MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2012) doi: 10.1111/1755-0998.12052

# Genomic resource development for shellfish of conservation concern

EMMA B. TIMMINS-SCHIFFMAN, CAROLYN S. FRIEDMAN, DAVE C. METZGER, SAMUEL J. WHITE and STEVEN B. ROBERTS
*University of Washington, School of Aquatic and Fishery Sciences, Box 355020, Seattle, WA 98195, USA*

## Abstract

**Effective conservation of threatened species depends on the ability to assess organism physiology and population demography. To develop genomic resources to better understand the dynamics of two ecologically vulnerable species in the Pacific Northwest of the United States, larval transcriptomes were sequenced for the pinto abalone, *Haliotis kamtschatkana kamtschatkana*, and the Olympia oyster, *Ostrea lurida*. Based on comparative species analysis the *Ostrea lurida* transcriptome (41 136 contigs) is relatively complete. These transcriptomes represent the first significant contribution to genomic resources for both species. Genes are described based on biological function with particular attention to those associated with temperature change, oxidative stress and immune function. In addition, transcriptome-derived genetic markers are provided. Together, these resources provide valuable tools for future studies aimed at conservation of *Haliotis kamtschatkana kamtschatkana*, *Ostrea lurida* and related species.**

*Keywords*: gene expression, high-throughput sequencing, Olympia oyster, pinto abalone, transcriptome

*Received 5 October 2012; revision received 16 November 2012; accepted 20 November 2012*

## Introduction

Effective conservation efforts are dependent on understanding how organisms respond to environmental conditions (Wikelski & Cooke 2006). One means to assess physiological responses is to evaluate changes at the molecular level. Often, valuable insight can be gained from assessing gene expression variation that would not be readily evident from ecological observations. For instance, in *Crassostrea virginica* larvae, specific gene expression patterns as determined through qPCR were found to signal an early mortality event (Genard *et al.* 2012). Quantitative PCR has also been used to characterize the effects of ocean acidification on the purple sea urchin *Strongylocentrotus purpuratus* (Stumpp *et al.* 2011), Mediterranean sea urchin *Paracentrotus lividus* (Martin *et al.* 2011), red abalone *Haliotis rufescens* (Zippay & Hofmann 2010) and red sea urchin *Strongylocentrotus franciscanus* (O'Donnell *et al.* 2009). These studies revealed that ocean acidification alters metabolism and calcification (Martin *et al.* 2011; Stumpp *et al.* 2011) and limits an effective heat shock response (O'Donnell *et al.* 2009).

Transcriptomic data can be used to increase our understanding not only of an organismal response but can also aid in characterizing population structure and dynamics. Understanding population genetic structure can further the knowledge needed to increase the success of conservation efforts (e.g. restoration, relocation). This includes providing necessary information to maintain effective genetic diversity (Reed & Frankham 2003) and, in some cases, ensuring native cohorts are not diminished (Frankham 2002). High-throughput sequencing of transcriptomes can reveal a breadth of markers (e.g. single nucleotide polymorphisms or SNPs) that are informative not only for population demographics but also for understanding ecological and evolutionary mechanisms. For example, high-throughput sequencing of transcriptomes has been used to identify SNPs in the nonmodel Glanville fritillary butterfly (*Melitaea cinxia*; Vera *et al.* 2008), species in the fish genus *Xiphophorus* (Shen *et al.* 2012) and oilseed rape (*Brassica napus*; Trick *et al.* 2009). The SNPs discovered in *B. napus* were successfully tested for use in linkage mapping (Trick *et al.* 2009). Although not necessarily required *a priori*, sufficient genomic resources can greatly increase a researcher's ability to plan, implement and monitor success of conservation efforts.

The overall objective of this effort was to develop genomic resources for two shellfish species of conservation concern in the coastal region of northwest North America: the pinto abalone, *Haliotis kamtschatkana kamtschatkana* (hereafter referred to as *H. kamtschatkana*), and the Olympia oyster, *Ostrea lurida*. The pinto (or northern) abalone is an IUCN listed species of concern (NOAA

Correspondence: Steven B. Roberts, Fax: 206-685-7471;
E-mail: sr320@uw.edu

© 2012 Blackwell Publishing Ltd

(National Oceanographic & Atmospheric Administration) 2007) distributed from Alaska through Point Conception, California. Harvest of pinto abalone has been outlawed along its range since the 1990s, except in Alaska where recreational and subsistence harvests are still permitted. Within the past two decades, significant declines have been reported in pinto abalone populations in the San Juan Archipelago (Washington, USA), with evidence of decreased recruitment and potential for inhibited capacity to reproduce based on low population density (Rothaus *et al.* 2008; Bouma *et al.* 2012). Greater than expected mortality of mature *H. kamtschatkana* has been observed in the southeast Queen Charlotte Islands, BC, suggesting that poaching is still a valid concern for this threatened species (Hankewich *et al.* 2008). In the southern end of their range, *H. kamtschatkana* are threatened by warming ocean temperatures, poaching and predation by sea otters (Rogers-Bennett 2007). Marine reserves have been shown to promote both population growth and reproductive output in pinto abalone (Wallace 1999).

Olympia oysters are the only native oyster species on the United States west coast and are another species of conservation concern with a coast wide distribution. In the state of Washington, *O. lurida* are listed as a candidate Species of Concern. Overexploitation of *O. lurida* in the late 1800s through the early 1900s led to significant reductions of natural populations (White *et al.* 2009). Water pollution and decreased availability of suitable habitat compounded the effects of overharvest and led to further declines in *O. lurida* populations along the coast (Gillespie 2009; Groth & Rumrill 2009; White *et al.* 2009). Olympia oyster populations are still found along the majority of their historical distribution, indicating a possibility for rebuilding the species, however, densities are much lower than they once were (Polson & Zacherl 2009). *O. lurida* restoration efforts have been increasing in recent years (McGraw 2009) and include efforts such as habitat remediation and hatchery supplementation of natural populations (e.g. Dinnel *et al.* 2009).

This study describes the larval transcriptomes of the pinto (northern) abalone, *Haliotis kamtschatkana kamtschatkana* and the Olympia oyster, *Ostrea lurida*, with the objective of developing genomic resources for further conservation and management purposes. Along with descriptions of the full transcriptomes and identification of genes associated with environmental stress response, we also identify putative genetic markers for both species.

## Materials and methods

### *Sampling and library construction*

Olympia oyster brood stock from south Puget Sound were conditioned and stimulated to mass spawn at the Taylor Shellfish Hatchery in Quilcene, WA, USA. Larvae were transferred to the University of Washington 12 h post spawning. Larvae (12 larvae/mL) were evenly distributed to six 4.5-L larval chambers. Larvae were sampled from all chambers by filtering them onto a 35 $\mu$m screen and flash freezing in liquid $N_2$ on days one, two and three post-fertilization. Total RNA was extracted from pooled larvae using Tri Reagent and following the manufacturer's protocol (Molecular Research Center, Cincinnati, OH, USA). Two RNA-seq libraries were constructed from pooled mRNA (13 $\mu$g per sample) and sequenced at the University of Washington High Throughput Genomics Unit (HTGU) on the Illumina Hi-Seq 2000 platform (Illumina, San Diego, CA, USA). Both libraries were made of pooled mRNA from three chambers across all sampling time points. Each library was run on a single lane.

Pinto abalone larvae were spawned (1 × 1 cross) at the Mukilteo Research Station in Mukilteo, WA and transported to the University of Washington at three days post fertilization. Larvae (2 larvae/mL) were held in eight 4.0-L chambers and on day four post fertilization, all larvae were sampled for sequencing by filtering them onto a 70-$\mu$m screen and flash freezing them in liquid $N_2$. Similar to the Olympia oyster samples, two RNA-Seq libraries were constructed from equal quantities (10 $\mu$g) of pooled mRNA samples (from four chambers each). RNA was extracted from pooled larvae using Tri Reagent and following the manufacturer's protocol (Molecular Research Center). Pinto abalone libraries were run together on one lane of the Illumina Hi-Seq 2000 (Illumina) at the University of Washington HTGU.

### *Sequence analysis*

Sequence analysis was performed with CLC Genomics Workbench version 5.0 (CLC bio, Katrinebjerg, Denmark). Quality trimming was performed using the following parameters: quality score 0.05 (Phred; Ewing & Green 1998; Ewing *et al.* 1998), number of ambiguous nucleotides < 2 on ends and reads shorter than 25 bp were removed. *De novo* assembly was performed on the quality trimmed reads using the following parameters: mismatch cost = 2, deletion cost = 3, similarity fraction = 0.9, insertion cost = 3, length fraction = 0.8 and minimum contig size of 200 bp. Sequences were annotated by comparing contiguous sequences to the UniProtKB/Swiss-Prot database (http://uniprot.org) using the BLASTx algorithm (Altschul *et al.* 1997) with a 1.0E-5 e-value threshold. Genes were classified according to their biological processes as determined by their Gene Ontology (GO) and classified into their parent categories (GO Slim).

© 2012 Blackwell Publishing Ltd

To quantify the completeness of the transcriptomes, contigs from both species were assessed to determine if they contained orthologs to proteins found in all Metazoa. Specifically, OrthoDB (Waterhouse *et al.* 2011, http://cegg.unige.ch/orthodb6) was used to obtain a suite of proteins from *Lottia gigantea* (the giant owl limpet) found as single copy, which have orthologs in all other metazoans in OrthoDB. Sequence comparisons (tBLASTn; Altschul *et al.* 1997) were performed to find matching contigs in *H. kamtschatkana* and *O. lurida* transcriptomes. An e-value threshold of 1.0E-10 was used. To determine what proportion of full-length coding regions were obtained, sequences were translated and aligned to corresponding *L. gigantea* proteins (Geneious Pro version 5.3.6; Kearse *et al.* 2012). Alignment parameters were as follows: BLOSUM 62 cost matrix, gap open penalty of 12, gap extension penalty of 3, global alignment with free end gaps, and 2 refinement iterations. Percent coverage of the *L. gigantea* protein by the *O. lurida* or *H. kamtschatkana* contig was then calculated.

Using a more global approach to assess the larval transcriptomes of *O. lurida* and *H. kamtschatkana*, all contigs were compared using the BLASTn algorithm (Altschul *et al.* 1997) to publicly available transcriptomes of other shellfish, as well as to zebrafish. Specific data sets used include transcriptomes of Pacific oyster, *Crassostrea gigas* (GigasDatabase version 8; Fleury *et al.* 2009); pearl oyster, *Pinctada fucata* (Predicted Transcripts from Genome Assembly ver 1.0; http://marinegenomics.oist.jp/genomes/download?project_id=0; Takeuchi *et al.* 2012); South African abalone, *Haliotis midae* transcriptome (Franchini *et al.* 2011); Manila clam, *Ruditapes philippinarum* (RuphiBase http://compgen.bio.unipd.it/ruphibase/); and zebrafish, *Danio rerio* (NCBI RefSeq mRNA, ftp://ftp.ncbi.nih.gov/refseq/D_rerio/).

To investigate the functions of contigs that did not have significant matches in the transcriptomes of the species listed above, gene enrichment analysis was performed. Gene ontology information associated with the enriched unique gene sets was identified using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) v. 6.7 (Huang *et al.* 2009a,b, http://david.abcc.ncifcrf.gov/). The background gene list was made from the UniProtKB/Swiss-Prot database accession numbers that corresponded to the entire transcriptome of either *H. kamtschatkana* or *O. lurida*.

Phylogenetic analysis of select stress response genes was conducted with sequences from the following species: *O. lurida*, *H. kamtschatkana*, *H. midae*, *P. fucata*, *C. gigas*, *R. philippinarum* and *D. rerio*. Target genes and corresponding sequence accession numbers for each species are provided in Table 1. Sequences were aligned in Geneious Pro version 5.3.6 (Kearse *et al.* 2012) and trimmed to the length of the shortest sequence. Parameters for

**Table 1** Contig identification and accession numbers for the sequences used to build phylogenetic trees. Numbers are specific to the database from which the contigs were taken (see Materials & Methods section)

| | Cathepsin L | GABA Receptor | Heat Shock Cognate 70 | HSP 70 | HSP 82 | Transmembrane Protein 85 | v-Type Proton ATPase |
|---|---|---|---|---|---|---|---|
| *O. lurida* | 19 125 | 127 | 510 | 14 040 | 402 | 2793 | 1442 |
| *H. kamtschatkana* | 4888 | 430 | 111 | 8247 | 639 | 3651 | 8496 |
| *C. gigas* | FU6OSJA02F1X0P | BQ427163 | AF144646 | AF144646 | BQ426606 | FP001678 | CU995554 |
| *P. fucata* | aug1.0_4948 | aug1.0_4097 | aug1.0_16548 | aug1.0_6473 | aug1.0_178011 | aug1.0_175952 | aug1.0_3996 |
| *H. midae* | 3098 | 1516 | 199 | 4428 | 379 | 4120 | 881 |
| *R. philippinarum* | 33 755 | 7131 | 147 | 147 | 846 | 7419 | 865 |
| *D. rerio* | NM_001002368 | XM_001332873 | NM_001110403 | NM_001110403 | NM_131310 | NM_205658 | NM_205554 |

© 2012 Blackwell Publishing Ltd

alignment included cost matrix 65% similarity (5.0/−4.0), gap open penalty of 12, and gap extension penalty of 3. Phylogenetic trees were made using Geneious tree builder with the genetic distance model HKY, neighbour-joining tree build method, *Danio rerio* as the outgroup, 100 × bootstrap and 50% support threshold.

### *Single nucleotide polymorphism (SNP) identification*

Single Nucleotide Polymorphism (SNP) detection was carried out on *O. lurida* and *H. kamtschatkana* assemblies using the following parameters: window length = 11, maximum gap and mismatch count = 2, minimum central base quality = 20, minimum coverage = 10, maximum coverage = 1000 and minimum variant frequency = 35% (Genomics Workbench v. 5; CLC bio).

## Results

### *Larval transcriptomes*

Two larval transcriptome libraries were constructed and sequenced for Olympia oyster (*O. lurida*) and pinto abalone (*H. kamtschatkana*). Given that quality and yield from complementary libraries were consistent, all sequencing reads were combined within each species. Following quality trimming, 387 720 843 sequencing reads remained from the combined *O. lurida* larval libraries. For *H. kamtschatkana*, 94 777 799 quality reads remained. Raw sequence data are available in the NCBI Short Read Archive (GenBank SRA Accession Number SRA057107).

*De novo* assembly of sequencing reads from each species generated 41 136 and 8641 contigs for *O. lurida* and *H. kamtschatkana* respectively (Data S1. *Ostrea lurida* transcriptome and Data S2. *Haliotis kamtschatkana* transcriptome). The average contig lengths were 611 bp for *O. lurida* and 401 bp for *H. kamtschatkana* (Table 2).

**Table 2** Sequencing statistic summaries for high-throughput sequencing of *O. lurida* and *H. kamtschatkana* transcriptomes

| | *O. lurida* | *H. kamtschatkana* |
|---|---|---|
| Sequencing | 388 473 974 | 95 128 627 |
| Quality Trimmed Sequencing Reads | 387 720 843 | 94 777 799 |
| Mapped Reads | 222 773 113 | 71 892 404 |
| Number of Contigs | 41 136 | 8641 |
| Mean Contig Length | 611 | 401 |
| N50 Contig Count | 8236 | 2727 |
| Annotate Contigs (UniProt) | 15 381 | 1277 |
| Contigs with GO Definitions | 8030 | 574 |

### *Transcriptome annotation*

A total of 15 381 (37%) *O. lurida* contigs were annotated based on the UniProtKB/Swiss-Prot database (Data S3. *Ostrea lurida* SPIDs) with associated gene ontology information available for 8030 contigs (Data S4. *Ostrea lurida* GO and Fig. 1A). A total of 1277 (15%) *H. kamtschatkana* contigs were annotated based on the UniProtKB/Swiss-Prot database (Data S5. *Haliotis kamtschatkana* SPIDs) with associated gene ontology information available for 574 contigs (Date S6. *Haliotis kamtschatkana* GO and Fig. 1B).

Of the 30 orthologous proteins found across Metazoa, and found as a single copy in *L. gigantea*, 28 were identified with sequence similarity matches in *O. lurida*. A majority of the matching *O. lurida* contigs (24) had e-values < 1.0E-50. The average coverage of *O. lurida* contigs based on the corresponding *L. gigantea* protein was 53.7%. In contrast, only two proteins found across Metazoa were identified in the *H. kamtschatkana* transcriptome.

### *Comparative transcriptome analysis*

Of the 41 136 contigs from the *O. lurida* larval library, transcriptomes from the *C. gigas* and *P. fucata* oysters had the highest similarity to *O. lurida*, 21 748 and 11 101 matching contigs, respectively (Fig. 2A). The *D. rerio* transcriptome had the next highest number of matches with 6015, followed in order by *H. midae, R. philippinarum* and *H. kamtschatkana* (Fig. 2A). A summary of matches to *O. lurida* contigs in each data set is provided with respective bit scores for each pairwise blast comparison (Data S7. *Ostrea lurida* bitscores).

Of the 8641 contigs from the *H. kamtschatkana* library, the *H. midae* database had the most matches with 3090 BLAST hits (Fig. 2B). Comparison with the other transcriptomic data sets all produced fewer than 1000 matches (Fig. 2B). A summary of matches to *H. kamtschatkana* contigs in each data set is provided with respective bit scores for each pairwise blast comparison (Data S8. *Haliotis kamtschatkana* bitscores).

There were 1315 genes without matches to other transcriptomes in the *O. lurida* transcriptome (3.2% of all the contigs) and 243 with no matches to other transcriptome in the *H. kamtschatkana* transcriptome (2.8% of all contigs). The *O. lurida* unique contigs were enriched in the GO Slim categories of developmental processes, cell organization and biogenesis, cell adhesion, transport, stress response, signal transduction and protein metabolism (Table 3). The *H. kamtschatkana* unique contigs were enriched in the GO Slim terms of stress response and RNA metabolism (Table 3).

Analysis using select stress response genes resulted in phylogenetic tree clustering that was consistent with taxonomic relationships. Specifically, *H. kamtschatkana*

© 2012 Blackwell Publishing Ltd

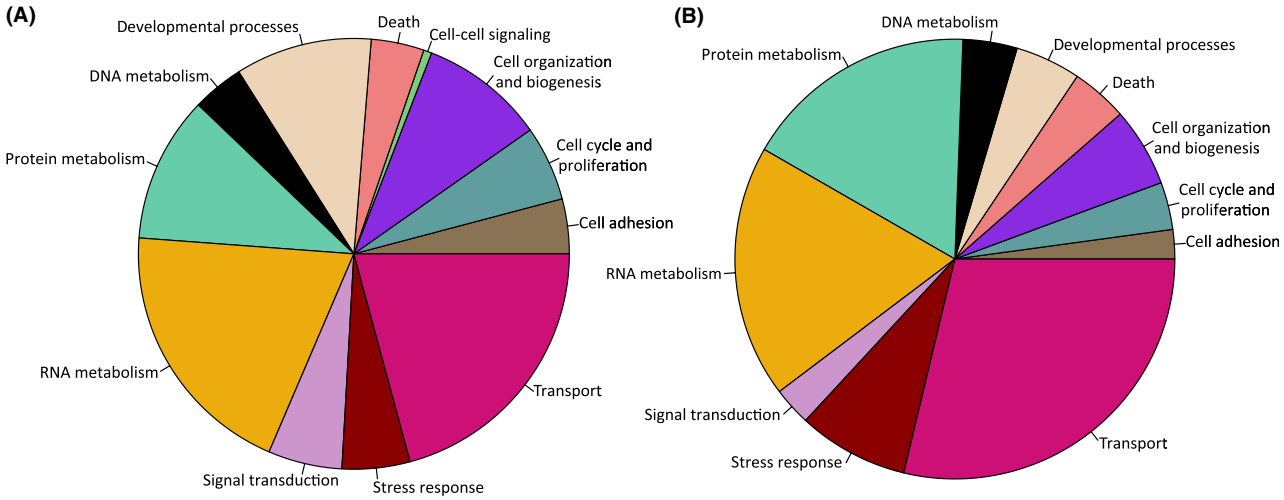


**Fig. 1** Functional classification of *O. lurida* contigs (panel A) and *H. kamtschatkana* contigs (panel B) based on GO Slim terms associated with Biological Processes. Other biological processes and other metabolic processes are not shown.

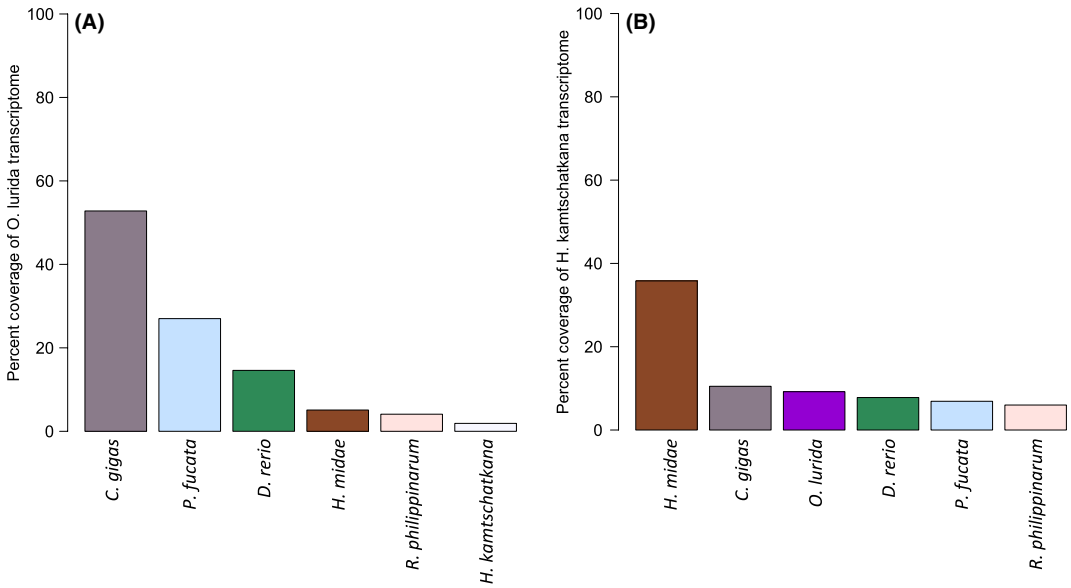


**Fig. 2** Percent coverage of *O. lurida* transcriptome (panel A) and *H. kamtschatkana* transcriptome (panel B) in comparison with the transcriptomes of closely related species. The *C. gigas* database has the most coverage of the *O. lurida* transcriptome (52.8%) and the *H. kamtschatkana* database covers the least (1.9%). The *H. midae* database has the most coverage of the *H. kamtschatkana* (35.8%) and the *R. philippinarum* database covers the least (6.0%). There are a total of 82 312 contigs in the *C. gigas* transcriptome database, 72 597 in the *P. fucata* database, 28 304 in the *D. rerio* database, 22 761 in the *H. midae* database, 32 606 in the *R. philippinarum* database, 8641 in the *H. kamtschatkana* database, and 41 136 in the *O. lurida* database.

and *H. midae* clustered together with high confidence (100%) except for the gene *hsp82* where *H. midae* and *R. philippinarum* clustered together with a bootstrap confidence of 50%, but within the same phyletic group as *H. kamtschatkana* (bootstrap confidence of 93%) (Fig. 3A). *O. lurida* consistently clustered with *C. gigas* with high bootstrap support (*v-type proton ATPase* 85%, *transmembrane protein* 85 99%, *hsp82* 66%, *heat shock 70 cognate* 67%, *GABA* 99%) (see Fig. 3A and C for *hsp82* and *transmembrane protein 85*, respectively). However, *O. lurida hsp70* clustered more closely with the corresponding gene in *P. fucata* (bootstrap confidence 68%) (Fig. 3B).

### *SNP Characterization*

In the *O. lurida* transcriptome, 59 221 putative SNPs were identified within 19 354 contigs (Data S9. *Ostrea lurida* SNPs). A subset of these putative SNPs (27 818) was

© 2012 Blackwell Publishing Ltd

**Table 3** Summary of unique genes and their functional categories for *O. lurida* and *H. kamtschatkana*. For each species, the total number of contigs corresponding to unique genes in enriched biological processes are given. Total number of contigs (for enriched and non-enriched biological processes) are in bold

| | Number of Contigs in Category | |
|---|---|---|
| | *O. lurida* | *H. kamtschatkana* |
| Number Unique Genes | 1315 | 243 |
| RNA metabolism | – | 5 |
| Stress response | 65 | 3 |
| Cell adhesion | 156 | – |
| Cell organization and biogenesis | 241 | – |
| Developmental processes | 328 | – |
| Protein metabolism | 268 | – |
| Signal transduction | 114 | – |
| Transport | 39 | – |

within annotated genes, corresponding to 6328 protein descriptions. In the *H. kamtschatkana* transcriptome, 6596 putative SNPs were identified within 3378 contigs (Data S10. *Haliotis kamtschatkana* SNPs). A subset of these putative SNPs (1038) was within annotated genes, corresponding to 442 protein descriptions. The functional distribution of SNPs among contigs (as determined by GO and GO Slim terms) was not significantly different from the annotation of the entire transcriptome for both *O. lurida* and *H. kamtschatkana*.

## Discussion

This research effort has increased the amount of genomic data available in *O. lurida and H. kamtschatkana*. At the time the project was undertaken there were only a few hundred publicly available sequences for the species combined. Here we provide 8641 *H. kamtschatkana* and 41 136 *O. lurida* contig sequences representing the larval transcriptomes. The *O. lurida* contigs represent a more complete transcriptome as compared to the *H. kamtschatkana* data set. This is not unexpected as approximately three times more reads were sequenced for *O. lurida* compared to *H. kamtschatkana*. Evidence to support the

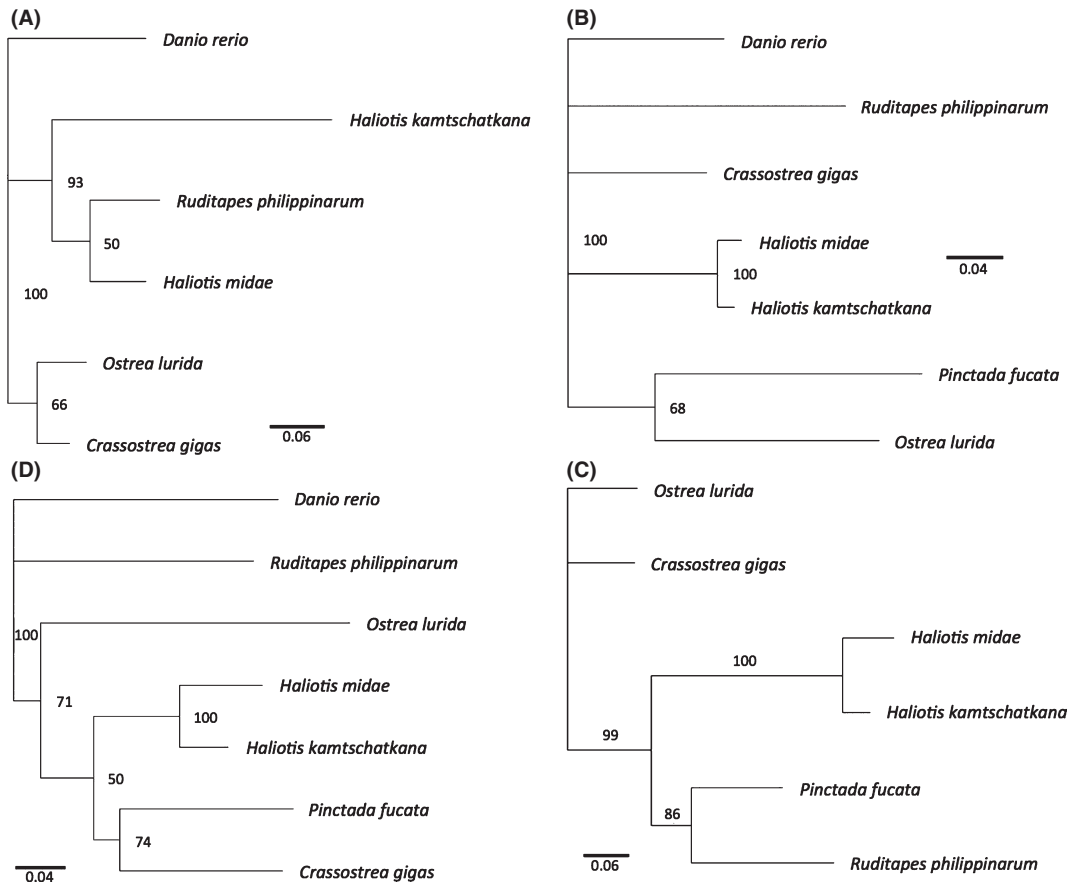


**Fig. 3** Phylogenetic trees of *hsp82* (panel A), *hsp70* (panel B), *transmembrane protein 85* (panel C), and *cathepsin-L* (panel D). Each tree includes sequences for all species (*O. lurida*, *H. kamtschatkana*, *C. gigas*, *H. midae*, *P. fucata*, *R. philippinarum*, and *D. rerio*) except when the sequence for a particular species did not fit in the alignment. Bootstrap consensus support is shown at the nodes.

© 2012 Blackwell Publishing Ltd

relative completeness of the *O. lurida* transcriptome includes the fact that the number of contigs from *de novo* assembly is similar to what has been reported in other species. Specific examples include *Ciona intestinalis* species B (27 426 contigs; Cahais *et al.* 2012), *Lepus granatensis* (38 540 contigs; Cahais *et al.* 2012) and *Radix balthica* (>20 000 contigs; Feldmeyer *et al.* 2011). Furthermore, using OrthoDB, we found evidence that over 90% of the proteins shared across Metazoa were present in the *O. lurida* transcriptome. Together, these data suggest that 14 gigabases of ultra short read (~36 bp) high-throughput transcriptome sequencing can produce a robust transcriptome. It would be expected that more sequences and longer read length would increase the number of full length sequences and facilitate identification of transcripts expressed at a minimal level. Although not complete in nature, both transcriptomes characterized here provide a wealth of information, evidenced in part by the number of sequences that did not have a significant match in cross-species comparisons.

In addition to characterizing the transcripts based on sequence homology, a suite of transcriptome derived SNPs are reported. Recently there have been several examples where transcriptome derived SNPs have provided unique insight into population dynamics. For example, high-throughput sequencing of the Pacific herring, *Clupea pallasii*, transcriptome yielded SNPs that were used to discriminate population structure between ocean basins (Roberts *et al.* 2012). Interestingly, one locus (*Cpa_APOB*) was a virus response gene and differentiation across populations at this locus suggested a disease pressure could contribute to the delayed recovery of Pacific herring (Roberts *et al.* 2012). In Pacific oysters, *Crassostrea gigas*, transcriptome derived SNPs have been used to create a linkage map and find quantitative trait loci linked to summer mortality and viral infection (Sauvage *et al.* 2010). Another example includes the application of SNPs derived from the Glanville fritillary butterfly, *Melitaea cinxia*, transcriptome (Vera *et al.* 2008) that were used in a separate large-scale study to characterize metapopulation structure (Wheat *et al.* 2011). It should be pointed out that while there are advantages to using transcriptome derived SNPs, there are also drawbacks. Foremost, these markers are characterized at the genomic level and the absence of introns in the transcriptome can contribute to marker 'drop-out' during validation (Seeb *et al.* 2011). Although it is beyond the scope of this effort to validate and use these markers to characterize population structure, the annotated SNPs provided for *H. kamtschatkana* and *O. lurida* will be a valuable resource in future efforts to examine restoration and conservation activity.

Through phylogenetic analysis and comparative species analysis, both *O. lurida* and *H. kamtschatkana* contigs have the overall greatest transcriptome similarity with closely related species, *C. gigas* and *H. midae* respectively. There was not a consistent or direct relationship between phylogenetic relatedness and the number of shared sequences. This is likely related to the incomplete nature of some of the transcriptome data sets used for comparison. For example, in both species the fully sequenced transcriptome of *Danio rerio* had a greater number of sequence matches than some invertebrates. Gene specific phylogenetic analysis generally demonstrated clustering along taxonomic lines, although there were some exceptions. Similarly, this is likely due to the incomplete nature of transcripts from some species, which limited the robustness of the comparison.

Even though most genes were shared across multiple, related species, both *O. lurida* and *H. kamtschatkana* had significant subsets of expressed genes that did not have matches in other species based on sequence similarity. For both species, these contigs were enriched for genes involved in the environmental stress response. The sequence dissimilarity across taxa is likely associated with species specific environmental interactions. In other words, it would be expected that the constitutively expressed, housekeeping genes would be conserved, whereas genes involved in transient processes (e.g. immune function) could diverge based on environmental pressure (e.g. disease, temperature). Another explanation for the relatively large number of unique sequences is that a majority of the other data sets were not specifically larval transcriptomes, but were derived from adult tissue. This is supported by the fact that the enriched biological process with the largest number of genes in *O. lurida* was developmental processes. It becomes possible that both species may possess some genes that have no homologues across taxa, or potential 'orphan' genes (Tautz & Domazet-Loso 2011). As additional sequence information is available for related species it is likely there will be greater attention directed towards these whole genome level evolutionary questions.

One important aspect of conservation is understanding how an organism responds to environmental change. As part of this sequencing effort we identified a suite of genes that are involved in the stress response that could be used in future studies to assess the physiological state and fitness of an organism. One group of genes that fits this description is molecular chaperones. *O. lurida* and *H. kamtschatkana* larvae express genes homologous to high molecular weight molecular chaperones, including *heat shock protein 70* (*hsp70*) (Ostrea_lur_contig14040, Haliotis_kam_contig8247), *hsp83* (Ostrea_lur_contig402, Haliotis_kam_contig7441), and *hsp90* (Ostrea_lur_contig7674, Haliotis_kam_contig1648). Chaperones in the HSP70 family are responsible for the proper folding of nascent polypeptides as well as providing a response to

© 2012 Blackwell Publishing Ltd

environmental stress in bivalves (reviewed in Hendrick & Hartl 1993; Kawabe & Yokoyama 2011; Chapman *et al.* 2011; Navarro *et al.* 2011; Genard *et al.* 2012). The HSP90 family, including HSP90 and HSP83, bind to various receptors and can be associated with signalling proteins (Hendrick & Hartl 1993; Parcellier *et al.* 2003). The expression of HSP90 can also be stress-induced and increased expression of its transcript is associated with stress-inducing dietary changes and other environmental changes in invertebrates (Parcellier *et al.* 2003; Fu *et al.* 2011; Wu *et al.* 2011; Zhang *et al.* 2011b). Global climate change is expected to significantly alter the temperature regimes and other environmental factors in the coastal ocean (IPCC (Intergovernmental Panel on Climate Change) 2007). Using general stress biomarkers such as heat shock proteins could be useful in monitoring resilience and adaptive potential in native invertebrate populations.

A second set of genes identified in *O. lurida* and *H. kamtschatkana* larval transcriptomes included genes associated with oxidative stress and antioxidant activity. Invertebrate exposure to pathogens and other common environmental stresses frequently results in increased expression of genes and proteins associated with oxidative stress. Oxidative stress is caused by proliferation of reactive oxygen species (ROS), which can promote apoptosis in cells (Wang *et al.* 1998) and lead to cellular aging and disease (Berlett & Stadtman 1997). A number of proteins scavenge ROS to minimize cellular damage, which makes them reliable biomarkers of a variety of stresses. Antioxidant genes identified in *O. lurida* and *H. kamtschatkana* included *peroxiredoxin-6* (Ostrea_lur_contig1185, Haliotis_kam_contig6404), *superoxide dismutase* (Ostrea_lur_contig1691, Haliotis_kam_contig8349), *catalase* (Ostrea_lur_contig24174), *cytochrome P450* (Ostrea_lur_contig24499, Haliotis_kam_contig5816), and *glutathione-s-transferase* (Ostrea_lur_contig27247, Haliotis_kam_contig5954). Peroxiredoxin-6 expression changes have been associated with the bivalve stress response (David *et al.* 2007; Genard *et al.* 2012). Superoxide dismutase and catalase have proven to play roles in *H. discus hannai* and *C. hongkongensis* immune responses (Wu *et al.* 2011; Zhang *et al.* 2011c). Cytochrome P450 and Glutathion-s-transferase are up-regulated in invertebrate responses to contaminants and metals, respectively (Livingstone *et al.* 1997; Chapman *et al.* 2011). All these antioxidant enzymes are reliable indicators of physiological stress caused by a harmful environment and can be applied as biomarkers to monitor populations of *O. lurida* and *H. kamtschatkana*.

Genes that code for proteins that accomplish a wide variety of roles in the innate immune response were also identified. These include *cathepsins* (Ostrea_lur_contig19125, Haliotis_kam_contig4888), *tumor necrosis factor* (TNF) *receptor* (Ostrea_lur_contig22192), *Toll-like receptor* (Ostrea_lur_contig811) and *mitogen activated protein kinase* (MAPK) (Ostrea_lur_contig10231). Some of these genes, like TNF receptor and MAPK, act upstream of pathways that affect the oxidative stress response (Wang *et al.* 1998). A subfamily of MAPKs, p38 enzymes have proven to be an important component of invertebrate response to a variety of environmental stressors, such as oxidative stress, hypoxia, osmotic stress and immune challenge (Canesi *et al.* 2002; Gaitanaki *et al.* 2004; Travers *et al.* 2009). TNF receptor is associated with several immune response pathways and plays a role in the invertebrate immune response (Betti *et al.* 2006; DeZoysa *et al.* 2009). Similarly, Toll-like receptors also play an integral role in signalling pathways during an immune response (reviewed in Barton & Medzhitov 2003). Homologues of toll-like receptors have been implicated in bacterial immune response in the Zhikong scallop, *Chlamys farreri* (Qiu *et al.* 2007) and in *C. gigas* (Zhang *et al.* 2011a). Cathepsins represent a wide class of enzymes, the roles of which include immune function, repair and apoptosis (reviewed in Conus & Simon 2010), roles that appear to be conserved in invertebrates (Griffitt *et al.* 2006; Meistertzheim *et al.* 2007). Although disease mortality is not considered a major culprit in the decreased population levels of *O. lurida* or *H. kamtschatkana,* these gene sequences will be important resources to examine immune function and resilience. In addition, these gene sequences could be used in isolating and characterizing homologues in closely related species. For example, the pathogen *Candidatus Xenohaliotis californiensis* has been identified as a major contributor to the decline of other *Haliotis* species such as the black abalone, *H. cracherodii* (Friedman & Finley 2003). Gene sequences from *H. kamtschatkana* could easily be used to study differential disease resistance in populations, which could directly aid in restoration activity.

In summary, we have characterized genomic resources for two species of conservation concern that, prior to this effort, had limited molecular information available. These data will provide essential information that could be used for future physiological (i.e. gene expression) and genetic based research carried out for planning, implementation and monitoring of conservation efforts.

## References

Altschul SF, Madden TL, Schäffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Barton GM, Medzhitov R (2003) Toll-like receptor signaling pathways. *Science*, **300**, 1524–1525.

Berlett BS, Stadtman ER (1997) Protein oxidation in aging, disease, and oxidative stress. *Journal of Biological Chemistry*, **272**, 20313–20316.

© 2012 Blackwell Publishing Ltd

Betti M, Ciacci C, Lorusso LC *et al*. (2006) Effects of tumor necrosis factor α (TNFα) on Mytilus haemocytes: role of stress-activated mitogen-activated protein kinases (MAPKs). *Biology of the Cell*, **98**, 233–244.

Bouma JV, Rothaus DP, Straus KM, Vadopalas B, Friedman CS (2012) Low juvenile Pinto Abalone *Haliotis kamtschatkana kamtschatkana* abundance in the San Juan Archipelago, Washington State. *Transactions of the American Fisheries Society*, **141**, 76–83.

Cahais V, Gayral P, Tsagkogeorga G *et al*. (2012) Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Molecular Ecology Resources*, **12**, 834–845.

Canesi L, Betti M, Ciacci C *et al*. (2002) Signaling pathways involved in the physiological response of mussel hemocytes to bacterial challenge: the role of stress-activated p38 MAP kinases. *Developmental & Comparative Immunology*, **26**, 325–334.

Chapman RW, Mancia A, Beal M *et al*. (2011) The transcriptomic responses of the eastern oyster, *Crassostrea virginica*, to environmental conditions. *Molecular Ecology*, **20**, 1431–1449.

Conus S, Simon H-U (2010) Cathepsins and their involvement in immune responses. *Swiss Medical Weekly*, **140**, w13042.

David E, Tanguy A, Moraga D (2007) Peroxiredoxin 6: a new physiological and genetic indicator of multiple environmental stress response in Pacific oyster *Crassostrea gigas*. *Aquatic Toxicology*, **84**, 389–398.

DeZoysa M, Jung S, Lee J (2009) First molluscan TNF- α homologue of the TNF superfamily in disk abalone: molecular characterization and expression analysis. *Fish & Shellfish Immunology*, **26**, 625–631.

Dinnel PA, Peabody B, Peter-Contesse T (2009) Rebuilding olympia oysters, Ostrea lurida Carpenter 1864, in Fidalgo Bay, Washington. *Journal of Shellfish Research*, **28**, 79–85.

Ewing B, Green P (1998) Base-calling of automated sequencer traces using *Phred* II. error probabilities. *Genome Research*, **8**, 186–194.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using *Phred* I. accuracy assessment. *Genome Research*, **8**, 175–185.

Feldmeyer B, Wheat CW, Krezdorn N, Rotter B, Pfenninger M (2011) 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*, **12**, 317–329.

Fleury E, Huvet A, Lelong C *et al*. (2009) Generation and analysis of 29,745 unique expressed sequence tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database, the gigasdatabase. *BMC Genomics*, **10**, 341–356.

Franchini P, van der Merwe M, Roodt-Wilding R (2011) Transcriptome characterization of the South African abalone *Haliotis midae* using sequencing-by-synthesis. *BMC Research Notes*, **4**, 59–70.

Frankham R (2002) Relationship of genetic variation to population size in wildlife. *Conservation Biology*, **10**, 1500–1508.

Friedman CS, Finley CA (2003) Anthropogenic introduction of the etiological agent of withering syndrome into northern California abalone populations via conservation efforts. *Canadian Journal of Fisheries and Aquatic Sciences*, **60**, 1424–1431.

Fu D, Chen J, Zhang Y, Yu Z (2011) Cloning and expression of a heat shock protein (HSP) 90 gene in the haemocytes of *Crassostrea hongkongensis* under osmotic stress and bacterial challenge. *Fish & Shellfish Immunology*, **31**, 118–125.

Gaitanaki C, Kefaloyianni E, Marmari A, Beis I (2004) Various stressors rapidly activate the p38-MAPK signaling pathway in *Mytilus galloprovincialis* (Lam.). *Molecular and Cellular Biochemistry*, **260**, 119–127.

Genard B, Moraga D, Pernet F, David E, Boudry P, Tremblay R (2012) Expression of candidate genes related to metabolism, immunity and cellular stress during massive mortality in the American oyster *Crassostrea virginica* larvae in relation to biochemical and physiological parameters. *Gene*, **499**, 70–75.

Gillespie GE (2009) Status of the olympia oyster, *Ostrea lurida* carpenter 1864, in British Columbia, Canada. *Journal of Shellfish Research*, **28**, 59–68.

Griffitt RJ, Chandler GT, Greig TW, Quattro JM (2006) Cathepsin B and glutathione peroxidase show differing transcriptional responses in the grass shrimp, *Palaemonetes pugio* following exposure to three xenobiotics. *Environmental Science and Technology*, **40**, 3640–3645.

Groth S, Rumrill S (2009) History of olympia oysters (*Ostrea lurida* Carpenter 1864) in oregon estuaries, and a description of recovering populations in Coos bay. *Journal of Shellfish Research*, **28**, 51–58.

Hankewich S, Lessard J, Grebeldinger E (2008) Resurvey of Northern Abalone, *haliotis kamtschatkana*, populations in Southeast Queen Charlotte Islands, British Columbia, May 2007. *Canadian Manuscript Report of Fisheries Aquatic Science*, **2839**, vii + 39 pp.

Hendrick JP, Hartl F (1993) Molecular chaperone functions of heat-shock proteins. *Annual Review of Biochemistry*, **62**, 349–384.

Huang DW, Sherman BT, Lempicki RA (2009a) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, **4**, 44–57.

Huang DW, Sherman BT, Lempicki RA (2009b) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, **37**, 1–13.

IPCC (Intergovernmental Panel on Climate Change) (2007) *Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Core Writing Team, Pachauri RK & Reisinger A (Eds) IPCC, Geneva, Switzerland, pp104.

Kawabe S, Yokoyama Y (2011) Novel isoforms of heat shock transcription factor 1 are induced by hypoxia in the Pacific oyster *Crassostrea gigas*. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, **315A**, 394–407.

Kearse M, Moir R, Wilson A *et al*. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Livingstone DR, Nasci C, Solé M *et al*. (1997) Apparent induction of a cytochrom P450 with immunochemical similarities to CYP1A in digestive gland of the common mussel (*Mytilus galloprovincialis* L.) with exposure to 2,2′,3,4,4′,5′-hexachlorobiphenyl and Arochlor 1254. *Aquatic Toxicology*, **38**, 205–224.

Martin S, Richier S, Pedrotti M *et al*. (2011) Early development and molecular plasticity in the Mediterranean sea urchin *Paracentrotus lividus* exposed to CO2-driven acidification. *Journal of Experimental Biology*, **214**, 1357–1368.

McGraw K (2009) The Olympia Oyster, *Ostrea lurida* Carpenter 1864 along the West Coast of North America. *Journal of Shellfish Research*, **28**, 5–10.

Meistertzheim A, Tanguy A, Moraga D, Thébault M (2007) Identification of differentially expressed genes of the Pacific oyster Crassostrea gigas exposed to prolonged thermal stress. *FEBS Journal*, **274**, 6392–6402.

Navarro A, Faria M, Barata C, Piña B (2011) Transcriptional response of stress genes to metal exposure in zebra mussel larvae and adults. *Environmental Pollution*, **159**, 100–107.

NOAA (National Oceanographic and Atmospheric Administration) (2007) *Species of Concern NOAA National Marine Fisheries Service, Pinto abalone Haliotis kamtschatkana*. NOAA, Silver Spring, MD.

O'Donnell MJ, Hammond LM, Hofmann GE (2009) Predicted impact of ocean acidification on a marine invertebrate: elevated $CO_2$ alters response to thermal stress in sea urchin larvae. *Marine Biology*, **156**, 439–446.

Parcellier A, Burbuxani S, Schmitt E, Solary E, Garrido C (2003) Heat shock proteins, cellular chaperones that modulate mitochondrial cell death pathways. *Biochemical and Biophysical Research Communications*, **304**, 505–512.

Polson MP, Zacherl DC (2009) Geographic distribution and intertidal population status for the olympia oyster, *Ostrea lurida* carpenter 1864, from Alaska to Baja. *Journal of Shellfish Research*, **28**, 68–77.

Qiu L, Song L, Xu W, Ni D, Yu Y (2007) Molecular cloning and expression of a toll receptor gene homologue from Zhikong Scallpo, *Chlamys farreri*. *Fish & Shellfish Immunology*, **22**, 451–466.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230–237.

Roberts SB, Hauser L, Seeb LW, Seeb JE (2012) Development of genomic resources for Pacific herring through targeted transcriptome pyrosequencing. *PLoS ONE*, **7**, e30908. doi:10.1371/journal.pone.0030908.

© 2012 Blackwell Publishing Ltd

Rogers-Bennett L (2007) Is climate change contributing to range reductions and localized extinctions in northern (*Haliotis kamtschatkana*) and flat (*Haliotis walallensis*) abalones? *Bulletin of Marine Science*, **81**, 283–296.

Rothaus DP, Vadopalas B, Friedman CS (2008) Precipitous declines in pinto abalone (*Haliotis kamtschatkana kamtschatkana*) abundance in the San Juan Archipelago, Washington, USA, despite statewide fishery closure. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 2703–2711.

Sauvage C, Boudry P, deKoning DJ, Haley CS, Heurtebise S, Lapègue S (2010) QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*). *Animal Genetics*, **41**, 390–399.

Seeb JE, Pascal CE, Grau ED *et al.* (2011) Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. *Molecular Ecology Resources*, **11**, 335–348.

Shen Y, Catchen J, Garcia T *et al.* (2012) Identification of transcriptome SNPs between *Xiphophorus* lines and species for assessing allele specific gene expression within F1 interspecies hybrids. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **1**, 102–108.

Stumpp M, Dupont S, Thorndyke MC, Melzner F (2011) CO2 induced seawater acidification impacts sea urchin larval development II: Gene expression patterns in pluteus larvae. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **160**, 320–330.

Takeuchi T, Kawashima T, Koyanagi R *et al.* (2012) Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. *DNA Research*, **19**, 117–130.

Tautz D, Domazet-Loso T (2011) The evolutionary origin of orphan genes. *Nature Reviews*, **12**, 692–702.

Travers M, leBouffant R, Friedman CS *et al.* (2009) Pathogenic *Vibrio harveyi*, in contrast to non-pathogenic strains, intervenes with the p38 MAPK pathway to avoid an abalone haemoctye immune response. *Journal of Cellular Biochemistry*, **106**, 152–160.

Trick M, Long Y, Meng J, Bancroft I (2009) Single nucleotide polymorphis (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnology Journal*, **7**, 334–346.

Vera JC, Wheat C, Fescemyer HW *et al.* (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology*, **17**, 1636–1647.

Wallace SS (1999) Evaluating the effects of three forms of marine reserve on Northern Abalone populations in British Columbia, Canada. *Conservation Biology*, **13**, 882–887.

Wang X, Martindale JL, Liu Y, Holbrook NJ (1998) The cellular response to oxidative stress: influences of mitogen-activated protein kinase signaling pathways on cell survival. *Biochemical Journal*, **15**, 291–300.

Waterhouse RM, Adobnov EM, Tegenfeldt F, Li J, Kriventseva EV (2011) OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. *Nucleic Acids Research*, **39**, 930, doi:10.1093/nar/gk

Wheat CW, Fescemyer HW, Kvist J *et al.* (2011) Functional genomics of life history variation in a butterfly metapopulation. *Molecular Ecology*, **20**, 1813–1828.

White J, Ruesink JL, Trimble AC (2009) The nearly forgotten oyster: *Ostrea lurida* carpenter 1864 (olympia oyster) history and management in Washington State. *Journal of Shellfish Research*, **28**, 43–49.

Wikelski M, Cooke SJ (2006) Conservation physiology. *Trends in Ecology & Evolution*, **21**, 28–46.

Wu C, Zhang W, Mai K, Xu W, Zhong X (2011) Effects of dietary zinc on gene expression of antioxidant enzymes and heat shock proteins in hepatopancreas of abalone *Haliotis discus hannai*. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **154**, 1–6.

Zhang L, Li L, Zhang G (2011a) A *Crassostrea gigas* toll-like receptor and comparative analysis of TLR pathway in invertebrates. *Fish & Shellfish Immunology*, **30**, 653–660.

Zhang W, Wu C, Mai K, Chen Q, Xu W (2011b) Molecular cloning, characterization and expression analysis of heat shock protein 90 from Pacific Abalone, *Haliotis discus hannai* Ino in response to dietary selenium. *Fish & Shellfish Immunology*, **30**, 280–286.

Zhang Y, Fu D, Yu F, Liu Q, Yu Z (2011c) Two catalase homologs are involved in host protection against bacterial infection and oxidative stress in *Crassostrea hongkongensis*. *Fish & Shellfish Immunology*, **31**, 894–903.

Zippay ML, Hofmann GE (2010) Effect of pH on gene expression and thermal tolerance of early life history stages of red Abalone (*Haliotis rufescens*). *Journal of Shellfish Research*, **29**, 429–439.

E.T.S. performed the analyses and contributed to writing of the manuscript. C.S.F. and S.B.R. were responsible for experimental design and provided funding for the research. S.B.R. helped to write the manuscript as well. D.C.M., and S.J.W. performed the lab work for the *O. lurida* and *H. kamtschatkana* parts of the experiment, respectively. All co-authors contributed comments and editing during the manuscript writing process.

## Data accessibility

mRNA Sequences: NCBI SRA SRA057107. Final RNA sequence assembly uploaded as online supplemental material. SNP data: uploaded as online supplemental material

## Acknowledgements

The authors would like to thank Taylor Shellfish Farms for supplying the Olympia oyster larvae and Puget Sound Restoration Fund, University of Washington, and Washington Department of Fish and Wildlife for supplying the pinto abalone larvae. We would also like to thank Liza Ray who helped to execute the pinto abalone part of this experiment. Lisa Crosson, Samantha Brombacker, and Elene Dorfmeier helped with larval care and sampling. We are grateful to Josh Bouma who provided comments during the writing process. This work was funded from Washington Sea Grant, University of Washington, pursuant to National Oceanographic and Atmospheric Administration Award No. NA10OAR4170057 Projects R/LME/N—3 and R/OCEH-4. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies. The U.S. government is authorized to reproduce and distribute this paper for governmental purposes.

## Supporting information

Additional supporting information may be found in the online version of this article:

**Data S1.** *Ostrea lurida* transcriptome. Assembled contigs of *O. lurida* transcriptome sequencing.

**Data S2.** *Haliotis kamtschatkana* transcriptome. Assembled contigs of *H. kamtschatkana* sequencing.

**Data S3.** *Ostrea lurida* SPIDs. BLASTx results for *O. lurida* contig search against the UniProtKB/Swiss-Prot database. BLAST e-values and gene descriptions are also given.

© 2012 Blackwell Publishing Ltd

**Data S4.** *Ostrea lurida* GO. Gene Ontology annotations of *O. lurida* contigs. GO annotations are made based on associations with a Swiss-Prot ID.

**Data S5.** *Haliotis kamtschatkana* SPIDs. BLASTx results for *H. kamtschatkana* contig search against the UniProtKB/Swiss-Prot database. BLAST e-values and gene descriptions are also given.

**Data S6.** *Haliotis kamtschatkana* GO. Gene Ontology annotations of *H. kamtschatkana* contigs. GO annotations are made based on associations with a Swiss-Prot ID.

**Data S7.** *Ostrea lurida* bitscores. Bit scores for BLASTn results of *O. lurida* contigs against species-specific databases of other closely related species.

**Data S8.** *Haliotis kamtschatkana* bitscores. Bit scores for BLASTn results of *H. kamtschatkana* contigs against species-specific databases of other closely related species.

**Data S9.** *Ostrea lurida* SNPs. SNP information for putative SNPs identified in the *O. lurida* transcriptome. Contig numbers are listed in the leftmost column, followed by SNP location and allele. Annotations of the contigs, as determined through a BLASTx against the UniProtKB/Swiss-Prot database, are given along with the e-value for the BLAST result.

**Data S10.** *Haliotis kamtschatkana* SNPs. SNP information for putative SNPs identified in the *H. kamtschatkana* transcriptome. Contig numbers are listed in the leftmost column, followed by SNP location and allele. Annotations of the contigs, as determined through a BLASTx against the UniProtKB/Swiss-Prot database, are given along with the e-value for the BLAST result.

© 2012 Blackwell Publishing Ltd