in various GO terms, and genes associated with stress responses and degradation of xenobiotics. The expression profiles of certain selected genes in each tissue type were further validated using real-time quantitative polymerase chain reaction assays and a similar tissue-specific trend was seen.

**Conclusions:** The extensive sequence data generated from this study will provide a valuable molecular resource for facilitating environmental studies with *P. viridis*, and highlight the importance of tissue-specific approaches in the future.

**Keywords:** *De novo* transcriptome, *Perna viridis*, Mussel, RNA-seq, Tissue specificity, Biomonitor

\* Correspondence: kmyleung@hku.hk

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

<sup>2</sup>State Key Laboratory in Marine Pollution, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China

Full list of author information is available at the end of the article

## Background

Over the past 30 years, mussels have served as useful bio-monitors in assessing the quality of our marine ecosystems with respect to the impact of chemicals [1]. Mytilid mussels, in particular *Mytilus* species, have been extensively applied to environmental monitoring programs and toxicological studies in temperate regions [2]. More recently, a tropical/subtropical member of mytilids, the green-lipped mussel *Perna viridis*, has been adopted as a biomonitor throughout the Indo-Pacific region [3] for assessing a wide range of chemical pollutants [4-8]. Because of its wide distribution, and high tolerance to environmental stresses, this mussel is also regarded as highly invasive and a biofouling species [9].

In conventional biomonitoring, the body burden of common pollutants and a limited number of physiological and molecular biomarkers are used as the end-points for assessing the effects of chemical pollution on mussels [3]. To understand the toxic effects of chemical pollutants, and to develop an early warning signal of pollutant stress on a biomonitor, it is essential to have an in-depth knowledge of their toxic mechanisms at a molecular level. However, like many other non-model marine organisms, the limited genomic data available for *P. viridis* has hindered studies attempting to elucidate molecular mechanisms related to specific toxic stress responses.

A transcriptome of a given organism refers to a complete set of transcripts present in a cell type at a specific developmental stage or physiological condition [10]. The underlying molecular mechanisms and phenotypic plasticity of an organism against environmental influences can be interpreted from the dynamics of its transcriptome. Given the advent of next-generation sequencing (NGS) technology, high throughput RNA-sequencing analysis of transcripts with greater gene coverage, higher sensitivity, and better reproducibility of transcriptomes that are highly complex, can now be applied to any genome [10]. RNA-sequencing analysis has been applied to several marine bivalve species, including the Mediterranean mussel *Mytilus galloprovincialis* [11], the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus* [12], the Manila clam *Ruditapes philippinarum* [13], the Yesso scallop *Mizuhopecten yessoensis* (formerly known as *Patinopecten yessoensis*) [14], the pearl oyster *Pinctada fucata* [15] and the Pacific oyster

*Crassostrea gigas* [16]. Some of these studies have highlighted the species specificity of the transcriptome, and also demonstrated specific transcript profiles among different tissues. The extensive and novel genomic data generated from these non-model animals have revealed much about their underlying molecular mechanisms in response to different environmental stressors [13,17].

Compared with other mytilids, genomic and protein databases are largely unavailable for *P. viridis*. To date, the number of sequence entries available for this species in the National Center for Biotechnology Information (NCBI) is limited, with 451 nucleotide sequences, 4 expressed sequence tags (ESTs), 252 protein sequences, and 1 high-throughput DNA and RNA sequence read archive (SRA) (Table 1, Data search performed on 15 May 2014). Molecular characterization of stress-associated genes, such as metallothioneins [18] and glutathione S-transferases [19], and the adhesion associated proteins such as the 3, 4-dihydroxyphenyl-L-alanine (DOPA)-containing proteins from the byssus threads [20] have been conducted for *P. viridis*. A recent study revealed the adhesive byssal complex and various adhesive proteins in the foot tissue of *P. viridis*, using RNA-sequencing and proteomics [21]. A number of microsatellite markers were also characterized for genetic investigations [22,23]. *P. viridis* has the shortest reported mitochondrial genome (16,014 bp) among marine mussels, and differs from that of other *Mytilus* species in terms of genome organization [24]. A cytogenetic study suggested that *P. viridis* could be a more primitive species of the Mytilidae, that diverged from other members of the *Perna* and *Mytilus* genera [25]. Such distinctions indicate that the existing genomic data from other Mytilidae member may not be suitable to serve as a basis for homologous references in *P. viridis*.

Given the importance of *P. viridis* to ecotoxicology and environmental applications, a comprehensive genomic database is urgently needed for this species to advance future marine environmental studies in the Indo-Pacific region. In the current study, we applied Illumina sequencing to characterize a *de novo* transcriptome assembled from a mixed sample of three target tissues (adductor muscle, gills and hepatopancreas) of male and female *P. viridis* collected from the field and from various designated exposure scenarios. Comparative

**Table 1 Sequence entries for Mytilidae including *Perna viridis* and three *Mytilus* species in the NCBI database**

Mussel	EST entries	Nucleotide entries	Protein entries	SRA
Mytilidae	71,328	54,548	12,803	65
<i>Perna viridis</i>	4	451	252	1
<i>Mytilus californianus</i>	42,365	945	969	0
<i>Mytilus galloprovincialis</i>	19,756	41,167	2,417	16
<i>Mytilus edulis</i>	5,352	5,301	3,496	45

Note: Search was performed on 15 May 2014.

analyses were also performed to determine any tissue- and sex-specific transcriptome profiles.

Results and discussion

Sequencing and assembly

An overall profile of the transcriptome of *P. viridis* was obtained through Illumina sequencing of six cDNA libraries (2 sexes × 3 tissues) and pooled mRNA samples from the adductor muscle, gills and hepatopancreas of male and female. The number of reads generated for each target tissue ranged from 84,607,612–102,182,018 with a Q30 values (denoting the accuracy of a base call to be 99.9%) ranging from 83.78–89.19%. The highest GC content was seen in the adductor muscle (42%), followed by the hepatopancreas (40%), and the gills (39%) (Additional file 1).

A global *de novo* transcriptome of *P. viridis* was assembled from the pool of the reads for the six cDNA libraries using Trinity software. We obtained 544,272,542 high-quality reads after filtering and subjecting to *de novo* assembly, generating 233,257 contigs. After clustering with CD-Hit, the resulting assembled transcriptome comprised 192,879 assembled transcripts and 150,111 assembled unigenes (Table 2). The lengths of the assembled transcripts ranged from 201–43,014 bp, with an average length of 1,264 bp and an N50 length of 2,868 bp (Table 2). The read data for the six samples was deposited in the NCBI SRA database (BioProject Accession Number SRP043984).

Sequence abundance

The expression levels of assembled unigenes/transcripts were measured using fragments per kilobases per million reads (FPKM). The numbers of assembled unigenes/transcripts that were expressed (cut-off value at FPKM ≥ 0.05) in male and female *P. viridis* were similar, however strong variations were observed for unigene/transcript abundance among tissues (Figure 1). The largest transcriptome size was obtained from the gills, followed by the hepatopancreas, and then the adductor muscle. The numbers of assembled unigenes/transcripts in the gills and

hepatopancreas were very similar, while those detected in the adductor muscle were comparatively low. A recent study outlined the potential errors in using whole organism or composite structures (organs) in RNA-sequencing analysis as a common practice for small invertebrates like insects or larvae. The same researchers demonstrated the significance of using a structure/organ of interest to correctly interpret specific expression of genes [26]. We have demonstrated the possibility of high variability in transcriptome sizes occurring in different tissues of *P. viridis*.

Functional annotation

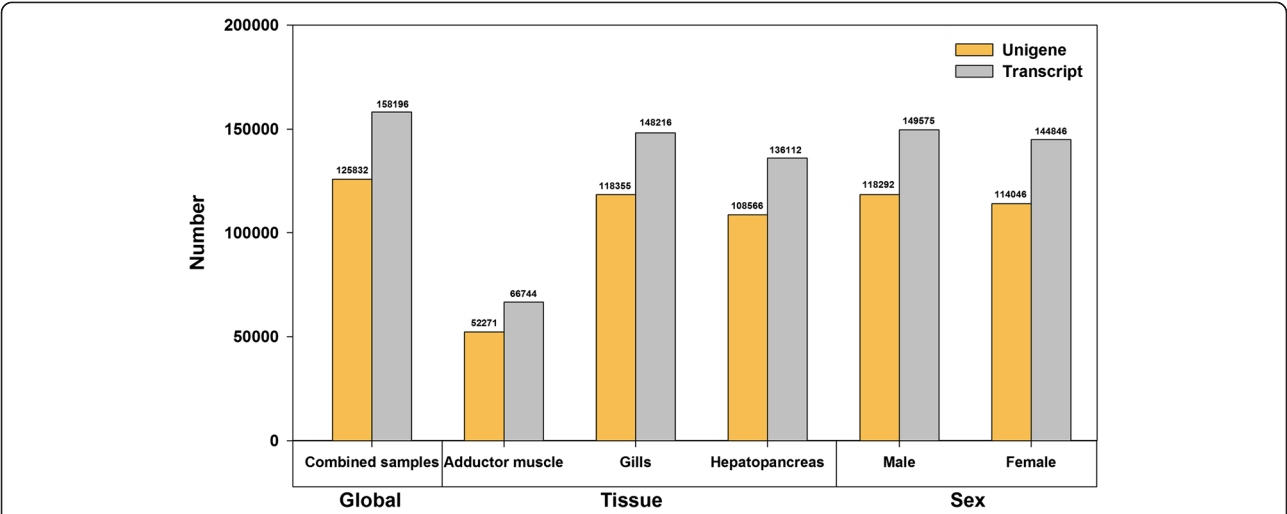
Annotation results for the non-redundant assembled transcripts/unigenes are summarized in Table 3. We performed a BLAST search against the NCBI nr database and the three molluscan EST databases; 44,713 assembled transcripts and 25,319 assembled genes with at least one BLAST hit were returned, with 41,686 assembled transcripts and 23,392 assembled unigenes that had been annotated by the NCBI nr database. The BLAST search of the 27,651 assembled transcripts had their best match with sequences from the Pacific oyster, *Crassostrea gigas* (Additional file 2).

Annotation results from the BLAST search for *de novo* transcriptomes of the non-model species were shown to be highly dependent on the availability of annotated sequence information for a reference genome, and the sizes of their contig sequences [27,28]. The number of transcripts obtained for *P. viridis* from this study was comparatively higher than for other non-model bivalve species subjected to NGS technology. There were 38,942 reported sequences (14,638 contigs plus 24,304 singletons) [14] and 14,240 transcripts in the Yesso scallop *Mizuhopecten yessoensis* [29], 22,023 transcripts in the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus* [12], 18,196 contigs in the striped venus *Chamelea gallina* [27], 9,747 transcripts in the Manila clam *Ruditapes philippinarum* [13] and 1,023 contigs in the Mediterranean mussel *Mytilus galloprovincialis* [11]. At the gene and transcript levels, the annotation results for *P. viridis* that we obtained were comparable to transcriptomes obtained from multiple organs and at various developmental stages of the Pacific oyster *C. gigas* (45,771 transcripts [17] and 28,027 protein-coding predicted genes [16]). The improvement in the annotation of assembled transcripts of *P. viridis* could be explained by the higher sequence coverage obtained because of better sequence length (average contig length: 1,264 bp; N50: 2,868 bp), and the substantial annotated genomic data available from the recently sequenced genome of *C. gigas* [16].

A total of 93,668 assembled transcripts and 66,692 assembled unigenes were predicted to have putative functions in protein domains based on InterProScan analysis. Gene Ontology (GO) functional terms were assigned

Table 2 Summary of the assembled transcriptome of *Perna viridis*

Item	Number
Total number of filtered reads	544,272,542
Total number of assembled contigs	233,257
Number of assembled transcripts after clustering	192,879
Number of assembled unigenes after clustering	150,011
N50 (bp)	2,868
Minimum contig length (bp)	201
Maximum contig length (bp)	43,014
Average contig length (bp)	1,264
Total length in contigs (bp)	295,064,579



**Figure 1 Assembled unigenes and transcripts for combined samples, individual tissue types and by sex.** The expression level of assembled transcripts/unigenes was measured using fragments per kilobases per million reads (FPKM). An assembled transcript/unigene was considered expressed using an FPKM cut-off value  $\geq 0.05$ .

according to NCBI nr annotations and the InterProScan results. From the global transcriptome, 21,262 assembled transcripts and 11,947 assembled unigenes were associated with at least one well-defined GO term. These were further classified into 31 functional groups (GO term of level 2) according to three main functional categories: biological process; cellular component; and molecular function (Additional file 3). The number of genes assigned to “metabolic process” (24.4%) was the highest under the biological process classification, while genes annotated with functions related to “cell” (64.8%) and “binding” (57.7%) were dominant for the cellular component and molecular function categories, respectively.

The composition of various GO terms for genes from individual tissue types and sexes can be seen in Additional file 4. Variations in the GO term compositions across the three tissue types and between the sexes were not apparent. Kinoshita et al. [15] also reported that there was no difference in GO composition of the annotated genes from the pallium, mental edge and pearl sac of the pearl oyster *Pinctada fucata*. However, a strong tissue-specific expression pattern was found among the three different tissues of *P. viridis* that we examined, without any apparent sex-specific patterns when multivariate analyses

were applied (Figure 2A). A recent transcriptomic study on the Mediterranean mussel *M. galloprovincialis* also showed significant tissue-specific differences among four tissues with respect to transcript expression profiles in metabolic pathways [11]. We have successfully demonstrated that GO compositions did not vary among different tissues, but expression profiles exhibited distinctive tissue-specific patterns that were probably associated with the unique functions of the adductor muscle, gills, and the hepatopancreas.

For KEGG annotation, 5,131 assembled transcripts and 3,181 assembled unigenes were assigned to 329 KEGG pathways under the main groups: metabolism (1,126 genes); genetic information processing (928 genes); environmental information processing (597 genes); cellular processes (618 genes); organismal systems (746 genes); and human diseases pathway (874 genes) (Additional file 5). The functional annotations for GO terms and KEGG pathways offer a useful resource for future studies of specific pathways, cellular structures and protein functions of *P. viridis*. Subgroups of genes potentially involved in stress-associated responses and detoxification of xenobiotics are of particular interest to environmental researchers and ecotoxicologists.

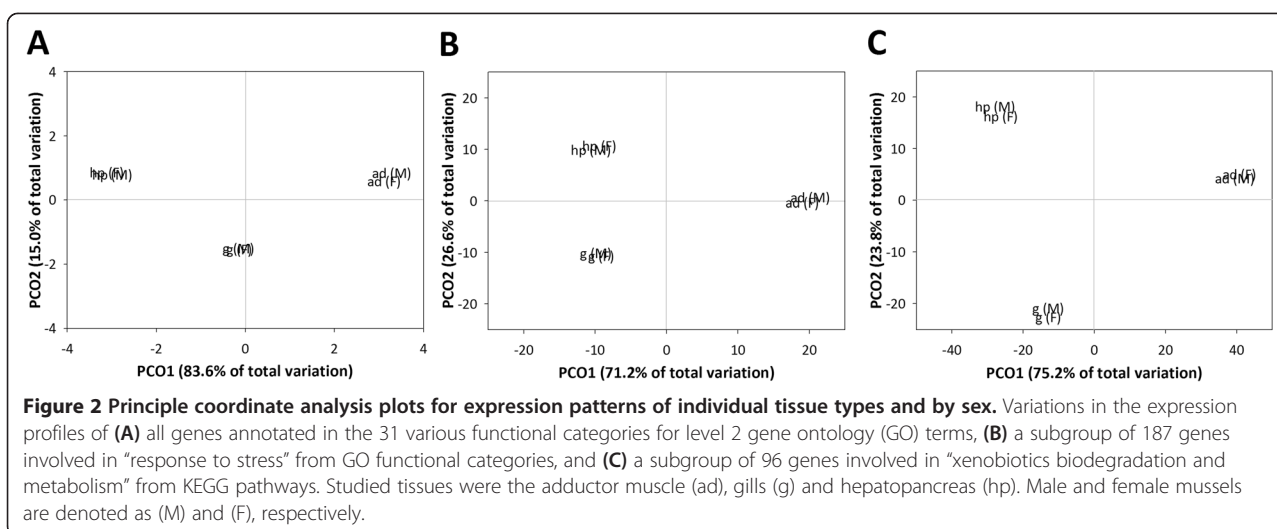
**Table 3 Summary of the annotation results**

Item	Number	
	Transcript	Unigene
With BLAST hit	44,713	25,319
With InterProScan result	93,668	66,692
With GO annotation	21,262	11,947
With KEGG annotation	5,131	3,181
KEGG pathways involved	329	329

**Functional genes related to stress-associated response and xenobiotic degradation**

*Tissue-specific expression patterns*

Based on GO term assignment and KEGG annotation, 187 assembled unigenes were annotated in “response to stress” under the level 3 classification for “biological process”, while 96 assembled unigenes were associated with the pathways of “xenobiotics biodegradation and metabolism” under the KEGG metabolism pathway. It is



believed that genes from these two groups are highly relevant to the responses of *P. viridis* upon challenge with chemicals or other environmental factors. Multivariate analysis using principle coordinate analysis plot (PCO) revealed that expression patterns of the genes from these two categories exhibited distinct tissue-specific patterns (Figure 2B and C). Additionally, the top 20 most expressed assembled unigenes of each tissue type from the two categories representing > 90% and > 60% of genes expressing from the “response to stress” and “xenobiotics biodegradation and metabolism” groups, respectively, are summarized (Tables 4 and 5). Compared with the top 20 genes under “response to stress” in each tissue, relatively higher expression levels were reported for genes in the hepatopancreas. In “xenobiotics biodegradation and metabolism”, the top 20 genes in the gills and hepatopancreas exhibited higher expression levels than those in the adductor muscle (Figure 3).

The expression of some of the selected assembled unigenes involved in stress-associated responses and xenobiotic degradation were further validated using quantitative PCR (qPCR) assays (Additional file 6). Most of these genes also demonstrated tissue-specific expression patterns with generally higher expression levels noticed in the hepatopancreas and/or the gills compared with those in the adductor muscle.

#### Heat-shock protein (HSP) genes

Heat-shock proteins (HSPs) are a group of stress-inducible chaperones that play crucial roles in modulating stress responses and offering cytoprotection [30]. HSPs in bivalves can be induced by various environmental stressors and chemical contaminants [31-34]. A number of HSPs were abundant among the top expressed genes in the “response to stress” category. These included *HSP71*, *HSP90*, *HSP60*, *HSP24.1*, *HSP22*, and other smaller HSPs.

Tissue-specific expression patterns of HSPs have been reported in bivalve species. At basal levels, the *HSP70* gene in the oyster *Crassostrea hongkongensis* was more highly expressed in muscle than in other tissues such as gills, the digestive gland, mantle and heart. This gene was induced differentially and in a time-dependent manner in various tissues, with a faster and more potent response noticed in the digestive gland upon heat stress [32]. Expression of some of the identified HSP genes in individual tissues were also validated using qPCR analysis (Additional file 6). *HSP60* was expressed to a greater extent in the gills and hepatopancreas than in the adductor muscle, and was generally up-regulated by chemical exposure (Table 4, Additional file 6). The small HSP encoded by *HSP22* was present at high levels in the adductor muscle, and the protein encoded by *HSP24.1* was highly expressed in the gills and hepatopancreas but generally down-regulated by chemical stressors (Table 4, Additional file 6). A previous study suggested that some of the small HSPs in the mussel would play an essential chaperone role in maintaining cytoskeletal structural elements under adverse conditions induced by a combination of stressors when other HSPs, such as *HSP70* and *HSP90*, were suppressed [35]. Other genes encoding HSPs were also annotated in the present study (Additional file 7), including members from the *HSP40*, *HSP60*, *HSP70*, *HSP90* classes; a total of 19 members of HSPs were annotated for *P. viridis*.

#### Cytochrome P450 (CYP) genes

With respect to detoxification, members involved in the metabolism of xenobiotics via the cytochrome P450 pathway were particularly abundant among the top expressed genes under the KEGG pathways of “xenobiotics biodegradation and metabolism”, making up 40%, 60% and 50% of that category in the adductor muscle, gills and hepatopancreas of *P. viridis* respectively. The CYP genes encode



**Table 4 20 most abundant annotated genes involved in “response to stress” from individual tissues**

BLAST top-hit description	FPKM*		
	Adductor muscle	Gills	Hepatopancreas
2'-5'-oligoadenylate synthetase 3	0.78	7.25	<b>38.09</b>
Activator of 90 kDa heat shock protein ATPase-like protein 1	<b>16.40</b>	<b>28.60</b>	<b>26.20</b>
B Chain B, Solution Structure Of S5a Uim-1 ubiquitin complex	<b>1693.91</b>	<b>1439.25</b>	<b>1004.83</b>
Catalase	10.49	18.55	<b>57.91</b>
Defensin-B	0.70	5.25	<b>5169.03</b>
Eosinophil peroxidase	0.28	<b>57.76</b>	1.04
Epididymal secretory glutathione peroxidase	0.01	1.76	<b>339.67</b>
Flap endonuclease 1-A	8.54	<b>20.45</b>	13.54
Glucose-regulated protein 94	<b>136.47</b>	<b>119.02</b>	0.19
Glutathione peroxidase 3	0.06	1.62	<b>872.96</b>
Growth arrest and DNA-damage-inducible protein GADD45 alpha	<b>29.06</b>	<b>54.65</b>	11.08
Growth arrest and DNA-damage-inducible protein GADD45 gamma	<b>32.61</b>	10.91	5.17
Heat shock factor binding protein 1	<b>23.82</b>	<b>354.84</b>	<b>138.25</b>
Heat shock protein 60	<b>15.56</b>	<b>57.83</b>	<b>51.70</b>
Heat shock protein 71	<b>1664.60</b>	<b>1676.22</b>	<b>1521.06</b>
Heat shock protein 90	<b>375.18</b>	<b>712.16</b>	<b>633.78</b>
Histone H3.3	<b>143.32</b>	<b>781.19</b>	<b>437.40</b>
Hypothetical protein CGI_10001640	<b>25.81</b>	<b>89.80</b>	<b>117.91</b>
Hypothetical protein CGI_10003818	0.22	0.10	<b>35.65</b>
Hypothetical protein CGI_10015993	<b>11.76</b>	1.71	5.98
Peroxidasin	0.22	<b>41.80</b>	0.27
Radical S-adenosyl methionine domain-containing protein 2	1.94	13.21	<b>26.31</b>
Ras-related C3 botulinum toxin substrate 1, partial	<b>15.05</b>	<b>26.70</b>	15.43
Retinal dehydrogenase	<b>21.62</b>	9.43	<b>43.25</b>
Selenium-dependent glutathione peroxidase	2.95	<b>26.33</b>	<b>68.00</b>
Small heat shock protein 22	<b>1023.03</b>	0.90	1.16
Small heat shock protein 24.1	<b>277.20</b>	<b>443.02</b>	<b>518.55</b>
Stress response protein nhaX	<b>61.17</b>	0.19	0.20
Unnamed protein product	5.23	<b>100.72</b>	13.63
UPS-like protein isoform 1	<b>26.24</b>	<b>42.42</b>	20.43
USP-like protein isoform 2	<b>23.19</b>	<b>73.67</b>	<b>120.01</b>
UV excision repair protein RAD23 homolog B-like	<b>48.10</b>	<b>46.30</b>	<b>36.59</b>

Note: \*Expression levels are presented as FPKM values. The top 20 genes for each tissue are in boldface type.

one of the largest groups of enzymes that catalyze the oxidation of organic substances; they and play a central role in the Phase I detoxification system, and have a range of diverse functions in endogenous metabolism [36]. *CYPs* are classified into different clans according to homology between families; 11 clans were reported for animals [37], with at least four main clans (*CYP2*, *CYP3*, *CYP4* clans and mitochondrial clan) being found in protostomes including mollusks [36,38]. In our study, 14 families (33 members) of *CYP* genes were annotated for *P. viridis* (Additional file 8). Given the important role of *CYP* genes

in the degradation of xenobiotics and other diverse functions, the understanding of this vast group of genes is important in environmental studies. In recent years, the increasing deposition of sequences through large-scale EST and genome projects has facilitated the identification of *CYP* genes in bivalve species. Zanette et al. [39] described *CYP* genes for some mussel and oyster species, included 58 members in *Mytilus californianus*, 12 members in *M. galloprovincialis*, and 14 members in *Crassostrea virginica*. Guo et al. [40] also identified 33 families of *CYP* genes, comprising 88 non-redundant members in the

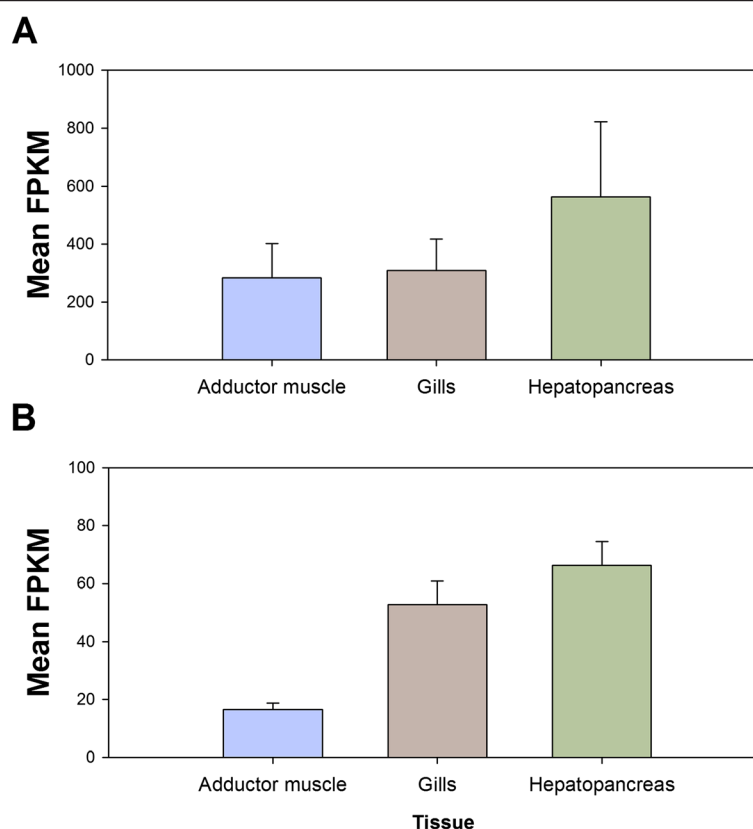
**Table 5 20 most abundant annotated genes involved in “xenobiotics biodegradation and metabolism” in individual tissues**

BLAST top-hit description	FPKM*		
	Adductor muscle	Gills	Hepatopancreas
3-ketoacyl-CoA thiolase, mitochondrial	2.68	18.29	<b>44.68</b>
Acetyl- acetyltransferase mitochondrial	<b>16.01</b>	<b>21.24</b>	10.54
Alcohol dehydrogenase class-3	<b>53.46</b>	<b>30.34</b>	<b>81.41</b>
Aldehyde mitochondrial	<b>7.55</b>	<b>47.66</b>	<b>42.06</b>
ATPase family AAA domain-containing protein 2B	<b>11.38</b>	<b>39.79</b>	<b>56.63</b>
Carbonyl reductase	<b>16.1</b>	<b>68.81</b>	<b>59.44</b>
Cytochrome p450	0.62	11.38	<b>35.83</b>
Cytochrome p450 2c8	1.01	<b>40.79</b>	28.08
Cytochrome p450 2 k1-like	1.65	<b>27.14</b>	12.77
Cytochrome p450 3a-like isoform 1	2.76	4.11	<b>46.21</b>
Cytochrome p450 4f22	<b>14.48</b>	5.51	<b>35.58</b>
Dihydropyrimidinase	<b>21.62</b>	17.10	2.35
Dimethylaniline monooxygenase	1.52	4.14	<b>46.37</b>
Fumarylacetoacetase	<b>8.75</b>	18.67	8.63
Glutathione s-transferase alpha	<b>18.34</b>	<b>34.55</b>	<b>136.66</b>
Glutathione s-transferase pi 1	<b>16.12</b>	15.73	<b>44.15</b>
Glutathione s-transferase sigma 2	0.58	<b>176.58</b>	2.80
Glutathione s-transferase zeta	<b>20.35</b>	<b>21.10</b>	12.28
Glutathione-s-transferase omega class	0.54	<b>56.96</b>	<b>56.97</b>
Hydroxyacyl-coenzyme a mitochondrial	<b>16.01</b>	<b>20.28</b>	<b>56.62</b>
Hypothetical protein BRAFLDRAFT_86061	<b>19.69</b>	<b>65.68</b>	<b>70.07</b>
Hypothetical protein CAPTEDRAFT_149199	<b>8.35</b>	<b>68.49</b>	32.48
Microsomal glutathione s-transferase 2	5.76	<b>20.62</b>	26.62
Microsomal glutathione s-transferase 3	<b>16.04</b>	<b>44.71</b>	26.77
Microsomal glutathione-s-transferase	2.66	<b>38.78</b>	32.88
Monoamine oxidase A	<b>11.64</b>	2.73	12.58
Predicted protein	<b>9.98</b>	<b>99.6</b>	21.26
Regucalcin	0.07	0.26	<b>149.99</b>
Sodium- and chloride-dependent GABA transporter ine	0.62	<b>63.33</b>	<b>143.35</b>
Sorbitol dehydrogenase	<b>7.25</b>	3.22	<b>48.91</b>
Thymidine phosphorylase	4.77	13.95	<b>36.81</b>
Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase	2.67	5.90	<b>38.77</b>
Trifunctional enzyme subunit beta, mitochondrial-like	<b>26.54</b>	<b>68.98</b>	<b>95.02</b>
Uridine-cytidine kinase-like 1	<b>9.03</b>	16.26	0.35

Note: \*Expression levels are presented as FPKM values. The top 20 genes for each tissue are in boldface.

Zhikong scallop *Azumapecten farreri* (formerly known as *Chlamys farreri*) via RNA sequencing analysis. With the completion of the *C. gigas* genome project, the number of CYP genes identified in oysters has further increased to 136 genes [16]. The majority of annotated CYP genes (11 members) in *P. viridis* were identified as members of clan 2 (Additional file 8).

The CYP genes of marine invertebrates, including mollusks, polychaetes and crustaceans, have been observed in a wide range of tissues and are generally enriched in organs associated with food processing [36,41]. CYPs show tissue specificity with respect to their functions; for instance, CYPs in hepatic tissues are generally involved in the detoxification of a wide range of xenobiotics [36]. A



**Figure 3 Mean expression levels of the 20 most abundant annotated genes from individual tissue types.** Sub-groupings are based on the annotation to (A) "response to stress" and (B) "xenobiotics biodegradation and metabolism" from KEGG pathways. Expression levels are presented as mean FPKM values + S.E. for the top 20 most abundant annotated genes.

member from *CYP* family 3 in the xenobiotics degradation pathway was one of the top 20 expressed genes in the hepatopancreas of *P. viridis*. The mRNA expression levels of this *CYP* were highest in this tissue (Additional file 6). A member of the *CYP12* family belonged to the mitochondrial clan and showed high expression levels in the gills, and was up-regulated by chemical stressors (Additional file 6). Although the function of *CYP12* genes has not been extensively studied in mussels, previous reports suggest that these genes are capable of metabolizing xenobiotics such as pesticides in arthropods [42,43].

#### Glutathione S-transferase (GST) genes

Another predominant group contributing to the KEGG pathways of "xenobiotic biodegradation and metabolism" was the group of genes encoding GSTs. These genes comprised 25, 35 and 15% of the top expressed genes in the gills, adductor muscle, and hepatopancreas, respectively (Table 5). The GST proteins are a group of multi-functional enzymes that catalyze the conjugation and detoxification of xenobiotics during phase II enzyme detoxification [44]. Three major families of GSTs (cytosolic, mitochondrial, and microsomal) and at least 15 different

classes of GSTs (alpha, beta, delta, epsilon, kappa, lambda, mu, omega, phi, pi, sigma, tau, theta, zeta, and rho) have been identified in numerous organisms based on substrate specificity, and catalytic and immunological properties [44]. To date, only a limited number of GSTs have been characterized in mussels. GST pi and omega were identified in *P. viridis* [19], while alpha, pi and sigma classes have been identified in *M. galloprovincialis* [45,46], and pi class GST was found in *M. edulis* [47]. Characterization of GSTs has been conducted for other marine bivalves, such as oysters (*C. gigas* [48] and *C. ariakensis* [49]), and clams (*R. philippinarum* [50-52], *Laternula elliptica* [53,54], *Mercenaria mercenaria* [55], *Solen grandis* [56] and *A. farreri* [57]). From our analysis 19 GSTs were annotated for *P. viridis*, with most of them matching entries for *C. gigas* and *M. galloprovincialis* (Additional file 9). At least six classes of GSTs (alpha, pi, sigma, omega, theta and zeta), along with microsomal GSTs, and the C- and N-terminal domains were annotated for *P. viridis* (Additional file 9).

The tissue-specific expression patterns of GST genes in bivalves have been previously shown [19,51]. A recent study on *P. viridis* showed that GST pi and omega



classes showed higher transcriptional levels in the hepatopancreas than in the gonad, gills and mantle [19]. It was also reported that *GST* pi and *GST* sigma 1 in *M. galloprovincialis* were more highly expressed in the hepatopancreas than in the gills; *GST*s alpha, sigma 2 and sigma 3 were found to be more abundant in hemocytes than in the hepatopancreas, gills, mantle, gonad and muscle [45,46]. We also noticed strong tissue-specific expression patterns for the *GST* genes of *P. viridis* (Table 5). These strong tissue-specific expression patterns were also demonstrated by gene expression analysis for some of the selected *GST* members from the three tissues we examined (Additional file 6). Studies involving microarrays have also shown that differential expression of various *GST* genes from the gills and digestive gland of *R. philippinarum* were detected in samples collected from polluted areas [13]. Our novel findings regarding various *GST* genes in *P. viridis* clearly show that tissue- and class-specificity should be considered in future application of *GST* genes as biomarkers in biomonitoring.

## Conclusions

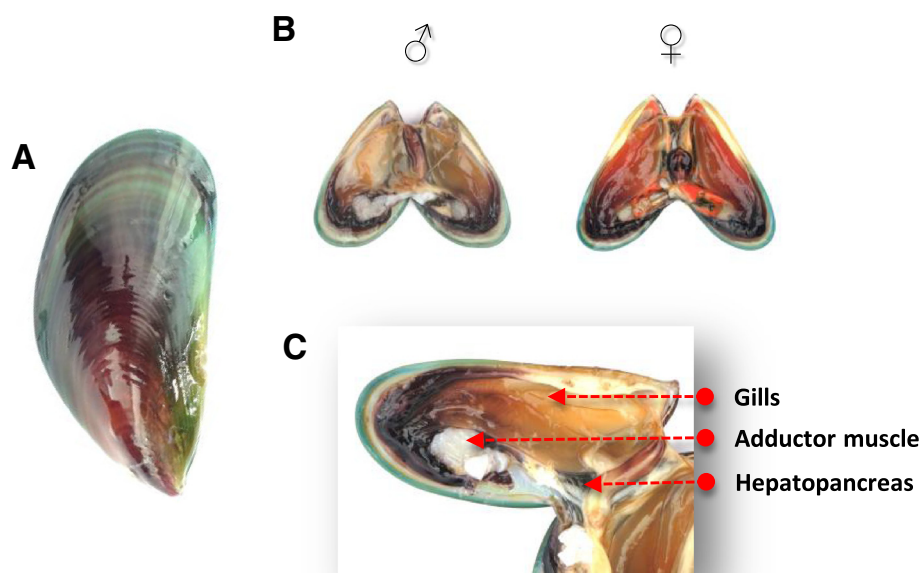
A *de novo* transcriptome of *P. viridis* has been successfully generated; this transcriptome was based on the mRNA from three target tissues (adductor muscle, gills and hepatopancreas). Tissue-specific patterns were associated with transcriptome size, expression profiles of genes in GO terms, and subgroups of genes associated with stress responses and degradation of xenobiotics. Our findings provide an informative transcriptomic platform for further

in-depth mechanistic environmental studies using the biomonitor *P. viridis*, and at the same time emphasizes the advantage of tissue-specific approaches in future investigations.

## Methods

### Sample collection and preparation

We obtained adult *P. viridis* (4–5 cm shell length; Figure 4) for RNA sequencing from the field (field-type) and from mariculture farms. Animals from mariculture farms were exposed to various physical and chemical stressors (exposed-type) so as to obtain a wide spectrum of transcripts associated with environmental conditions. A similar approach was employed in the generation of other transcriptome databases [58]. According to the Animals (Control of Experiments) Ordinance, Chapter 340 (Department of Health, Hong Kong SAR), there is no requirement for permission to perform experiments involving mussels (invertebrate). The field-type samples were collected from different locations in Hong Kong waters. Locations encompassed the western (Butterfly Beach, 22°22'19.73"N and 113°57'38.31"E), southern (Po Toi, 22°9'49.61"N and 114°15'7.73"E) and eastern waters (Tung Lung Island, 22°15'24.16"N and 114°17'13.76"E; Sam Mun Tsai, 22°27'8.87"N and 114°12'28.90"E) as each area had different hydrographic condition. The field-type samples were dissected to obtain target tissues (gills, adductor muscle and hepatopancreas) within 24 h of collection. Sex was determined by examining the external characteristic of the gonad, that is a female gonad



**Figure 4 Morphology of *Perna viridis*.** (A) The external morphology of the mussel. (B) The internal morphology of male and female mussels. (C) The three tissues examined were the gills, adductor muscle and hepatopancreas.

was in bright orange brick red colour while male gonad was in milky colour [9].

The exposed-type mussels were obtained from a mariculture farm at Yung Shue O mariculture zone, Sai Kung, Hong Kong (22°25'26"N, 114°16'45"E). Acclimation and exposure treatments were conducted at the Swire Institute of Marine Science (SWIMS; The University of Hong Kong). Prior to exposure experiments, mussels were acclimated in running seawater for 7 days, and fed with the diatom *Thalassiosira pseudonana*. Mussels were then acutely exposed (24 h) to different individual model chemicals (metals, endocrine disruptors, organic pollutants and engineered nano-particles) and to different individual physical conditions (temperature, salinity, dissolved oxygen level and pH). Sublethal concentrations were used for chemical exposure, with concentrations set at the corresponding median effective concentration or 10% of the corresponding median lethal concentration available for *P. viridis* or other marine bivalves from the USEPA's ECOTOX database (<http://cfpub.epa.gov/ecotox/>) and literatures. For the exposure experiments, mussels were treated with different types of chemical: metals, 730 µg/L CdCl<sub>2</sub>, (ECOTOX) and 65 µg/L CuSO<sub>4</sub> (ECOTOX); endocrine disruptors, 0.32 µg/L triphenyltin chloride (TPTCl; ECOTOX) [59] and 40 µg/L dichlorodiphenyltrichloroethane (DDT; ECOTOX); organic pollutants, 5 µg/L polybrominated diphenyl ether 47 (PBDE 47; ECOTOX) [60] and 75 µg/L Benzo[a]pyrene (ECOTOX); and engineered nano particle, 13 mg/L nZnO (ECOTOX) designed to provide 600 µg/L zinc ions. For nZnO treatment, the concentration of zinc ions was estimated from the calculated dissolution rate of nZnO [61]. Mussels were kept in test solutions in 0.45-µm filtered artificial seawater (FASW; salinity: 30 ± 0.5‰; temperature: 25 ± 0.5°C and pH 8.0 ± 0.1). Organic compounds (TPTCl, DDT, PBDE 47 and Benzo[a]pyrene) were delivered into FASW using dimethylsulfoxide (DMSO) as a solvent at a final concentration of 0.01% (v/v).

Additionally, mussels were exposed to a range of: temperatures (15, 25 and 30°C) in FASW (30 ± 0.5‰; pH 8.0 ± 0.1); salinities (10, 20 and 30‰) in FASW (25 ± 0.5°C; pH 8.0 ± 0.1); dissolved oxygen (DO; 2.5 and 4.0 mg/L) in FASW (30 ± 0.5‰; 25 ± 0.5°C; pH 8.0 ± 0.1); and pH (7.5, 7.8 and 8.1) in FASW (30 ± 0.5‰; 25 ± 0.5°C). The mussels used in all treatments were kept at a density of one mussel per 250 mL of FASW. After 24 h of exposure, mussels were immediately dissected to obtain target tissues.

Two mussels, a male and a female, from each field site and from each exposure treatment ( $n = 44$  mussels) were chosen for dissection to obtain the tissue samples of adductor muscle (ad), gills (g) and hepatopancreas (hp). These three types of tissue we examined were common targets that have been used in environmental studies [3].

Dissected tissue samples were immediately fixed in RNAlater® (Applied Biosystems, Warrington, UK), and then stored at -20°C until required.

#### RNA preparation and sequencing

Total RNA for each sample was extracted with the RNeasy® Mini Kit and digested with DNase I following the manufacturer's instructions (QIAGEN, Hilden, Germany). Integrity and size distribution of RNA samples were verified using an Agilent 2100 Bioanalyser (Agilent Technologies, Germany). Samples with an RNA Integrity Number ≥ 8.0 were used for cDNA library preparation. The concentration of RNA in each extracted sample was measured using a Qubit 1.0 Fluorometer (Invitrogen, Carlsbad, USA).

Six cDNA libraries were generated from the three target tissues (male-hp, female-hp, male-g, female-g, male-ad, and female-ad), with each library comprising pooled samples of field-type and exposed-type mussels. The cDNA libraries were assigned to six individual sequencing lanes, and one dedicated PhiX control lane for matrix and phasing calculations was applied during sequencing. The cDNA libraries were prepared using Illumina® TruSeq RNA Sample Preparation Kits v2 (Catalog # RS-122-2001), and the Illumina sequencing was performed at the Centre of Genomic Sciences (The University of Hong Kong). In brief, poly-A containing mRNA was collected using poly-T oligo-attached magnetic beads. The purified mRNA was broken down into short fragments and was applied as a template to synthesize the first-strand cDNA using a random hexamer-primer and reverse transcriptase (SuperScript® II Reverse Transcriptase; Invitrogen, Catalog # 18064014). For second-strand cDNA synthesis, the mRNA template was removed and a replacement strand was generated to form double-stranded (ds) cDNA. The ds cDNA underwent end repair, 3' adenylation and indexed adaptor ligation, and the adaptor-ligated libraries were enriched by 10 cycles of PCR. Qualified and quantified libraries were then sequenced using an Illumina Genome Analyzer IIx (GAIIx; Illumina, California, USA). Image analysis and base calling was performed with SCS2.8/RTA1.8. FASTQ file generation and removal of failed reads were performed using CASAVA version 1.8.2.

#### De novo assembly and annotation

The raw reads from the six samples were merged, and then each read was pre-processed by the trimming of adaptors and the removal of 10 bp from the 3' end. Low-quality reads were filtered out based on three criteria: reads with more than 5% unknown "N" bases; reads having more than 50% bases with a quality value less than 10; and reads < 35 bp. The resulting high-quality reads were then used for *de novo* assembly with Trinity version trinityrnaseq\_r2012-10-25 [62] using default parameters.

Trinity used the greedy k-mer extension (k-mer 25) to assemble raw reads into unique sequences of transcripts, followed by clustering of the contigs to construct complete de Bruijn graphs, and finally processed the individual graphs to generate full-length transcripts. The assembled transcripts were further clustered using CD-HIT [63] to reduce redundancy. Briefly, CD-HIT employed a 'longest sequence first' algorithm by identifying a representative sequence and then clustered each subsequent sequence as a redundant or a new cluster sequence based on a similarity score set at  $\geq 95\%$  for our analysis. The longest transcript in each gene cluster defined by Trinity was selected as a "unigene" for downstream analysis.

The annotation of assembled transcripts/unigenes was performed based on sequence similarity searches using BLASTx against the NCBI nr database (08 February 2013), and BLASTn and tBLASTn searches against the EST databases from the molluscan genome projects of *Aplysia californica* (GenBank Accession Number AASC00000000, 255,605 EST entries), *Crassostrea gigas* (AFTI00000000, 206,647 EST entries) and *Lottia gigantea* (AMQO00000000, 252,091 EST entries), with an E-value threshold of  $1 \times 10^{-5}$  and a high-scoring segment pairs cut-off length of 33. Functional classification was carried out based on GO using Blast2GO [64]. GO terms were assigned and annotated to the transcripts/unigenes according to three main categories: cell component; molecular function; and biological process. The presence of a conserved domain in transcript/unigene was analysed using InterProScan [65] within Blast2GO. KEGG pathways were assigned to the transcripts using the online KEGG Automatic Annotation Server (KAAS; <http://www.genome.jp/kegg/kaas/>) using the bi-directional best-hit method.

#### Transcriptome quantification for individual tissues and sexes

The high-quality reads from the six libraries (2 sexes  $\times$  3 tissues) were individually mapped to the global assembled *de novo* transcriptome to estimate the abundance of transcripts/unigenes using RNA-Seq by Expectation Maximization (RSEM) version 1.2.3 [66] with a pipeline script provided by Trinity. High-quality reads were mapped to the transcripts and the resulting alignment file was inputted into RSEM to calculate the expression value for each transcript. Expression level was presented in a normalized expressed value as FPKM, which corrects for read differences in libraries and gene length. A transcript was only considered expressed when the FPKM cut-off value was  $\geq 0.05$ .

#### Real-time qPCR assays

Tissue-specific expression patterns for 15 selected genes (Additional file 10) associated with stress responses/

detoxification were revealed using qPCR analysis. Mussel samples were chosen from selected treatments (i.e., control, 730  $\mu\text{g/L}$   $\text{CdCl}_2$ , 13  $\text{mg/L}$   $\text{nZnO}$ , 40  $\mu\text{g/L}$  DDT, and 0.32  $\mu\text{g/L}$  TPTCI) of the exposed-type that used in the library generation. Primer pairs for qPCR assays were designed using Primer3 version 0.4.0 (<http://bioinfo.ut.ee/primer3-0.4.0/>). For amplification, 1 mM reverse and forward primers and 2  $\mu\text{L}$  of cDNA were used per 20  $\mu\text{L}$  reaction on a CFX96™ Real-Time System (Bio-Rad, Hercules, USA) with iQ™ SYBR® green supermix (Bio-Rad, Philadelphia, USA), following the manufacturer's recommended protocol and default settings. Thermal cycling conditions involved 3 min of pre-heating at 95°C, followed by 40 cycles of amplification (10 s at 95°C, 30 s at 60°C); a melt curve analysis was performed after the 40th cycle (10 s at 95°C, then 65–95°C in 0.5°C increments every 5 s). Relative expression levels of target genes were calculated using the  $2^{-\Delta\Delta C_T}$  method [67], with the 18S rRNA gene used as a reference and pooled samples from all tissues as calibrator samples. The 18S gene was shown to be a suitable reference gene in mussels and its expression was stable among different chemical treatments [18,68]. Expression of the 18S gene was not significantly different among the various control and treatment groups (ANOVA,  $p > 0.05$ ), and between sexes ( $t$ -test,  $p > 0.05$ ).

#### Availability of supporting data

All supporting data are included as additional files. The raw sequencing reads were submitted to the NCBI Sequence Read Archive (Accession No. SRP043984).

#### Additional files

**Additional file 1:** Summary of raw reads data for *Perna viridis*.

**Additional file 2:** BLAST top-hit species distribution for assembled transcripts of *Perna viridis*. Note: *Hydra vulgaris* was formerly classified as *Hydra magnipapillata*.

**Additional file 3:** Gene ontology (GO) annotations for the global transcriptome of *Perna viridis*. GO terms were annotated at level 2 of classification according to three main categories (biological process, cellular component, and molecular function).

**Additional file 4:** Gene ontology (GO) annotations for the transcriptome of individual tissues and by sex. GO terms were annotated at level 2 of classification according to three main categories (biological process, cellular component, and molecular function).

**Additional file 5:** Distribution of genes from *Perna viridis* with putative protein annotations assigned to the KEGG pathways.

**Additional file 6:** Real-time qPCR analysis. Relative mRNA expression levels of selected stress-associated genes from the gills, adductor muscle and hepatopancreas of *Perna viridis* after five different treatments (control; 730  $\mu\text{g/L}$   $\text{CdCl}_2$ ; 13  $\text{mg/L}$   $\text{nZnO}$ ; 40  $\mu\text{g/L}$  DDT; or 0.32  $\mu\text{g/L}$  TPTCI). *HSP*, heat shock protein; *GSTa*, glutathione-S-transferase alpha-class; *GSTp*, glutathione-S-transferase pi-class; *GSTs*, glutathione-S-transferase sigma-class; *GSTo*, glutathione-S-transferase omega-class; and *CYP*, cytochrome p450. Expression levels are presented as mean relative mRNA expression level  $\pm$  S.E. ( $n = 4$ ). Significant differences are denoted by uppercase letters (A, B and C) for tissues and lowercase letters (a, b and c) for treatments within each tissue (one-way ANOVA, SNK test;  $p < 0.05$ ).



**Additional file 7: List of heat-shock protein (HSP) genes annotated in *Perna viridis*.**

**Additional file 8: List of cytochrome p450 (CYP) genes annotated in *Perna viridis*.**

**Additional file 9: List of glutathione S-transferase (GST) genes annotated in *Perna viridis*.**

**Additional file 10: List of oligonucleotide primers used in real-time qPCR assays.**

## Abbreviations

Bp: Base pairs; Cd: Cadmium; CdCl<sub>2</sub>: Cadmium chloride; CD-HIT: Cluster Database at High Identity with Tolerance; C<sub>t</sub>: Threshold cycle; CuSO<sub>4</sub>: Copper (II) sulphate; CYP: Cytochrome P450; DDT: Dichlorodiphenyltrichloroethane; DMSO: Dimethylsulfoxide; DO: Dissolved oxygen; DOPA: 3, 4-dihydroxyphenyl-L-alanine; ds: Double-stranded; EST: Expressed Sequence Tag; FASW: Filtered artificial seawater; FPKM: Fragments Per Kilo bases per Million reads; GAllx: Genome Analyzer Iix; GO: Gene Ontology; GST: Glutathione S-transferase; HSP10: 10 kDa heat shock protein; HSP22: Small heat shock protein 22; HSP24.1: Small heat shock protein 24.1; HSP60: Heat shock protein 60; HSP70: Heat shock cognate 70 kDa protein; HSP71: Heat shock 70 kDa protein 8; HSP90: Heat shock protein 90; HSP: Heat-shock protein; KASS: KEGG Automatic Annotation Server; KEGG: Kyoto Encyclopedia of Genes and Genomes; N50: 50% of the length of the assembled sequences; NCBI nr database: National Center for Biotechnology Information Non-redundant database; NGS: Next-generation sequencing; nZnO: Nano zinc oxide; PBDE 47: Polybrominated diphenyl ether 47; PCO: Principle coordinate analysis plot; PCR: Polymerase chain reaction; Q30: Quality score of 30; RT-qPCR: Real-time quantitative PCR; RSEM: RNA-Seq by Expectation Maximization; SRA: High-throughput DNA and RNA sequence read archive; SWIMS: Swire Institute of Marine Science; TPT: Triphenyltin; TPTCl: Triphenyltin chloride; Zn: Zinc.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

PTYL designed, performed and coordinated the experiments and wrote most of the manuscript. JCHL performed the experiments and participated in part of the writing of the manuscript. SSTM participated in gene expression analysis and data handling. JWQ and CKCW helped design and revise the manuscript. PKSL, LLC and KMYL approved the experimental design and revised the manuscript. KMYL was the principal investigator of this project. All authors read and approved the final manuscript.

## Acknowledgements

This work was supported by a research grant awarded to KMY Leung, a Seed Collaborative Research Fund (2011) from the State Key Laboratory in Marine Pollution via Innovation and Technology Commission. JCH Ip thanks the University of Hong Kong (HKU) for providing the Type B postgraduate studentship and support from the Research Grants Council of the Hong Kong SAR Government via a General Research Fund (Project No.: HKU 771212 M) awarded to KMY Leung. The authors thank the Centre for Genomic Sciences, HKU for the service support in Illumina sequencing and bioinformatics, Edward Lau for proofreading early drafts of this manuscript, and Adela Li and Andy Yi for their assistance in the laboratory exposure experiments.

## Author details

<sup>1</sup>The Swire Institute of Marine Science and School of Biological Sciences, The University of Hong Kong, Pokfulam, Hong Kong, China. <sup>2</sup>State Key Laboratory in Marine Pollution, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China. <sup>3</sup>Department of Biology, Hong Kong Baptist University, Kowloon, Hong Kong, China. <sup>4</sup>Department of Biology and Chemistry, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China.

Received: 21 May 2014 Accepted: 26 August 2014

Published: 19 September 2014

## References

- Goldberg ED, Bowen VT, Farrington JW, Harvey G, Martin JH, Parker PL, Risebrough RW, Robertson W, Schneider E, Gamble E: **Mussel watch.** *Environ Conserv* 1978, **5**(2):101–125.
- Livingstone DR, Chipman JK, Lowe DM, Minier C, Mitchelmore CL, Moore MN, Peters LD, Pipe RK: **Development of biomarkers to detect the effects of organic pollution on aquatic invertebrates: recent molecular, genotoxic, cellular and immunological studies on the common mussel (*Mytilus edulis* L.) and other mytilids.** *Int J Environ Pollut* 2000, **13**(1–6):56–91.
- Nicholson S, Lam PKS: **Pollution monitoring in Southeast Asia using biomarkers in the mytilid mussel *Perna viridis* (Mytilidae : Bivalvia).** *Environ Int* 2005, **31**(1):121–132.
- Yap CK, Shahbazi A, Zakaria MP: **Concentrations of heavy metals (Cu, Cd, Zn and Ni) and PAHs in *Perna viridis* collected from seaport and non-seaport waters in the Straits of Johore.** *Bull Environ Contam Toxicol* 2012, **89**(6):1205–1210.
- Shahbazi A, Zakaria MP, Yap CK, Tan SG, Surif S, Mohamed CAR, Sakari M, Bakhtiar AR, Bahry PS, Chandru K, Mirsadeghi SA: **Use of different tissues of *Perna viridis* as biomonitors of polycyclic aromatic hydrocarbons (PAHs) in the coastal waters of Peninsular Malaysia.** *Environ Forensics* 2010, **11**(3):248–263.
- Fang JKH, Wu RSS, Zheng GJ, Lam PKS, Shin PKS: **Seasonality of bioaccumulation of trace organics and lysosomal integrity in green-lipped mussel *Perna viridis*.** *Sci Total Environ* 2010, **408**(6):1458–1465.
- Tang CH, Wang WH: **Butyltin accumulation in two marine bivalves along a pollution gradient.** *Environ Toxicol Chem* 2008, **27**(10):2179–2185.
- Shek WM, Murphy MB, Lam JCW, Lam PKS: **Polycyclic musks in green-lipped mussels (*Perna viridis*) from Hong Kong.** *Mar Pollut Bull* 2008, **57**(6–12):373–380.
- Rajagopal S, Venugopalan VP, van der Velde G, Jenner HA: **Greening of the coasts: a review of the *Perna viridis* success story.** *Aquat Ecol* 2006, **40**(3):273–297.
- Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**(1):57–63.
- Craft JA, Gilbert JA, Temperton B, Dempsey KE, Ashelford K, Tiwari B, Hutchinson TH, Chipman JK: **Pyrosequencing of *Mytilus galloprovincialis* cDNAs: tissue-specific expression patterns.** *PLoS One* 2010, **5**(1):e8875.
- Bettencourt R, Pinheiro M, Egas C, Gomes P, Afonso M, Shank T, Santos RS: **High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*.** *BMC Genomics* 2010, **11**:559.
- Milan M, Coppe A, Reinhardt R, Cancela LM, Leite RB, Saavedra C, Ciofi C, Chelazzi G, Patarnello T, Bortoluzzi S, Bargelloni L: **Transcriptome sequencing and microarray development for the Manila clam, *Ruditapes philippinarum*: genomic tools for environmental monitoring.** *BMC Genomics* 2011, **12**:234.
- Hou R, Bao ZM, Wang S, Su HL, Li Y, Du HX, Hu JJ, Hu XL: **Transcriptome sequencing and *de novo* analysis for Yesso scallop (*Patinoptecten yessoensis*) using 454 GS FLX.** *PLoS One* 2011, **6**(6):e21560.
- Kinoshita S, Wang N, Inoue H, Maeyama K, Okamoto K, Nagai K, Kondo H, Hirono I, Asakawa S, Watabe S: **Deep sequencing of ESTs from nacreous and prismatic layer producing tissues and a screen for novel shell formation-related genes in the pearl oyster.** *PLoS One* 2011, **6**(6):e21238.
- Zhang GF, Fang XD, Guo XM, Li L, Luo RB, Xu F, Yang PC, Zhang LL, Wang XT, Qi HG, Xiong ZQ, Que HY, Xie YL, Holland PWH, Paps J, Zhu YB, Wu FC, Chen YX, Wang JF, Peng CF, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang ZY, Zhu QH, Feng Y, Mount A, Hedgecock D: **The oyster genome reveals stress adaptation and complexity of shell formation.** *Nature* 2012, **490**(7418):49–54.
- Zhao XL, Yu H, Kong LF, Li Q: **Transcriptomic responses to salinity stress in the Pacific oyster *Crassostrea gigas*.** *PLoS One* 2012, **7**(9):e46244.
- Leung PTY, Park TJ, Wang Y, Che CM, Leung KMY: **Isoform-specific responses of metallothioneins in a marine pollution biomonitor, the green-lipped mussel *Perna viridis*, towards different stress stimulations.** *Proteomics* 2014, **14**(15):1796–1807.
- Li ZZ, Chen R, Zuo ZH, Mo ZP, Yu A: **Cloning, expression and identification of two glutathione S-transferase isoenzymes from *Perna viridis*.** *Comp Biochem Physiol B Biochem Mol Biol* 2013, **165**(4):277–285.
- Zhao H, Sagert J, Hwang DS, Waite JH: **Glycosylated hydroxytryptophan in a mussel adhesive protein from *Perna viridis*.** *J Biol Chem* 2009, **284**(35):23344–23352.

21. Guerette PA, Hoon S, Seow Y, Raida M, Masic A, Wong FT, Ho VHB, Kong KW, Demirel MC, Pena-Francesch A, Amini S, Tay GZ, Ding D, Miserez A: **Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science.** *Nat Biotechnol* 2013, **31**(10):908–915.
22. Cao YY, Li ZB, Li QH, Chen XJ, Chen L, Dai G: **Characterization of eight novel microsatellite markers in the green-lipped mussel *Perna viridis* (Mytilidae).** *Genet Mol Res* 2013, **12**(1):344–347.
23. Ong CC, Yusoff K, Yap CK, Tan SG: **Genetic characterization of *Perna viridis* L. in peninsular Malaysia using microsatellite markers.** *J Genet* 2009, **88**(2):153–163.
24. Li XL, Wu XY, Yu ZN: **Complete mitochondrial genome of the Asian green mussel *Perna viridis* (Bivalvia, Mytilidae).** *Mitochondr DNA* 2012, **23**(5):358–360.
25. Iqbal A, Khan MS, Goswami U: **Cytogenetic studies in green mussel, *Perna viridis* (Mytiloida : Pteriomorpha), from West Coast of India.** *Mar Biol* 2008, **153**(5):987–993.
26. Johnson BR, Atallah J, Plachetzki DC: **The importance of tissue specificity for RNA-seq: highlighting the errors of composite structure extractions.** *BMC Genomics* 2013, **14**:586.
27. Coppe A, Bortoluzzi S, Murari G, Marino IAM, Zane L, Papetti C: **Sequencing and characterization of striped venus transcriptome expand resources for clam fishery genetics.** *PLoS One* 2012, **7**(9):e44185.
28. Men LN, Yan SC, Liu GJ: **De novo characterization of *Larix gmelinii* (Rupr.) Rupr. transcriptome and analysis of its gene expression induced by jasmonates.** *BMC Genomics* 2013, **14**:548.
29. Meng XL, Liu M, Jiang KY, Wang BJ, Tian X, Sun SJ, Luo ZY, Qiu CW, Wang L: **De novo characterization of Japanese scallop *Mizuhopecten yessoensis* transcriptome and analysis of its gene expression following cadmium exposure.** *PLoS One* 2013, **8**(5):e64485.
30. Feder ME, Hofmann GE: **Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology.** *Annu Rev Physiol* 1999, **61**:243–282.
31. Leung PTY, Wang Y, Mak SST, Ng WC, Leung KMY: **Differential proteomic responses in hepatopancreas and adductor muscles of the green-lipped mussel *Perna viridis* to stresses induced by cadmium and hydrogen peroxide.** *Aquat Toxicol* 2011, **105**(1–2):49–61.
32. Zhang ZH, Zhang QZ: **Molecular cloning, characterization and expression of heat shock protein 70 gene from the oyster *Crassostrea hongkongensis* responding to thermal stress and exposure of Cu<sup>2+</sup> and malachite green.** *Gene* 2012, **497**(2):172–180.
33. Woo S, Jeon HY, Kim SR, Yum S: **Differentially displayed genes with oxygen depletion stress and transcriptional responses in the marine mussel, *Mytilus galloprovincialis*.** *Comp Biochem Physiol Part D Genomics Proteomics* 2011, **6**(4):348–356.
34. Meng J, Zhu QH, Zhang LL, Li CY, Li L, She ZC, Huang BY, Zhang GF: **Genome and transcriptome analyses provide insight into the euryhaline adaptation mechanism of *Crassostrea gigas*.** *PLoS One* 2013, **8**(3):e58563.
35. Negri A, Oliveri C, Sforzini S, Mignione F, Viarengo A, Banni M: **Transcriptional response of the mussel *Mytilus galloprovincialis* (Lam.) following exposure to heat stress and copper.** *PLoS One* 2013, **8**(6):e66802.
36. Rewitz KF, Styrisshave B, Lobner-Olesen A, Andersen O: **Marine invertebrate cytochrome P450: Emerging insights from vertebrate and insect analogies.** *Comp Biochem Physiol C Toxicol Pharmacol* 2006, **143**(4):363–381.
37. Nelson DR: **Progress in tracing the evolutionary paths of cytochrome P450.** *Biochim Biophys Acta* 2011, **1814**(1):14–18.
38. Baldwin WS, Marko PB, Nelson DR: **The cytochrome P450 (CYP) gene superfamily in *Daphnia pulex*.** *BMC Genomics* 2009, **10**:169.
39. Zanette J, Goldstone JV, Baily ACD, Stegeman JJ: **Identification of CYP genes in *Mytilus* (mussel) and *Crassostrea* (oyster) species First approach to the full complement of cytochrome P450 genes in bivalves.** *Mar Environ Res* 2010, **69**:S1–S3.
40. Guo HH, Bao ZM, Du HX, Zhang LL, Wang S, Sun LY, Mou XY, Hu XL: **Identification of Cytochrome P450 (CYP) genes in Zhikong scallop (*Chlamys farreri*).** *J Ocean Univ* 2013, **12**(1):97–102.
41. Sole M, Livingstone DR: **Components of the cytochrome P450-dependent monooxygenase system and 'NADPH-independent benzo a pyrene hydroxylase' activity in a wide range of marine invertebrate species.** *Comp Biochem Physiol C Toxicol Pharmacol* 2005, **141**(1):20–31.
42. Wan PJ, Shi XQ, Kong Y, Zhou LT, Guo WC, Ahmat T, Li GQ: **Identification of cytochrome P450 monooxygenase genes and their expression profiles in cyhalothrin-treated Colorado potato beetle.** *Leptinotarsa decemlineata Pest Biochem Physiol* 2013, **107**(3):360–368.
43. Guzov VM, Unnithan GC, Chernogolov AA, Feyereisen R: **CYP12A1, a mitochondrial cytochrome P450 from the house fly.** *Arch Biochem Biophys* 1998, **359**(2):231–240.
44. Sheehan D, Meade G, Foley V, Dowd C: **Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily.** *Biochem J* 2001, **360**:1–16.
45. Hoarar P, Damians G, Romeo M, Gnassia-Barelli M, Bebianno MJ: **Cloning and expression of a GST-pi gene in *Mytilus galloprovincialis*. Attempt to use the GST-pi transcript as a biomarker of pollution.** *Comp Biochem Physiol C Toxicol Pharmacol* 2006, **143**(2):196–203.
46. Wang C, Zhao J, Mu C, Wang Q, Wu H, Wang C: **cDNA cloning and mRNA expression of four glutathione S-transferase (GST) genes from *Mytilus galloprovincialis*.** *Fish Shellfish Immunol* 2013, **34**(2):697–703.
47. Yang HL, Zeng QY, Li EQ, Zhu SG, Zhou XW: **Molecular cloning, expression and characterization of glutathione S-transferase from *Mytilus edulis*.** *Comp Biochem Physiol B Biochem Mol Biol* 2004, **139**(2):175–182.
48. Boutet I, Tanguy A, Moraga D: **Characterisation and expression of four mRNA sequences encoding glutathione S-transferases pi, mu, omega and sigma classes in the Pacific oyster *Crassostrea gigas* exposed to hydrocarbons and pesticides.** *Mar Biol* 2004, **146**(1):53–64.
49. Lin Q, Liang XF, Hu YL, Wang L, Liu XX: **Cloning and Sequence Analysis of Four Classes of Glutathione S-Transferase Genes in Jinjiang Oyster (*Crassostrea ariakensis*).** *Asian J Ecotoxicol* 2009, **4**(2):237–243.
50. Xu C, Pan L, Liu N, Wang L, Miao J: **Cloning, characterization and tissue distribution of a pi-class glutathione S-transferase from clam (*Venerupis philippinarum*): response to benzo alpha pyrene exposure.** *Comp Biochem Physiol C Toxicol Pharmacol* 2010, **152**(2):160–166.
51. Zhang LB, Qiu LH, Wu HF, Liu XL, You LP, Pei D, Chen LL, Wang Q, Zhao JM: **Expression profiles of seven glutathione S-transferase (GST) genes from *Venerupis philippinarum* exposed to heavy metals and benzo a pyrene.** *Comp Biochem Physiol C Toxicol Pharmacol* 2012, **155**(3):517–527.
52. Umasuthan N, Revathy KS, Lee Y, Whang I, Choi CY, Lee J: **A novel molluscan sigma-like glutathione S-transferase from Manila clam, *Ruditapes philippinarum*: Cloning, characterization and transcriptional profiling.** *Comp Biochem Physiol C Toxicol Pharmacol* 2012, **155**(4):539–550.
53. Kim M, Ahn I-Y, Cheon J, Park H: **Molecular cloning and thermal stress-induced expression of a pi-class glutathione S-transferase (GST) in the Antarctic bivalve *Laternula elliptica*.** *Comp Biochem Physiol A Mol Integr Physiol* 2009, **152**(2):207–213.
54. Park H, Ahn I-Y, Kim H, Lee J, Shin SC: **Glutathione S-transferase as a biomarker in the Antarctic bivalve *Laternula elliptica* after exposure to the polychlorinated biphenyl mixture Aroclor 1254.** *Comp Biochem Physiol C Toxicol Pharmacol* 2009, **150**(4):528–536.
55. Feng X, Singh BR: **Molecular identification of glutathione S-transferase gene and cDNAs of two isotypes from northern quahog (*Mercenaria mercenaria*).** *Comp Biochem Physiol B Biochem Mol Biol* 2009, **154**(1):25–36.
56. Yang J, Wei X, Xu J, Yang D, Liu X, Yang J, Fang J, Hu X: **A sigma-class glutathione S-transferase from *Solen grandis* that responded to microorganism glycan and organic contaminants.** *Fish Shellfish Immunol* 2012, **32**(6):1198–1204.
57. Miao J, Pan L, Liu N, Xu C, Zhang L: **Molecular cloning of CYP4 and GSTpi homologues in the scallop *Chlamys farreri* and its expression in response to Benzo a pyrene exposure.** *Mar Genom* 2011, **4**(2):99–108.
58. Feldmeyer B, Wheat CW, Krezdorn N, Rotter B, Pfenninger M: **Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (*Radix balthica*, Basommatophora, Pulmonata), and a comparison of assembler performance.** *BMC Genomics* 2011, **12**:317.
59. Yi AX, Leung KMY, Lam MHW, Lee JS, Giesy JP: **Review of measured concentrations of triphenyltin compounds in marine ecosystems and meta-analysis of their risks to humans and the environment.** *Chemosphere* 2012, **89**(9):1015–1025.
60. Barsiene J, Syvokiene J, Bjornstad A: **Induction of micronuclei and other nuclear abnormalities in mussels exposed to bisphenol A, diallyl phthalate and tetrabromodiphenyl ether-47.** *Aquat Toxicol* 2006, **78**:S105–S108.
61. Wong SWY, Leung PTY, Djuricic AB, Leung KMY: **Toxicities of nano zinc oxide to five marine organisms: influences of aggregate size and ion solubility.** *Anal Bioanal Chem* 2010, **396**(2):609–618.



62. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotechnol* 2011, **29**(7):644–U130.
63. Li WZ, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.** *Bioinformatics* 2006, **22**(13):1658–1659.
64. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674–3676.
65. Zdobnov EM, Apweiler R: **InterProScan - an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**(9):847–848.
66. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics* 2011, **12**:323.
67. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method.** *Methods* 2001, **25**(4):402–408.
68. Sheir SK, Handy RD, Henry TB: **Effect of pollution history on immunological responses and organ histology in the marine mussel *Mytilus edulis* exposed to cadmium.** *Arch Environ Contam Toxicol* 2013, **64**(4):701–716.

doi:10.1186/1471-2164-15-804

**Cite this article as:** Leung et al.: *De novo* transcriptome analysis of *Perna viridis* highlights tissue-specific patterns for environmental studies. *BMC Genomics* 2014 **15**:804.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

