

Comparative transcriptome profiling of heat stress response of the mangrove crab *Scylla serrata* across different sites

Anish M.S. Shrestha (✉ anish.ms.shrestha@gmail.com)

De La Salle University <https://orcid.org/0000-0002-9192-9709>

Crissa Ann I. Lilagan

De la Salle University

Joyce Emlyn B. Guiao

De la Salle University

Maria Rowena R. Romana-Eguia

Southeast Asian Fisheries Development Center Aquaculture Department

Ma. Carmen Ablan Lagman

De la Salle University

Research article

Keywords: mud crab, mangrove crab, *Scylla*, RNA-seq, transcriptome assembly, heat stress

Posted Date: December 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-124588/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Genomics on July 29th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07891-w>.

RESEARCH

Comparative transcriptome profiling of heat stress response of the mangrove crab *Scylla serrata* across different sites

Anish M.S. Shrestha^{1,2*}, Crissa Ann I. Lilagan^{2,3,4}, Joyce Emlyn B. Guiao^{1,5}, Maria Rowena R. Romana-Eguia^{6,7} and Ma. Carmen Ablan Lagman^{3,7}

Abstract

Background: The fishery and aquaculture of the widely distributed mangrove crab *Scylla serrata* is a steadily growing, high-value, global industry. Climate change poses a risk to this industry as temperature elevations are expected to threaten the mangrove crab habitat and the supply of mangrove crab seeds from the wild. It is therefore important to understand the genomic and molecular basis of how mangrove crab populations from sites with different climate profiles respond to heat stress. Towards this, we performed RNA-seq on the gill tissue of *S. serrata* individuals sampled from 3 sites (Cagayan, Bicol, and Bataan) in the Philippines, under normal and heat-stressed conditions. To compare the transcriptome expression profiles, we designed a 2-factor generalized linear model containing interaction terms, which allowed us to simultaneously analyze within-site response to heat-stress and across-site differences in the response.

Results: We present the first ever transcriptome assembly of *S. serrata* obtained from a massive data set containing 66 Gbases of cleaned RNA-seq reads. With lowly-expressed and short contigs excluded, the assembly contains roughly 17,000 genes with an N50 length of 2,366 bp. Based on sequence comparison to the fruitfly and shrimp proteomes, our assembly contains several thousands of almost full-length transcripts. Differential expression analysis found population-specific differences in heat-stress response. Within-site analysis of heat response showed 177, 755, and 221 differentially expressed (DE) genes in the Cagayan, Bataan, and Bicol group, respectively. Across-site analysis of difference in heat response showed that between Cagayan and Bataan, there were 389 differently differentially expressed (DDE) genes associated with 48 signalling and stress-response pathways; and between Cagayan and Bicol, there were 101 DDE genes affecting 8 pathways.

Conclusion: In light of previous work on climate profiling and on population genetics of marine species in the Philippines, our findings suggest that the variation in thermal response among populations might be derived from acclimatory plasticity due to pre-exposure to extreme temperature variations or from population structure shaped by connectivity which leads to adaptive genetic differences among populations.

Keywords: mud crab; mangrove crab; *Scylla*; RNA-seq; transcriptome assembly; heat stress

Background

Scylla serrata is a commercially important, portunid mangrove crab. Of the four species that currently constitute the genus *Scylla* [1, 2], *S. serrata* is the most widely distributed – occurring throughout coastal tropical and subtropical Indo-Pacific [3, 4]. It is also the preferred species for farming due to its fast growth rate and large size [5]. The fishery and aqua-

culture of mangrove crabs is a steadily growing, high-value, global industry [4]. In the Philippines alone, its production increased from 17,000 metric tons in 2013 to almost 22,000 metric tons in 2018 [6, 7].

Fluctuations in climate, particularly that of elevation in temperature, pose a threat to the mangrove crab industry. Climate change is expected to threaten the environment in which the mangrove crabs thrive and the supply of mangrove crab seeds from the wild. Adverse effects of high temperature to the development and survival of ectothermic crustaceans are well documented [8, 9, 10]. While little can be done to control

*Correspondence: anish.shrestha@dlsu.edu.ph

¹Bioinformatics Lab, Advanced Research Institute for Informatics, Computing, and Networking (AdRIC), De La Salle University, Manila, Philippines

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

temperature fluctuations, one viable mitigation option is to use crabs that are better adapted to temperature changes. In this regard, it is important to understand the differences in the molecular mechanisms by which mangrove crab populations from sites with different climate profiles respond to heat stress. This information would be useful, for example, in developing a base population for breeding programs.

Past studies on stress response in mangrove crabs have been limited to using qPCR to evaluate the expression level variations of a limited number of genes, such as the heat-shock proteins [11]. Similar studies in other crustaceans such as amphipods, clams, and shrimps have also been reported [12, 13, 14]. However, the qPCR approach is unable to detect responses that are multi-genic in nature or that involve new genes not yet known to be associated with stress response. Furthermore, previous studies have not looked into biogeographic region as a factor in stress response. While it is known that there exists a high degree of intra- and inter-species genetic diversity among mangrove crabs from different regions [15, 16], the connection between population differentiation and resilience to temperature stress has not been adequately explored.

In this study, we used RNA-seq to compare the transcriptome expression profiles of the gill tissue of 29 *S. serrata* samples from 3 different populations in the Philippines, under normal and heat-stressed conditions. Cagayan, Bicol, and Bataan are located in North, South, and Central Luzon, respectively. Cagayan and Bicol lie on the eastern coast facing the Philippine Sea, while Bataan faces the South China Sea to the west of the Philippines (Figure 1a). They are major sites where crabs are caught, farmed, and fattened. The three sites fall in different climate zones and witness different weather profiles [17]. Of particular relevance to this study is sea-surface temperature. We show in Figure 1, the distribution of historical sea-surface temperatures observed shown. Cagayan experiences a more extreme weather profile, with both a wider range of annual sea surface temperature (Figure 1b) and more extreme sea surface temperature anomalies (Figure 1c). Of the three, Bicol witnesses the least amount of variation.

To compare the transcriptome expression profiles of the 6 groups (3 sites \times 2 conditions), we designed a 2-factor generalized linear model containing interaction terms, which allowed us to simultaneously analyze within-site response to heat-stress and across-site differences in the response. In the within-site analysis, we identified 177, 755, and 221 differentially expressed (DE) genes in the Cagayan, Bataan, and Bicol group, respectively. In the across-site analysis, we identified 389 differently differentially expressed

(DDE) genes associated with 48 signalling and stress-response pathways between Cagayan and Bataan, and 101 DDE genes affecting 8 pathway between Cagayan and Bicol. Through homology search and gene ontology analysis, we discovered that many of these genes are known to be associated with stress response pathways. Well-annotated reference transcriptome or genome resources are unavailable for mangrove crabs and scant even for crustaceans in general [18]. As part of the RNA-seq data analysis procedure, we computed the first ever transcriptome assembly of any tissue of *S. serrata* containing roughly 17,000 genes and N50 length of 2,366 bp. The assembly contains several thousands of transcripts that have long, high-confident alignments against sequences from the fruit fly proteome, Uniprot Swiss-prot, and the shrimp proteome.

Results

Study design and data

Our raw data consists of a total of 725 million 100 bp-long paired-end RNA-seq reads, obtained from the gill tissue of 29 adult *S. serrata* individuals. These individuals come from 6 groups – a combination of 3 sites (Cagayan, Bicol, and Bataan) and 2 treatments (Control and Heat-stressed). Individuals were sequenced at an average depth of 25 million reads. After removing low-quality read segments and potential rRNA-derived reads, there were a total of 663 million reads remaining.

Transcriptome assembly

From the cleaned reads, we obtained a de-novo transcriptome assembly of size 402 million bases containing 505,246 contigs grouped into 349,812 ‘genes’. The N50 statistic was 1,467 bp when accounting for all contigs, and 678 bp when selecting only one longest isoform per gene. Applying CD-HIT to cluster sequences with 95% similarity, reduced the number of transcripts to 435,969 and genes to 345,516.

An overwhelming majority of the estimated genes were lowly-expressed and have short transcripts. This can be seen by considering the subset S_x of genes containing only the top highly expressed ones which account for x percent of the total expression – where gene expression is the sum of TMM value [19] across all isoforms and all samples. For $x = 90$, S_x contains 17,162 out of the initial 349,812 genes. Further, as shown in Figure 2, the ExN50 statistic – which is the N50 value for S_x with the length of a gene defined as the average length of its isoforms weighed by their expression – peaks at N50 length of 2,366 bp around $x = 90$.

Assembly quality assessment

We evaluated the quality and contiguity of the assembly using several metrics.

Perc. of protein length aligned	No. of fruitfly proteins	No. of shrimp proteins
91-100	2167	4015
81-100	3049	5229
71-100	3602	6118
61-100	4099	6940
51-100	4536	7794

Table 1 Presence of almost full-length transcripts in the assembly. The second (third) column shows the number of fruitfly (shrimp) proteins that align to a contig in the assembly such that the alignment covers the percentage in the first column of the protein length.

First, to check if sequences from paired-end reads appear as consistent pairs in the assembly, we used Bowtie2 [20] to align the RNA-seq reads to the assembly. An average of 86% of the read-pairs, across all samples, aligned concordantly, indicating that the read pairs are well-represented in the assembly.

Next, to test the presence of contigs that are likely full-length transcripts, we used Blastx to align the contigs to two reference proteomes from Uniprot. We chose the proteomes of the fruit fly (*D. melanogaster*, proteome ID: UP000000803) since it is a model organism for arthropods, and that of the whiteleg shrimp (*P. vannamei*, proteome ID: UP000283509), since it has one of the largest set of protein sequences in Uniprot for crustaceans. There are 13,790 genes in the former, and 25,399 in the latter. Table 1 summarizes the alignment results. Almost 15% of both the fruit fly genes and shrimp genes are represented in almost full length (91-100%) in the assembly. Roughly 20% of both gene sets have > 80% of the protein sequence locally aligned to a contig in the assembly, and this number rise to almost 30% at the >50% length coverage threshold (last row of table).

Finally, the contigs were subjected to a BUSCO analysis with the arthropoda_odb10 database, which contains 1013 BUSCO groups. We found 94.4% complete BUSCOs in the assembly, with only 4.3% missing and 1.3% fragmented. Of the complete BUSCOs, 38.3% were reported as single-copy and 56.1% duplicated, which is due to the presence of many isoforms per gene in the assembly.

Functional annotation of de-novo assembly

We used the Dammit pipeline [21] which uses LAST [22] to perform a conditional reciprocal best-hit homology search [23] of our contigs against the Swiss-Prot database, the fruit fly proteome, and the proteome of the whiteleg shrimp. Of the alignments produced by Dammit, we considered as a hit only those that have an E-value at most 10^{-10} and those that cover at least 50% of the protein length. Figure 3 summarizes the results of the database hits. There

were 6,356 transcripts that had a hit in each of the three databases. At the gene-level, this number reduced to 4,232.

Based on the fruitfly annotations, gene ontology and pathway annotation was performed using Omics-Box and Panther [24], respectively. A total of 4,299 transcripts were assigned to one of the three GO domains: biological processes (BP), molecular function (MF) and cellular component (CC). The top gene ontology annotations (BP, MF and CC) are shown in Figure 4. In the BP category, there was a high number of genes associated with oxidation-reduction process and protein transport and ubiquitination. These genes play a significant role in redox homeostasis, protein synthesis as well as maintaining protein structure. In the MF category, most genes were predominantly assigned to ATP binding, protein binding, RNA binding, GTP binding and calcium ion binding. These molecular activities are integral in regulation of calcium ion concentration, protein synthesis, as well as energy generation processes. In CC category, genes associated with the integral component of membrane, cytoplasm, nucleus, cytosol were found to be abundant in the transcriptome assembly.

Pathway analysis results are shown in Figure 5. We found that the highest number of genes were associated with Parkinson and Huntington disease, which are model diseases for neurodegenerative disorders. Genes associated with vital signaling pathways such as Wnt signaling pathway, EGF receptor signaling pathway, Integrin signaling pathway, PDGF signaling pathway were found to be abundant in the transcriptome. We also found a high number of genes associated with ubiquitin proteasome pathway. These most-represented pathways show an overview of the cellular processes which are involved in proper biological functioning as well as the processes which may respond in reaction to heat stress.

Differential gene expression analysis

Site-specific heat-stress response (main effect)

We first consider the genes which were found to be differentially expressed between the control and heat-stressed groups of a particular site. In the statistical model, these are the genes for which one of the main effect null hypotheses in Table 2 was rejected. We will henceforth refer to these site-specific differentially expressed genes as DE genes.

There were 177, 755, and 221 DE genes in the Cagayan, Bataan, and Bicol group, respectively. As shown in Figure 6, the number of up-regulated DE genes were 60, 394, 135 and down-regulated DE genes were 117, 361, and 86, respectively. The scatter plots in

Figure 6 show for each site, the average mean expression (as mean normalized read counts) versus the magnitude of differential expression (as \log_2 fold change). The DE genes are represented as non-grey dots. We call a gene to be *stress-response related* if its reciprocal-best-hit fruit fly gene is associated with a GO term that is a descendant of the GO term “response to stress” in the Gene Ontology graph. These genes can be loosely understood to be genes whose putative homologs in fruitfly have been known to be related to stress response. There were 24, 117, 39 stress-related DE genes in the Cagayan, Bataan, and Bicol group respectively, which are shown as red dots in the Figure 6.

The list of top 20 DE genes (by adjusted p-value) for each site and their expression profile is shown in the Supplementary Information.

Across-site difference in heat-stress response (interaction effect)

We next consider the genes for which there was an effect of site in how the expression levels varied between control and heat-stressed conditions. In other words, these are genes whose differential expression was different – either in magnitude (\log_2 fold-change) or the direction of regulation – between a site pair. In the statistical model, these are the genes for which an interaction effect null hypothesis in Table 2 was rejected. We will refer to them henceforth as differently differentially expressed (DDE) genes.

In the Cagayan-Bataan comparison, there were 389 DDE genes; and in the Cagayan-Bicol comparison, there were 101 DDE genes. The heatmaps of Figure 7 show the expression profiles of the DDE genes. Each row of the heatmap corresponds to a DDE gene, the colors represent deviation from row average of the variance-stabilized read counts – red for lower than average and green for higher than row average. The intensity of color represents the z-score. The horizontal black bars to the right indicate genes that are stress-response related. The entire list of DDE genes along with the interaction plots of the top 10 (by value of estimated interaction coefficient) are provided in the Supplementary Material.

GO analysis of DDE genes were performed using Blast2GO. Overall, most DDE genes were assigned to GO terms “cellular process”, “metabolic process” and “localization” in the biological process category. Next we restricted focus on the stress-related DDE genes. In the Cagayan-Bataan comparison, we observed that DDE genes associated with GO terms “activation of innate immune response”, “response to oxidative stress”, “long-chain fatty acid metabolic transprot” were upregulated in Cagayan and downregulated in Bataan;

whereas DDE genes responsible for “regulation of reactive oxygen species” and “cellular response to hypoxia” were downregulated in Cagayan and upregulated in Bicol. In the Cagayan-Bicol comparison, DDE genes associated with “regulation of reactive oxygen species” and “cellular response to hypoxia” were down-regulated in Cagayan and upregulated in Bicol; whereas DDE genes associated with “response to oxidative stress”, “activation of innate immune response” and “synaptic target recognition” were upregulated in Cagayan and downregulated in Bicol.

Using Panther, the DDE genes were assigned to 48 and 8 reactome pathways for the Cagayan-Bataan and Cagayan-Bicol comparison, respectively. A majority of DDE genes were mapped to pathways related to stress signal transduction, metabolic processes, immune response, response to oxidative stress and cell death. When restricting focus on the stress-related DDE genes, the enriched pathways were apoptotic pathways, oxidative stress-related pathways, and signal transduction pathways such that the Integrin signaling pathway, EGF receptor signaling pathway, Interferon-gamma signaling pathway, TGF-beta signaling pathway, Wnt signaling, and p53 pathway.

The list of top 20 DDE genes (by size of coefficient) for the two site-pairs and the corresponding interaction plots of gene expression is shown in the Supplementary Information.

Discussion

*The first *S. serrata* transcriptome assembly*

We present the first transcriptome assembly of any tissue of *S. serrata*. Several quality assessment metric attest to the reliability of the assembly. The source reads are well represented in the assembly, with an average of 86% of the read pairs aligning concordantly to the assembly. Table 1 shows that more than 4000 protein sequences of the proteome of the closely related whiteleg shrimp *P. vannamei*, align almost fully to distinct contigs in the assembly. The soundness of the assembly is further supported by BUSCO analysis, which found almost all of the 1013 arthropod BUSCOs in the assembly.

Pathway analysis results showed that our assembly is enriched in genes associated with several vital signalling pathways such as the Wnt signaling pathway, EGF receptor signaling pathway, Integrin signaling pathway, and PDGF signaling. These signalling pathways are mostly evolutionarily conserved and are responsible for proper biological functioning and development. Interestingly, these pathways are also known to function in response to heat-stress [25, 26, 27, 28]. High number of genes associated with ubiquitin proteasome pathway may indicate activated mechanism

for protein degradation, whereas apoptosis signaling pathway may indicate potential severe protein damage which leads to programmed cell death. Surprisingly, we also found that a high number of genes were associated with Parkinson and Huntington disease, which are model diseases for neurodegenerative disorders. This may be indicative of potential neural function damage in which several cellular processes are involved.

A few transcriptome assemblies of other *Scylla spp.* have been previously reported. These include the assemblies of the testis of *S. olivacea* [29], and testis, ovary, and infected hemocytes of *S. paramamosain*, and megalopa stage of *S. paramamosain* [30, 31, 32]. Compared to previous assemblies, our assembly was computed using the largest amount of input data so far – roughly 66 Gbases collected from 29 individuals. Functional analysis of the assembly revealed enrichment in pathways that have not been reported in any of the previous mangrove crab assemblies. This hints at a vast transcriptomic complexity in mangrove crabs; and our *S. serrata* transcriptome assembly adds to the scant but growing genomic database resources of mangrove crabs and crustaceans, in general, allowing for future investigations to shed light into the functional genomics of *Scylla spp.*

Within-site main-effect of heat stress suggests differences in response between East and West Philippine mangrove crab populations

The number of DE genes can be indicative of the degree of modifications in cellular processes that the organisms has to make in response to heat stress. In this regard, Figure 6 shows the stark differences in how the three sites responded differently to heat stress. The remarkably high number of DE genes in the Bataan population, may be a reflection of the drastic adjustments in metabolic processes to cope with the stress or it could be a manifestation of the damaging effects of heat stress.

In contrast, the Cagayan group had a 4.5-fold lower number of DE genes and the magnitude of expression level differences, measured as \log_2 fold change, were also overall lower compared to Bataan. As illustrated in Figure 1, Cagayan experiences a highly variable temperature profile. It is possible that the constant exposure to highly variable temperature prompted an increase in heat tolerance in this group which resulted in a lesser number of DE genes observed. It has been hypothesized that response to acute stress may be dependent on the magnitude of stress organisms experience throughout their lifetime, with organisms that have existed in an environment with high temperature variations having broad thermal tolerance and high capacity for acclimation [33, 34] In addition to this, the

fewer number of DE genes coupled with higher percentage of downregulated genes observed in Cagayan population is analogous to the observations in a study done by Poong et al. [35], wherein it is considered a strategy for ‘energy and resources saving’ of heat-acclimated cells. Owing to the broad heat tolerance of the Cagayan population due to natural exposure to highly variable temperature, transcription and translation of non-essential proteins has become unnecessary which consequently saves cell energy and resources.

Interestingly, the Bicol group, whose population does not seem to have natural exposure to extreme temperature, demonstrated a similar response as that of Cagayan, in terms of the number of DE genes. We postulate that this similarity may be a result of population connectivity due to the established East-West separation in the Philippines. Studies on population structure of marine fishes and invertebrates in the Philippines have revealed a strong East-West separation – for example in rabbitfish [36], damselfish [37], and seahorse [38]. A study on population structure of wild *S. serrata* in the Philippines using microsatellite and mitochondrial markers also shows distinct East-West groupings: east coast populations of Cagayan and Quezon, and west coast populations of Pangasinan and Bataan [39]. For the highly dispersive *S. serrata*, the limitations on gene flow between the east and west coasts due to sea-surface currents and geographic barriers may play a significant role in shaping the adaptive genetic differences among populations in response to thermal stress.

Analysis of interaction effect shows differences in biological pathways and genes responsible for difference in heat stress response

While the site-specific main effect analysis provides a global quantitative view of the differences in heat response, the DDE genes provide insights into the cellular processes that differentiate the heat-stress response across sites. We found that there were 48 pathways that differentially activated between Cagayan and Bataan, and 8 pathways between Cagayan and Bicol. Of particular interest are the oxidative stress and cell apoptosis pathways.

Under normal temperature conditions, the mild production of reactive oxygen species (ROS) controlled by the antioxidant defense mechanisms is helpful for proper cell function [40]. However, various stress conditions such as elevated temperature could trigger the overproduction of ROS which in turn results in cell oxidative stress [41]. Oxidative stress pathway was observed in the Cagayan-Bataan comparison with genes associated in this pathway being upregulated in Bataan but downregulated in Cagayan. In a study on

human umbilical vein endothelial cells, it was shown that heat stress increases the ROS production and induces mitochondrial p53 translocation and Ca²⁺ accumulation [42]. In our study, the gene associated with the p53 pathway was upregulated in Bataan and downregulated in Cagayan, and genes responding to calcium ions were expressed at various levels in 3 sites. Moreover, several studies have linked ROS-induced oxidative stress to cell death in heat-exposed cells. Different pathways are responsible for cell apoptosis which include extrinsic apoptotic pathway via death receptors and intrinsic apoptotic pathway involving the mitochondria and endoplasmic reticulum [42]. Cell apoptosis pathways were observed to differently activated in both Cagayan-Bataan and Cagayan-Bicol comparisons. Genes associated in this pathway were downregulated in Cagayan and upregulated in Bataan and Cagayan. Downregulation of genes associated with oxidative stress and cell apoptosis in the Cagayan population may be attributed to the thermotolerance of the population which indicates that such changes are part of the acclimation process rather than an onset of cell death. The downregulation of these genes may be a strategy to save energy and resources assuming that the population is already acclimated to elevated temperature.

Downregulation of genes associated with oxidative stress and cell apoptosis in the Cagayan population may be attributed to the thermotolerance of the population which indicates that such changes are part of the acclimation process rather than an onset of cell death. The downregulation of these genes may be a strategy to save energy and resources assuming that the population is already acclimated to elevated temperature.

Implications for fisheries and aquaculture management of *S. serrata* populations

Aquaculture production of *S. serrata* in the Philippines reached up to 22,000 metric tons valued at Php 9 billion in 2018. As *S. serrata* is a preferred species for trading and farming, this species is subject to over-exploitation for commercial purposes. Moreover, frequent exposure to environmental stressors could lead to population decline for this economically valuable crustacean. A study on domestication of mangrove crabs in the Philippines by Quinitio et al. [43] highlights the need for management of *S. serrata* populations due to a documented decrease in size, yield and catch per unit effort (CPUE) in key collection areas which are attributed to overexploitation and environmental factors.

Most mangrove crab farms in the Philippines practice a long-term grow-out approach [44] which is highly

dependent on crab seeds collected from the wild. Current hatchery and nursery production of *S. serrata* juveniles are not enough to support the needs of the mangrove crab industry in view of the species' vulnerability to variations in environmental conditions (i.e. salinity, temperature, dissolved oxygen) especially for the early larval stages. With this, knowledge on the impact of environmental conditions particularly temperature variation to the populations of *S. serrata* is critical to support their population and maintain sustainable mangrove crab aquaculture.

The findings of this study show that population-specific response of *S. serrata* to temperature variation is due to the physiological adaptation to their respective thermal optima which is derived from the local environment conditions. While our results indicate that *S. serrata* populations are capable of acclimatory plasticity, it may be metabolically expensive for the organisms to survive in a new environment with a new thermal regime. Hence in screening for potential sources of mangrove crabs for broodstock development either for aquaculture or for supportive breeding in stock management, it has been shown that one can explore the use of stocks that have broad heat tolerance and high capacity for acclimation, which may be hallmarks of success in adaptation to changing thermal environments [45].

Conclusion

In this study, we attempted to understand the genomic and molecular mechanisms that differentiate the heat-stress response of mangrove crab populations from sites with varying climate profiles. As part of the bioinformatic analysis, we constructed a first de-novo transcriptome assembly of *S. serrata*. Several quality assessment metric point to a sound and informative assembly, which adds to the growing genomic database resources of mangrove crabs. Results from the differential expression analysis showed genes and pathways responsible for population-specific response to heat-stress in the 3 populations. Based on previous work population structure of marine species, our findings suggest that population-specific heat-stress response might be attributed to acclimatory plasticity due to pre-exposure to extreme temperature variations or might be due to the physiological adaptation to their respective thermal optima which is derived from the local environment conditions. Our results may serve as a basis for hypothesis testing regarding the impact of population connectivity to the adaptive genetic differences among populations. Future work may focus on genetic analysis to confirm genetic signatures for thermal adaptation among populations.

Methods

Sample collection and thermal stress exposure

Adult *S. serrata* (90–120 mm carapace width) were collected from three sites of varying temperature profiles: Bataan, Cagayan and Bicol, Philippines. All crab samples were color-tagged according to collection sites and were subjected to formalin bath (100 ppm formalin solution) in one large tank for 1 hour. A total of six groups (3 sites × 2 conditions) were considered in the experiment, with 9 individuals per group. Prior to thermal stress exposure, all crab samples were acclimatized at an ambient temperature of $26 \pm 2^\circ\text{C}$ for 48 hours. The control condition was set at $26 \pm 2^\circ\text{C}$ for the whole duration of the study, whereas the heat stress condition was increased at a rate of $2^\circ\text{C}/24\text{ hr}$ until the maximum experimental temperature of 32°C was reached and kept for 72 hours. All crab samples were fed with the same diet and exposed to a salinity of $30 \pm 2\text{ ppt}$. After the heat-stress experiment, gill tissue samples were harvested, placed in RNALater (Ambion) and stored at -80°C until ready for processing.

RNA isolation, library construction, and RNA Sequencing

Tissue samples from each site and experimental condition were randomly selected and sent to an external service provider (Macrogen, Inc., South Korea) for RNA extraction, library construction, and sequencing. Total RNA was extracted from gill tissue of 30 crab samples (5 individuals from each group) following the TRIzol (Invitrogen) protocol. Quality of the extracted RNA was measured through RNA Integrity Number using 2200 TapeStation (Agilent). A total of 29 cDNA libraries (omitting one low-quality sample from Bataan-Experimental group) were constructed from poly(A)-captured mRNAs using TruSeq RNA Sample Prep Kit (Illumina, Inc.). A non-stranded RNA sequencing was performed for each library using the Illumina Hi-Seq 2000 in paired-end mode and 101 bp length.

De-novo transcriptome assembly

SortmeRNA [46] (version 2.1b) was used to remove potential rRNA-derived reads by aligning the reads against the following databases: silva-bac-16s-id90, silva-bac-23s-id98, silva-arc-16s-id95, silva-arc-23s-id98, silva-euk-18s-id95, silva-euk-28s-id98, silva-euk-28s, rfam-5s-database-id98, rfam-5.8s-database-id98.

Trimmomatic [47] (version 0.39) was used to remove adapters and filter or trim low-quality using the following parameters: LEADING:3 TRAILING:3 AVGQUAL:20 MINLEN:35. Only those read pairs for which both members of the pair survive the filtering were kept for further analysis.

Trinity [48] (version 2.8.5), in its default settings, was applied to the entire pool of filtered reads from all samples to obtain a transcriptome assembly.

To reduce computation time of the downstream annotation pipeline, we applied CD-HIT to cluster contigs with a sequence similarity of at least 95%.

Transcript abundance estimation

The *align_and_estimate_abundance.pl* script provided by Trinity were used to invoke the combination of RSEM [49] (version 1.3.3) and Bowtie2 [20] (version 2.3.5) to compute transcript abundance for each sample. The *abundance_estimates_to_matrix.pl* script was executed to integrate raw counts at the gene-level across samples, and also to compute cross-sample normalized TMM counts. The former was used for differential gene expression analysis, while the latter for computing ExN50 statistics and for exploratory analysis prior to differential gene expression analysis.

Annotation

The Dammit annotation pipeline was used to perform conditional reciprocal best-hit search using LAST in OrthoDB, Swiss-Prot, and the Uniprot reference proteomes of the fruitfly *D. melanogaster* and the whiteleg shrimp *P. vannamei*.

Differential gene expression analysis

Differential gene expression analysis was performed on the subset of genes which had a reciprocal-best-hit alignment to a protein in the fruitfly proteome, such that the alignment had an E-value of at most 10^{-10} and it covered at least half of the protein length. There were 6,898 such genes in the assembly. While this is a conservative choice, it allows for a downstream functional analysis using the rich pathway and GO database of the fruit fly leading to more meaningful functional interpretation of the differential expressed genes.

We used DeSeq2 [50] (version 1.28.1) to perform differential gene expression analysis on the set containing Trinity-genes which found a hit in the *D. melanogaster* proteome with an E-value of at most 10^{-10} and the alignment covering at least 50 percent of the protein length. Transcript-level count data produced by RSEM were aggregated at the gene-level using tximport [51] (version v1.16.1). Exploratory data analysis and quality control procedures taken prior to differential gene expression analysis are described in the Supplementary Information.

For each gene, we set the DeSeq2 negative binomial regression model to a 2-factor model with interaction terms: \sim site + treatment + site : treatment, where site has 3 levels (CAG (Cagayan), BAT (Bataan), BIC

Effect type	Statistical null hypothesis	Biological null hypothesis
Main	$\beta_E = 0$	Not differentially expressed between control and heat-stressed CAG groups.
Main	$\beta_E + \beta_{BAT.E} = 0$	Not differentially expressed between control and heat-stressed BAT groups.
Main	$\beta_E + \beta_{BIC.E} = 0$	Not differentially expressed between control and heat-stressed BIC groups.
Interaction	$\beta_{BAT.E} = 0$	Effect of heat stress is not different between CAG and BAT.
Interaction	$\beta_{BIC.E} = 0$	Effect of heat stress is not different between CAG and BIC.

Table 2 For each gene, based on the model in Equation 1, we test the statistical hypotheses shown in the second column, which corresponds to the biological hypotheses in the third column.

(Bicol)) and treatment has 2 levels (Control (C) and Heat-stressed (E)). For the treatment factor, we chose Control as the reference level. For the site factor, we chose Cagayan as the reference level, since given its temperature extremes it faces, our prior expectation was for the Cagayan group to be the most thermally resilient. Our results suggest that this indeed seems to be the case.

With CAG and C set as reference levels for site and treatment, respectively, the regression model can be expressed as:

$$\log(q_i) = \beta_0 + \beta_{BAT}X_{BAT} + \beta_{BIC}X_{BIC} + \beta_E X_E \quad (1)$$

$$+ \beta_{BAT.E}X_{BAT.E} + \beta_{BIC.E}X_{BIC.E},$$

where q_i is the mean of negative binomial distribution for sample i normalized for sequencing depth, X s are dummy variables that take values of 0 or 1 depending on the group membership of sample i , and β s are the coefficients of the model. The hypotheses tests performed on the model coefficients are shown in Table 2

Dispersion estimates were shrunk towards the dispersion of similarly expression genes using DESeq2's default parametric method for shrinking the dispersion. Likewise, for visualization and ranking purposes, estimated log fold change (LFC) was shrunk towards the more likely LFC estimate by pooling information from all genes [52]. To account for multiple testing due to the large number of genes as well as multiple comparisons per gene, we applied the Benjamini-Hochberg procedure [53] with the false discovery rate threshold at 0.1.

Abbreviations

- DE: Differentially expressed
- DDE: Differently differentially expressed
- TMM: Trimmed mean of M values
- GO: Gene ontology
- BUSCO: Benchmarking Universal Single-Copy Orthologs

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated during the current study are available in the DDBJ Sequence Read Archive under the accession number DRA010977. The computational pipeline used for data analysis is available at the following URL: https://bitbucket.org/s_serrata/dge/src/master/

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the Department of Science and Technology Philippine Council for Agriculture and Aquatic Resources Research and Development (DOST-PCARRD) Mangrove Crab Program to De La Salle University-Manila through MCAL. AMSS was partially funded by University Research Coordination Office, De La Salle University-Manila. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

MCAL and MRRR conceived, planned, and supervised the experiments. CAIL, MCAL, and MRRR performed the experiments. AMSS, JEBG, and CAIL carried out the bioinformatic analysis. AMSS and CAIL wrote the manuscripts with inputs from all authors. All authors read and approved the final manuscript.

Acknowledgements

The authors are thankful to Maria Victoria Faith Garcia, Biena Marie Joaquin, and Karen Perez for their work in experiment set-up and data gathering activities, and to Dan Joseph Logronio and Joseph De La Cruz for running the field experiments. AMSS is thankful to Hugues Richard and Avi Belbase for helpful discussions.

Author details

- ¹Bioinformatics Lab, Advanced Research Institute for Informatics, Computing, and Networking (AdRIC), De La Salle University, Manila, Philippines. ²Software Technology Department, College of Computer Studies, De La Salle University, Manila, Philippines. ³Practical Genomics Laboratory, Center for Natural Science and Environment Research, De La Salle University, Manila, Philippines. ⁴Department of Biological Sciences, College of Science University of Santo Tomas Manila, Philippines. ⁵Mathematics and Statistics Department, College of Science, De La Salle University, Manila, Philippines. ⁶Aquaculture Department, Southeast Asian Fisheries Development Center Binangaoan, 1940 Rizal, Philippines. ⁷Biology Department, College of Science, De La Salle University, Manila, Philippines.

References

1. Vincencruz-Abeledo, C.C., Lagman, M.C.A.: A revised dichotomous key for the mangrove crab genus *scylla* de haan, 1833 (brachyura, portunidae). *Crustaceana* **91**(7), 847–865 (2018)
2. Keenan, C., Davie, P., Mann, D.: A revision of the genus *scylla* de haan, 1833 (crustacea: Decapoda: Brachyura: Portunidae). *The Raffles bulletin of zoology* **46**, 1–29 (1998)
3. Alberts-Hubatsch, H., Lee, S.Y., Meynecke, J.-O., Diele, K., Nordhaus, I., Wolff, M.: Life-history, movement, and habitat use of *scylla serrata* (decapoda, portunidae): current knowledge and future challenges. *Hydrobiologia* **763**(1), 5–21 (2015)
4. Shelley, C., Lovatelli, A.: Mud crab aquaculture - a practical manual. Technical report, FAO Fisheries and Aquaculture Technical Paper No. 567 (2011)
5. Williams, M.J., Primavera, J.H.: Choosing tropical portunid species for culture, domestication and stock enhancement in the indo-pacific **14** (2001)
6. Bureau of Fisheries and Aquatic Resources, Philippines: Philippine Fisheries Profile 2018 (2018)
7. Philippines Statistics Authority: Fisheries Statistics of the Philippines 2016 - 2018 (2019)

8. Azra, M.N., Chen, J.-C., Hsu, T.-H., Ikhwanuddin, M., Abol-Munafi, A.B.: Growth, molting duration and carapace hardness of blue swimming crab, portunus pelagicus, instars at different water temperatures. *Aquaculture Reports* **15**, 100226 (2019)
9. Kuhn, A.A., Darnell, M.Z.: Elevated temperature induces a decrease in intermolt period and growth per molt in the lesser blue crab callinectes similis williams, 1966 (decapoda: Brachyura: Portunidae). *Journal of Crustacean Biology* **39**(1), 22–27 (2019)
10. Sokolova, I., Lannig, G.: Interactive effects of metal pollution and temperature on metabolism in aquatic ectotherms: implications of global climate change. *Climate Research* **37**(2-3), 181–201 (2008)
11. Yang, Y., Ye, H., Huang, H., Li, S., Liu, X., Zeng, X., Gong, J.: Expression of hsp70 in the mud crab, scylla paramamosain in response to bacterial, osmotic, and thermal stress. *Cell Stress and Chaperones* **18**(4), 475–482 (2013)
12. Drozdova, P., Bedulina, D., Madyarova, E., Rivarola-Duarte, L., Schreiber, S., Stadler, P.F., Luckenbach, T., Timofeyev, M.: Description of strongly heat-inducible heat shock protein 70 transcripts from baikal endemic amphipods. *Scientific Reports* **9**(1) (2019)
13. Lin, X., Wu, X., Liu, X.: Temperature stress response of heat shock protein 90 (hsp90) in the clam paphia undulata. *Aquaculture and Fisheries* **3**(3), 106–114 (2018)
14. Loc, N.H., MacRae, T.H., Musa, N., Abdullah, M.D.D.B., Wahid, M.E.A., Sung, Y.Y.: Non-lethal heat shock increased hsp70 and immune protein transcripts but not vibrio tolerance in the white-leg shrimp. *PLoS ONE* **8**(9), 73199 (2013)
15. Imai, H., Numachi, K.-i.: Intra- and interspecific genetic variability and relationships among mud crabs, scylla spp (decapoda: Portunidae), demonstrated by RFLP analysis of mitochondrial DNA. *The journal of animal genetics* **29**(2), 3–11 (2002)
16. Rumisha, C., Huyghe, F., Rapanuel, D., Mascaux, N., Kochzius, M.: Genetic diversity and connectivity in the east african giant mud crab scylla serrata: Implications for fisheries management. *PLOS ONE* **12**(10), 0186817 (2017)
17. Corporal-Lodangco, I.L., Leslie, L.M.: Defining philippine climate zones using surface and high-resolution satellite data. *Procedia Computer Science* **114**, 324–332 (2017)
18. Thomas, G.W.C., Dohmen, E., Hughes, D.S.T., Murali, S.C., Poelchau, M., Glastad, K., Anstead, C.A., Ayoub, N.A., Batterham, P., Bellair, M., Binford, G.J., Chao, H., Chen, Y.H., Childers, C., Dinh, H., Doddapaneni, H.V., Duan, J.J., Dugan, S., Esposito, L.A., Friedrich, M., Garb, J., Gasser, R.B., Goodman, M.A.D., Gundersen-Rindal, D.E., Han, Y., Handler, A.M., Hatakeyama, M., Hering, L., Hunter, W.B., Ioannidis, P., Jayaseelan, J.C., Kalra, D., Khila, A., Korhonen, P.K., Lee, C.E., Lee, S.L., Li, Y., Lindsey, A.R.I., Mayer, G., McGregor, A.P., McKenna, D.D., Misof, B., Mundasa, M., Munoz-Torres, M., Muzny, D.M., Niehuis, O., Osuji-Lacy, N., Palli, S.R., Panfilio, K.A., Pechmann, M., Perry, T., Peters, R.S., Poynton, H.C., Prpic, N.-M., Qu, J., Rotenberg, D., Schal, C., Schoville, S.D., Scully, E.D., Skinner, E., Sloan, D.B., Stouthamer, R., Strand, M.R., Szucsich, N.U., Wijeratne, A., Young, N.D., Zattara, E.E., Benoit, J.B., Zdobnov, E.M., Pfrender, M.E., Hackett, K.J., Werren, J.H., Worley, K.C., Gibbs, R.A., Chipman, A.D., Waterhouse, R.M., Bornberg-Bauer, E., Hahn, M.W., Richards, S.: Gene content evolution in the arthropods. *Genome Biology* **21**(1) (2020)
19. Robinson, M.D., Oshlack, A.: A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**(3), 25 (2010)
20. Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L.: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**(3), 25 (2009)
21. Dammit Pipeline. <https://github.com/dib-lab/dammit>
22. Kielbasa, S.M., Wan, R., Sato, K., Horton, P., Frith, M.C.: Adaptive seeds tame genomic sequence comparison. *Genome Research* **21**(3), 487–493 (2011)
23. Aubry, S., Kelly, S., Kümpers, B.M.C., Smith-Unna, R.D., Hibberd, J.M.: Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of c4 photosynthesis. *PLoS Genetics* **10**(6), 1004365 (2014)
24. Mi, H., Muruganujan, A., Casagrande, J.T., Thomas, P.D.: Large-scale gene function analysis with the panther classification system. *Nature protocols* **8**, 1551–1566 (2013)
25. Stupack, D.G.: Get a ligand, get a life: integrins, signaling and cell survival. *Journal of Cell Science* **115**(19), 3729–3738 (2002)
26. Zheng, L., Ishii, Y., Tokunaga, A., Hamashima, T., Shen, J., Zhao, Q.-L., Ishizawa, S., Fujimori, T., Nabeshima, Y.-i., Mori, H., Kondo, T., Sasahara, M.: Neuroprotective effects of PDGF against oxidative stress and the signaling pathway involved. *Journal of Neuroscience Research*, (2009)
27. Duffy, D.J., Millane, R.C., Frank, U.: A heat shock protein and wnt signaling crosstalk during axial patterning and stem cell proliferation. *Developmental biology* **362**, 271–281 (2012)
28. Sun, L., Lamont, S.J., Cooksey, A.M., McCarthy, F., Tudor, C.O., Vijay-Shanker, K., DeRita, R.M., Rothschild, M., Ashwell, C., Persia, M.E., Schmidt, C.J.: Transcriptome response to heat stress in a chicken hepatocellular carcinoma cell line. *Cell Stress and Chaperones* **20**(6), 939–950 (2015)
29. Waiho, K., Fazhan, H., Shahreza, M.S., Moh, J.H.Z., Noorbaiduri, S., Wong, L.L., Sinnasamy, S., Ikhwanuddin, M.: Transcriptome analysis and differential gene expression on the testis of orange mud crab, scylla olivacea, during sexual maturation. *PLOS ONE* **12**(1), 0171095 (2017)
30. Gao, J., Wang, X., Zou, Z., Jia, X., Wang, Y., Zhang, Z.: Transcriptome analysis of the differences in gene expression between testis and ovary in green mud crab *scylla paramamosain*. *BMC Genomics* **15**(1), 585 (2014)
31. Ma, H., Ma, C., Li, S., Jiang, W., Li, X., Liu, Y., Ma, L.: Transcriptome analysis of the mud crab (scylla paramamosain) by 454 deep sequencing: Assembly, annotation, and marker discovery. *PLoS ONE* **9**(7), 102668 (2014)
32. Zhang, Y., Wu, Q., Fang, S., Li, S., Zheng, H., Zhang, Y., Ikhwanuddin, M., Ma, H.: mrna profile provides novel insights into stress adaptation in mud crab megalopa, *scylla paramamosain* after salinity stress. *BMC Genomics* **21**(1) (2020)
33. Barshis, D.J., Ladner, J.T., Oliver, T.A., Seneca, F.O., Taylor-Knowles, N., Palumbi, S.R.: Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1387–1392 (2013)
34. Hopkin, R.S., Qari, S., Bowler, K., Hyde, D., Cuculescu, M.: Seasonal thermal tolerance in marine crustacea. *Journal of Experimental Marine Biology and Ecology* **331**(1), 74–81 (2006)
35. Poong, S.-W., Lee, K.-K., Lim, P.-E., Pai, T.-W., Wong, C.-Y., Phang, S.-M., Chen, C.-M., Yang, C.-H., Liu, C.-C.: RNA-seq-mediated transcriptomic analysis of heat stress response in a polar chlorella sp. (trebouxiphyceae, chlorophyta). *Journal of Applied Phycology* **30**(6), 3103–3119 (2018)
36. Magsino, R.M., Juñio-Méñez, M.A.: The influence of contrasting life history traits and oceanic processes on genetic structuring of rabbitfish populations *siganus argenteus* and *siganus fuscescens* along the eastern philippine coasts. *Marine Biology* **154**(3), 519–532 (2008)
37. Raynal, J., Crandall, E., Barber, P., Mahardika, G., Lagman, M., Carpenter, K.: Basin isolation and oceanographic features influencing lineage divergence in the humbug damselfish *dascyllus aruanus* in the coral triangle. *Bulletin of Marine Science* **90**(1), 513–532 (2014)
38. LOURIE, S.A., GREEN, D.M., VINCENT, A.C.J.: Dispersal, habitat differences, and comparative phylogeography of southeast asian seahorses (syngnathidae: Hippocampus). *Molecular Ecology* **14**(4), 1073–1094 (2005)
39. Vince Cruz, C.C.E., Ablan-Lagman, M.C.: Population structure of scylla serrata from microsatellite and mtDNA markers. In: In E. T. Quinitio, F. D. Parado-Estepa, & R. M. Coloso (Eds.)Philippines : In the Forefront of the Mud Crab Industry Development : Proceedings of the 1st National Mud Crab Congress, Tigbauan, Iloilo, Philippines, pp. 1–12 (2015). Aquaculture Department, Southeast Asian Fisheries Development Center
40. Akbarian, A., Michiels, J., Degroote, J., Majeddin, M., Golian, A., De Smet, S.: Association between heat stress and oxidative stress in poultry: mitochondrial dysfunction and dietary interventions with phytochemicals. *Journal of animal science and biotechnology* **7**, 37 (2016)
41. Slimen, I.B., Najar, T., Ghram, A., Dabbebi, H., Mrad, M.B., Abdrrabbah, M.: Reactive oxygen species, heat stress and oxidative-induced mitochondrial damage. a review. *International*

- Journal of Hyperthermia **30**(7), 513–523 (2014)
42. Gu, Z.T., Li, L., Wu, F., Zhao, P., Yang, H., Liu, Y.S., Geng, Y., Zhao, M., Su, L.: Heat stress induced apoptosis is triggered by transcription-independent p53, ca(2+) dyshomeostasis and the subsequent bax mitochondrial translocation. *Scientific reports* **5**, 11497 (2015)
 43. Quinitio, E.T., de la Cruz, J.J., Eguia, M.R.R., Parado-Estepa, F.D., Pates, G., Lavilla-Pitogo, C.R.: Domestication of the mud crab *scylla serrata*. *Aquaculture International* **19**(2), 237–250 (2010)
 44. Quinitio, E.T.: Overview of the mud crab industry in the philippines. In: In E. T. Quinitio, F. D. Parado-Estepa, & R. M. Coloso (Eds.) Philippines : In the Forefront of the Mud Crab Industry Development : Proceedings of the 1st National Mud Crab Congress, Tigbauan, Iloilo, Philippines, pp. 1–12 (2015). Aquaculture Department, Southeast Asian Fisheries Development Center
 45. Tepolt, C.K., Somero, G.N.: Master of all trades: thermal acclimation and adaptation of cardiac function in a broadly distributed marine invasive species, the european green crab, *carcinus maenas*. *Journal of Experimental Biology* **217**(7), 1129–1138 (2014)
 46. Kopylova, E., Noé, L., Touzet, H.: Sortmerna: fast and accurate filtering of ribosomal rnas in metatranscriptomic data. *Bioinformatics* (Oxford, England) **28**, 3211–3217 (2012)
 47. Bolger, A.M., Lohse, M., Usadel, B.: Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* (Oxford, England) **30**, 2114–2120 (2014)
 48. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A.: De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols* **8**(8), 1494–1512 (2013)
 49. Li, B., Dewey, C.N.: RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**(1) (2011)
 50. Love, M.I., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**(12) (2014)
 51. Soneson, C., Love, M.I., Robinson, M.D.: Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2015)
 52. Stephens, M.: False discovery rates: a new deal. *Biostatistics*, 041 (2016)
 53. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**(1), 289–300 (1995)
 54. Pebesma, E., Bivand, R.: Classes and methods for spatial data in r. *R News* **5** (2005)
 55. Global Administrative Areas (2018). GADM database of Global Administrative Areas version 3.6. www.gadm.org
 56. Huang, B., Thorne, P.W., Banzon, V.F., Boyer, T., Chepurin, G., Lawrimore, J.H., Menne, M.J., Smith, T.M., Vose, R.S., Zhang, H.-M.: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate* **30**(20), 8179–8205 (2017)

Figure Legends

Figure 1: (a) The 3 regions in the Philippines, Bataan (BAT), Bicol (BIC), and Cagayan (CAG), from where the *S. serrata* individuals were sampled. The map was produced using the R package sp [54] based on data from the Data of Global Administrative Areas, GADM version 3.6. [55]. (b) Distribution of monthly average temperature between 1900 – 2019 and (c) Distribution of monthly average sea-surface temperature anomaly. Sea-surface temperature data was obtained from ERSST version 5 [56].

Figure 2: ExN50, which is the N50 length when considering only the top expressing genes that account for x percentage of expression, shown here for different values of x .

Figure 3: Venn diagram illustrating the number of hits found in Swiss-Prot, the fruitfly *D. melanogaster* proteome, and the shrimp *P. vannamei*

proteome, for transcripts (right) and for genes (left), where a gene is said to have a hit if at least one of its isoforms in the assembly has a hit.

Figure 4: Gene ontology annotation of the de-novo transcriptome assembly. Number of transcripts assigned to top gene ontology terms for each of the three GO domains: Biological Process (BP), B) Molecular Function (MF) and C) Cellular Component (CC).

Figure 5: Top ten representative pathways of the de-novo transcriptome assembly annotated from the PANTHER database.

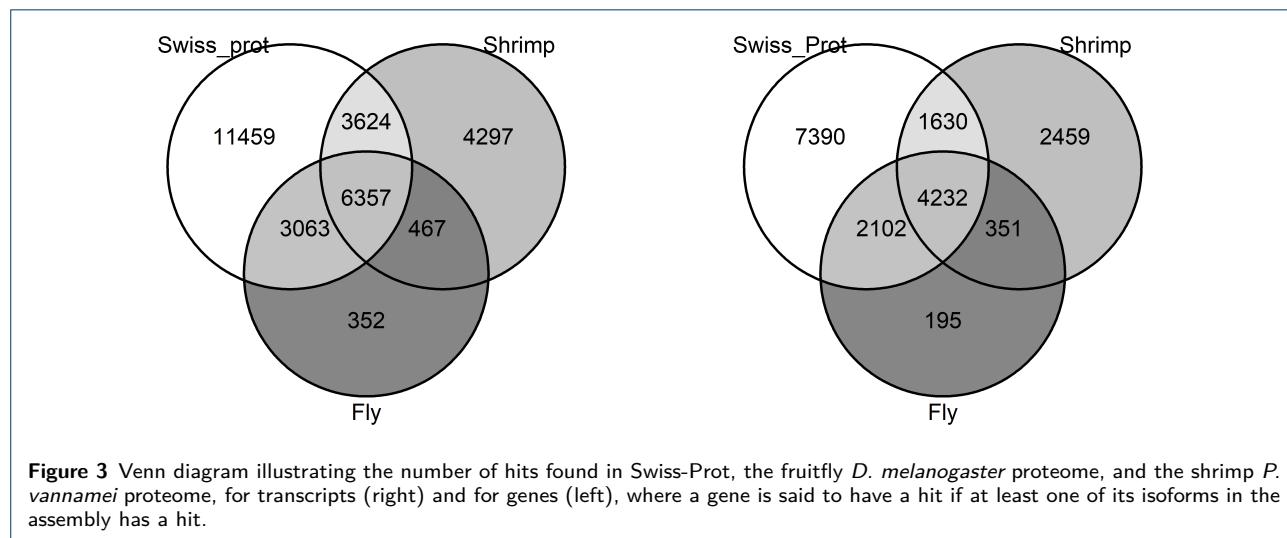
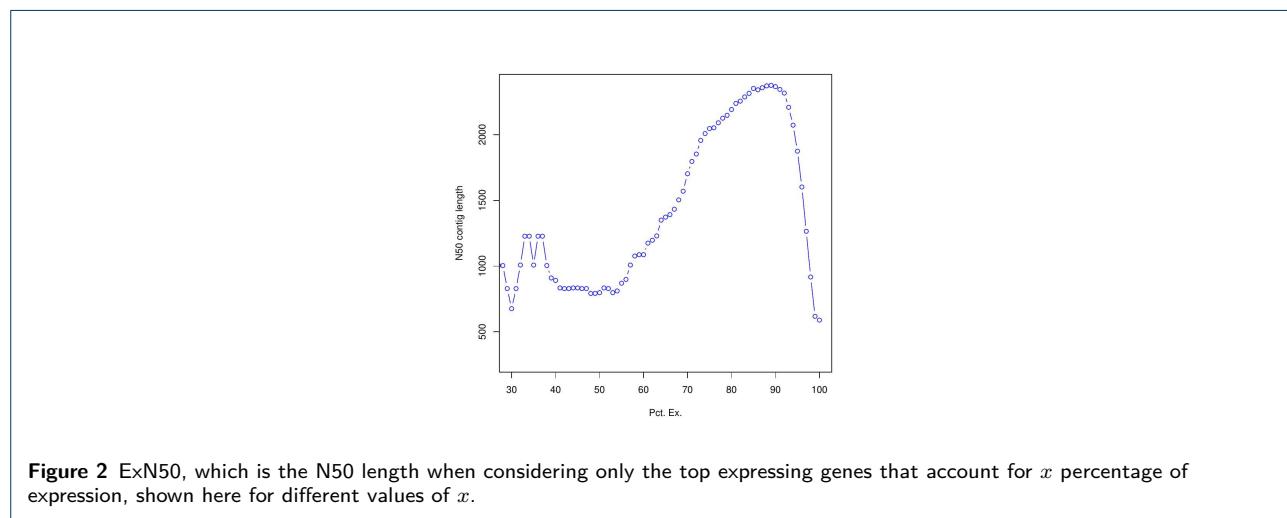
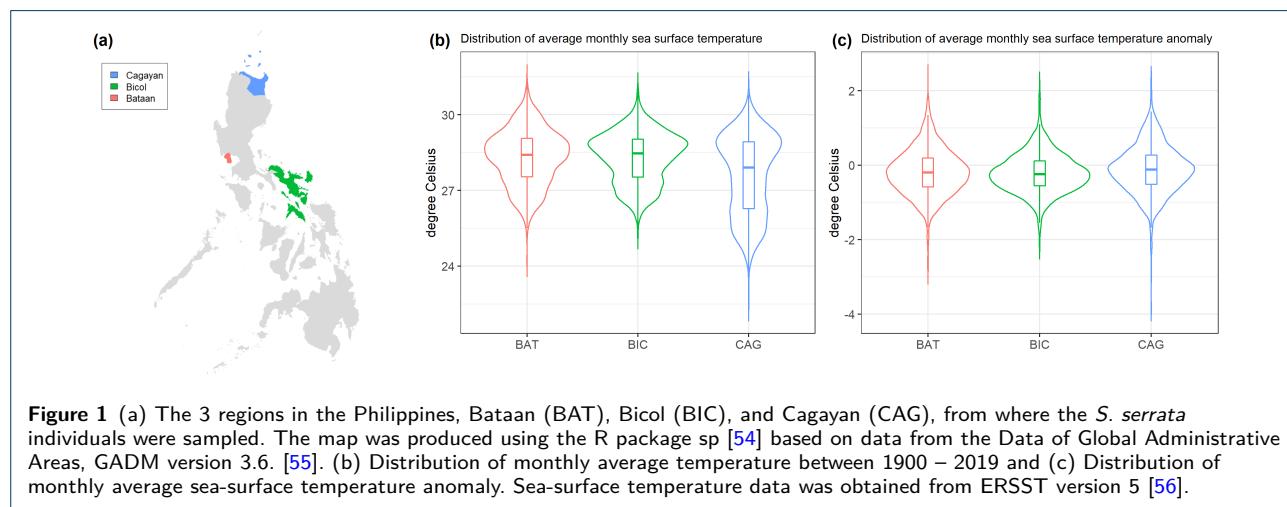
Figure 6: DE genes. Note: CAG = Cagayan, BAT = Bataan, BIC = Bicol (a) Bar graph showing number of DE genes classified according to the direction of regulation (b) Scatter plot of genes showing mean expression (measured as mean of normalized read counts) vs. log fold change for the 3 sites. Non-grey dots represent DE genes, red dots represent stress-related DE genes.

Figure 7: DDE genes. Note: CAG = Cagayan, BAT = Bataan, BIC = Bicol, C = Control, E = Heat-stressed. Heat map showing expression profile of the DDE genes. Each row corresponds to a DDE gene, each column to a site-treatment combination, and the color of each cell represents expression level – green for higher than row average, red for lower than row average, and the intensity of color represents the z-score of the normalized, variance-stabilized read counts.

Figures

Additional Files

Additional file 1 — Supplementary Information



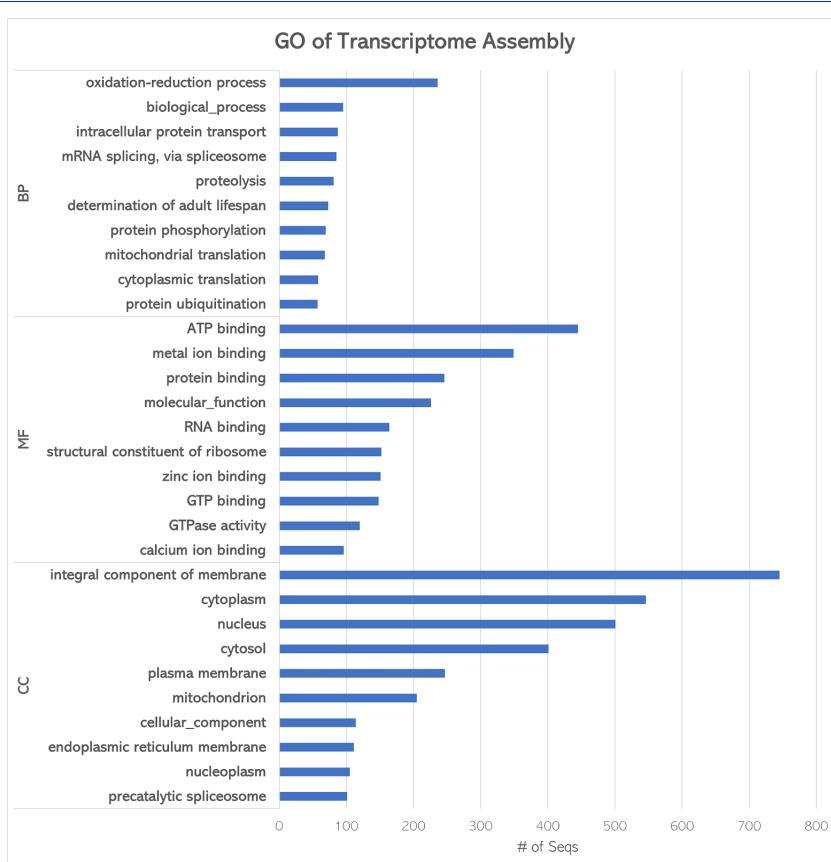


Figure 4 Gene ontology annotation of the de-novo transcriptome assembly. Number of transcripts assigned to top gene ontology terms for each of the three GO domains: Biological Process (BP), B) Molecular Function (MF) and C) Cellular Component (CC).

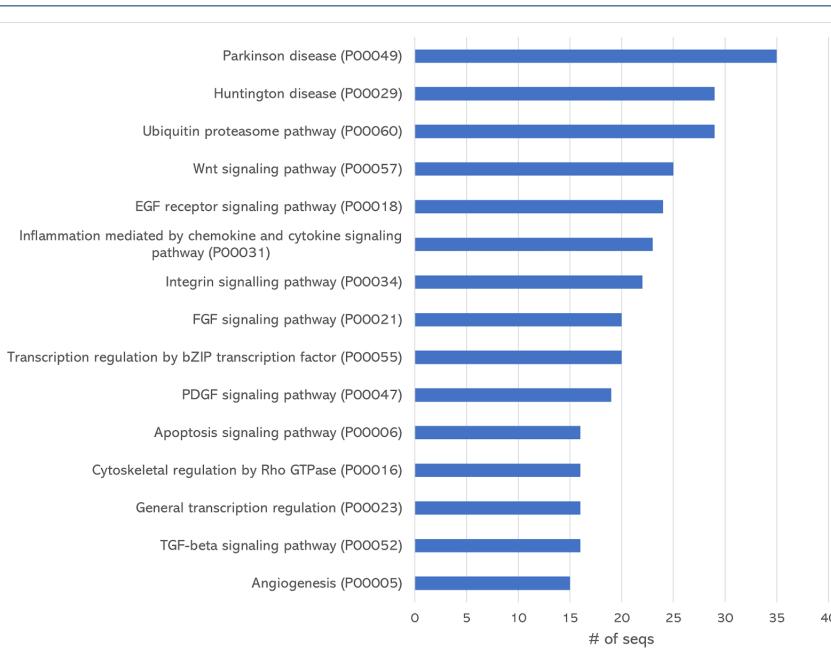


Figure 5 Top ten representative pathways of the de-novo transcriptome assembly annotated from the PANTHER database.

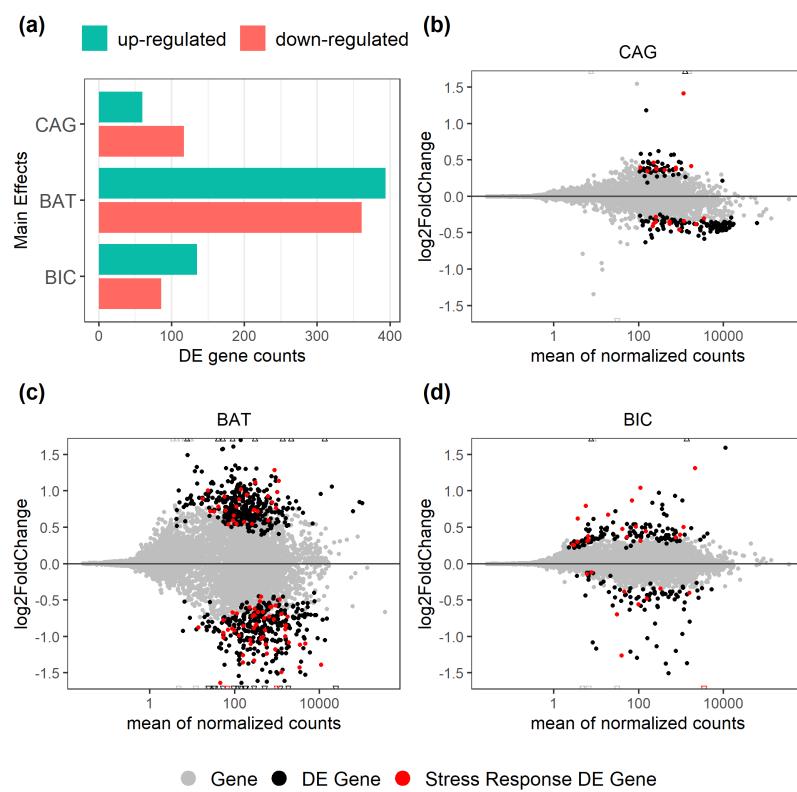


Figure 6 DE genes. Note: CAG = Cagayan, BAT = Bataan, BIC = Bicol (a) Bar graph showing number of DE genes classified according to the direction of regulation (b) Scatter plot of genes showing mean expression (measured as mean of normalized read counts) vs. log fold change for the 3 sites. Non-grey dots represent DE genes, red dots represent stress-related DE genes.

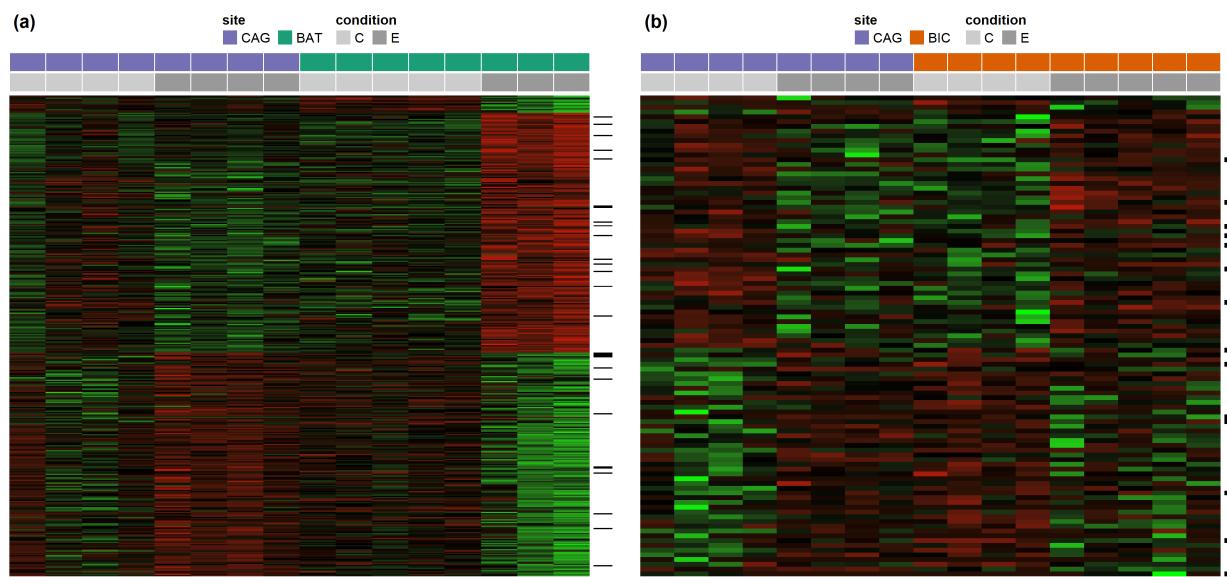


Figure 7 DDE genes. Note: CAG = Cagayan, BAT = Bataan, BIC = Bicol, C = Control, E = Heat-stressed. Heat map showing expression profile of the DDE genes. Each row corresponds to a DDE gene, each column to a site-treatment combination, and the color of each cell represents expression level – green for higher than row average, red for lower than row average, and the intensity of color represents the z-score of the normalized, variance-stabilized read counts.

Figures

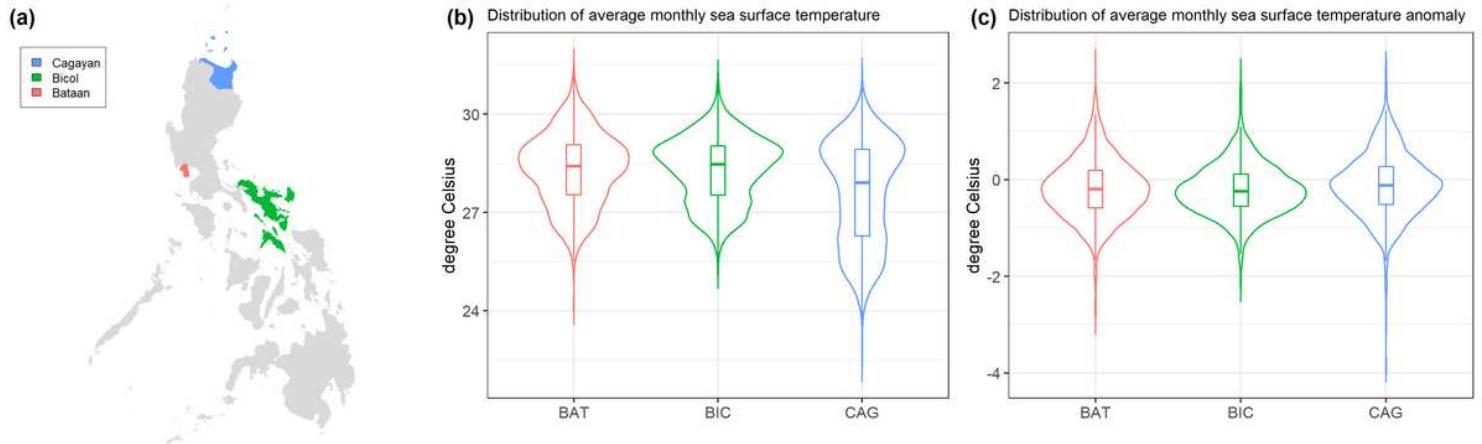


Figure 1

(a) The 3 regions in the Philippines, Bataan (BAT), Bicol (BIC), and Cagayan (CAG), from where the *S. serrata* individuals were sampled. The map was produced using the R package sp [54] based on data from the Data of Global Administrative Areas, GADM version 3.6. [55]. (b) Distribution of monthly average temperature between 1900 – 2019 and (c) Distribution of monthly average sea-surface temperature anomaly. Sea-surface temperature data was obtained from ERSST version 5 [56].

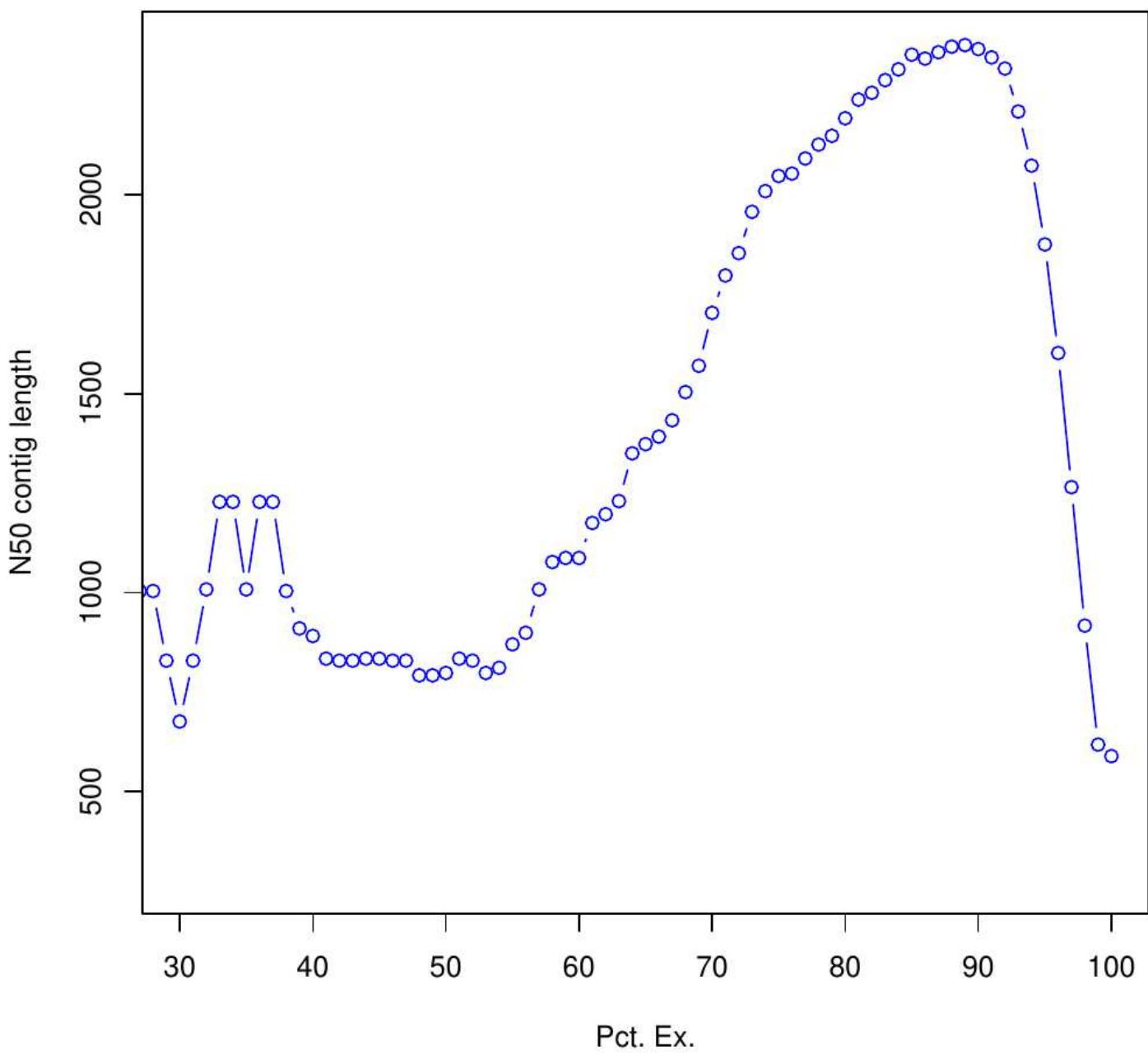


Figure 2

ExN50, which is the N50 length when considering only the top expressing genes that account for x percentage of expression, shown here for different values of x.

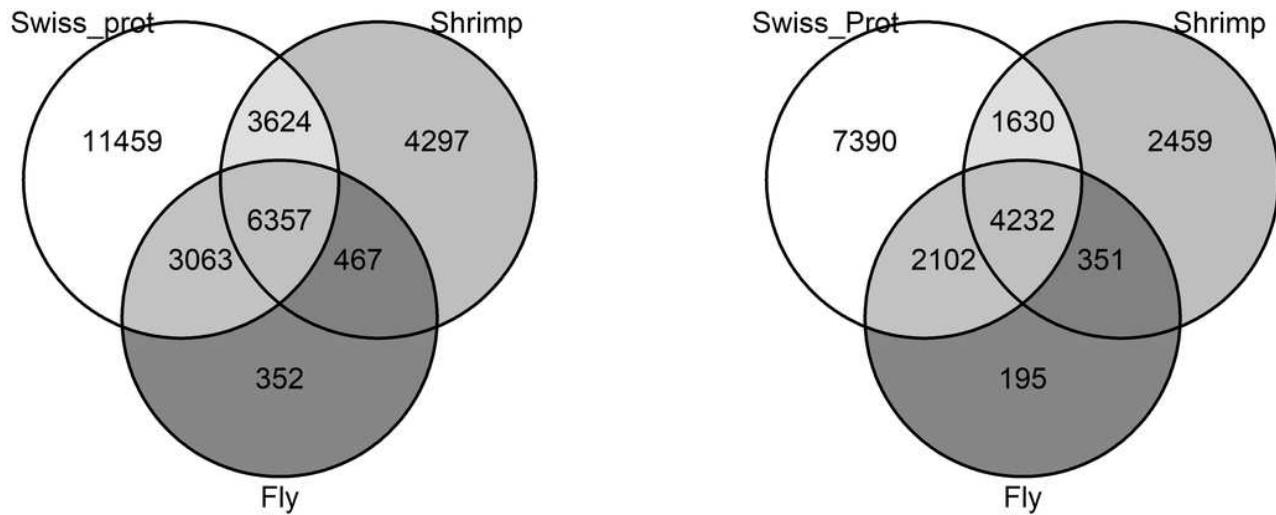


Figure 3

Venn diagram illustrating the number of hits found in Swiss-Prot, the fruitfly *D. melanogaster* proteome, and the shrimp *P. vannamei* proteome, for transcripts (right) and for genes (left), where a gene is said to have a hit if at least one of its isoforms in the assembly has a hit.

GO of Transcriptome Assembly

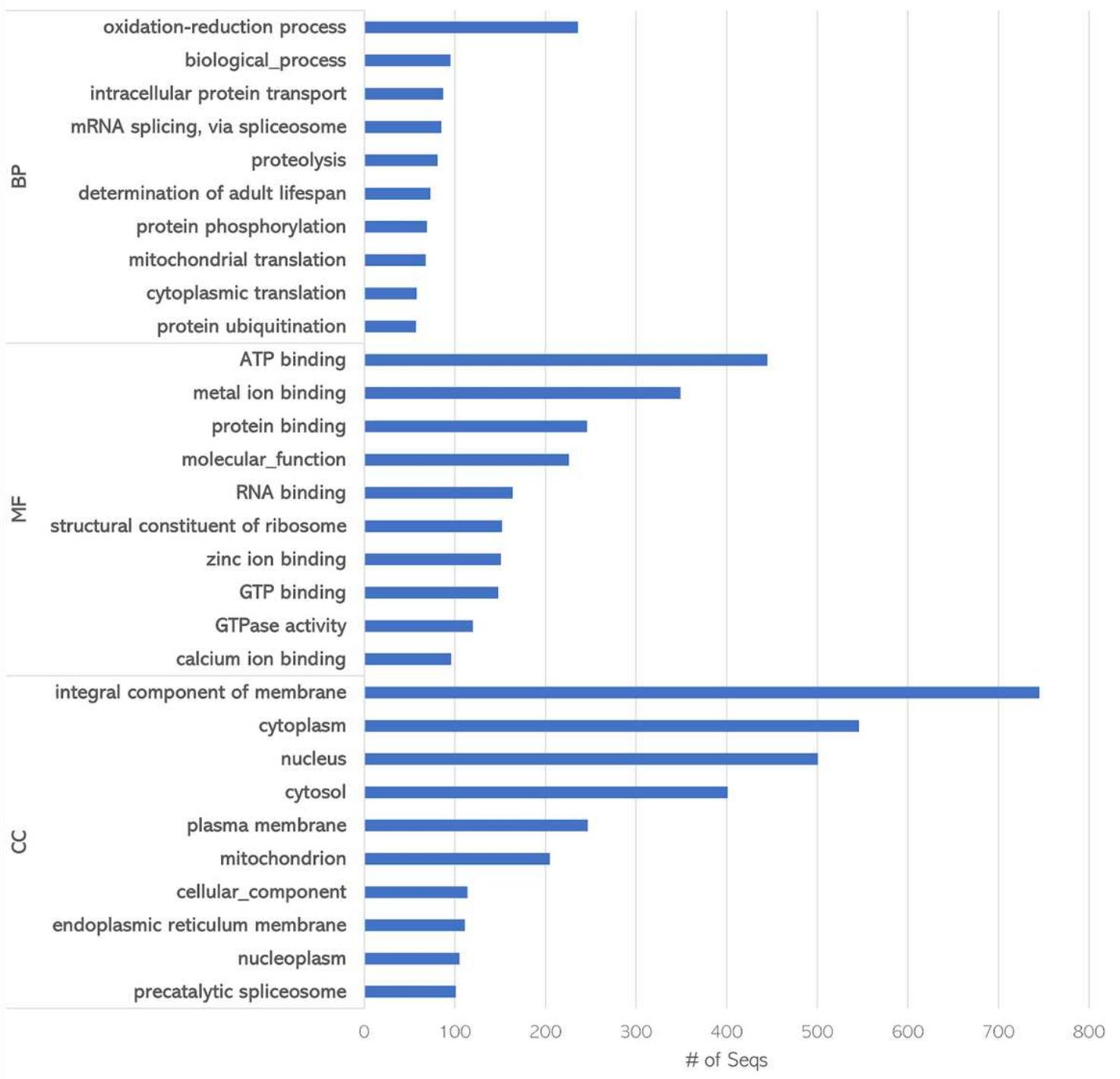


Figure 4

Gene ontology annotation of the de-novo transcriptome assembly. Number of transcripts assigned to top gene ontology terms for each of the three GO domains: Biological Process (BP), B) Molecular Function (MF) and C) Cellular Component (CC).

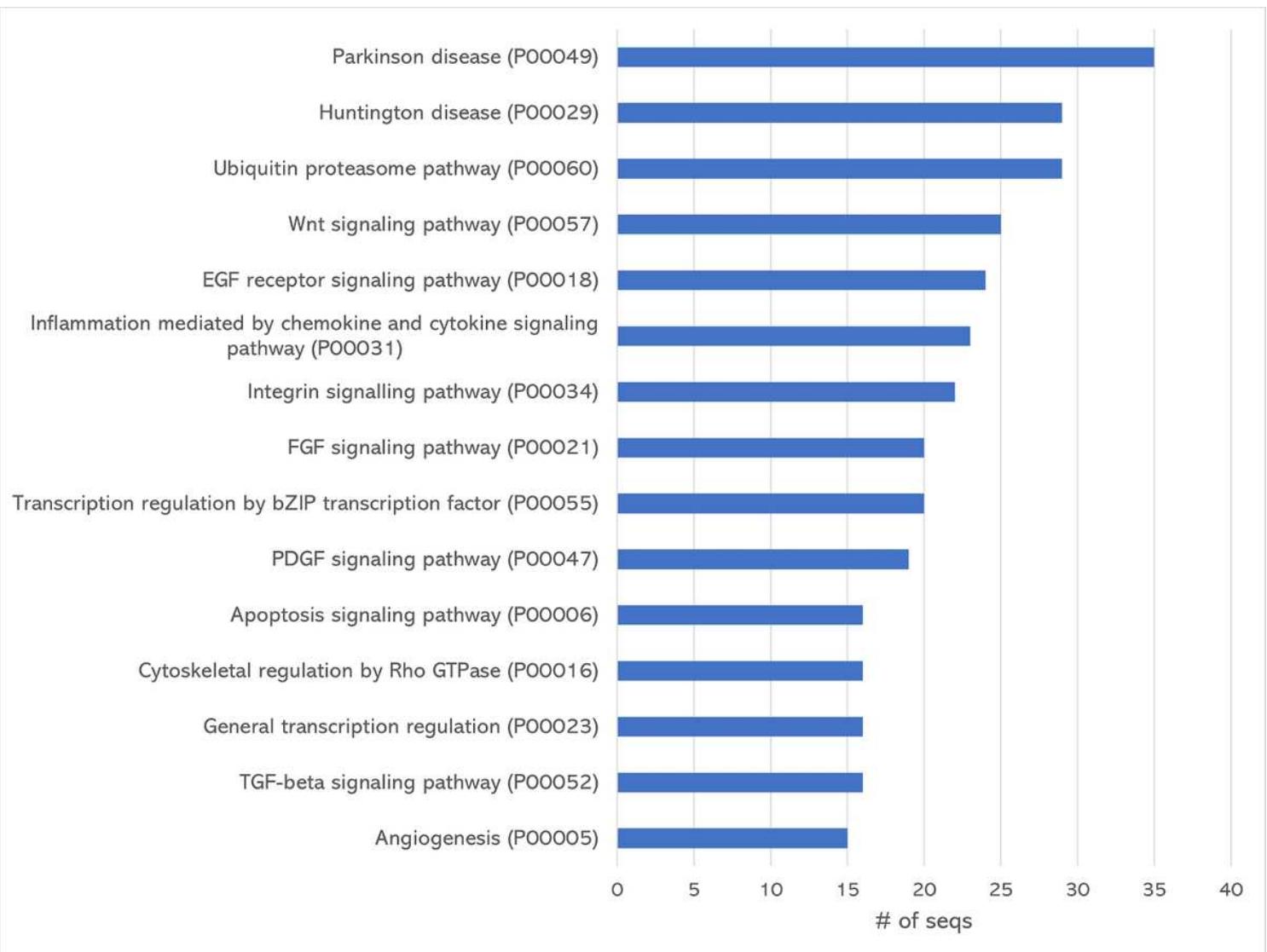


Figure 5

Top ten representative pathways of the de-novo transcriptome assembly annotated from the PANTHER database.

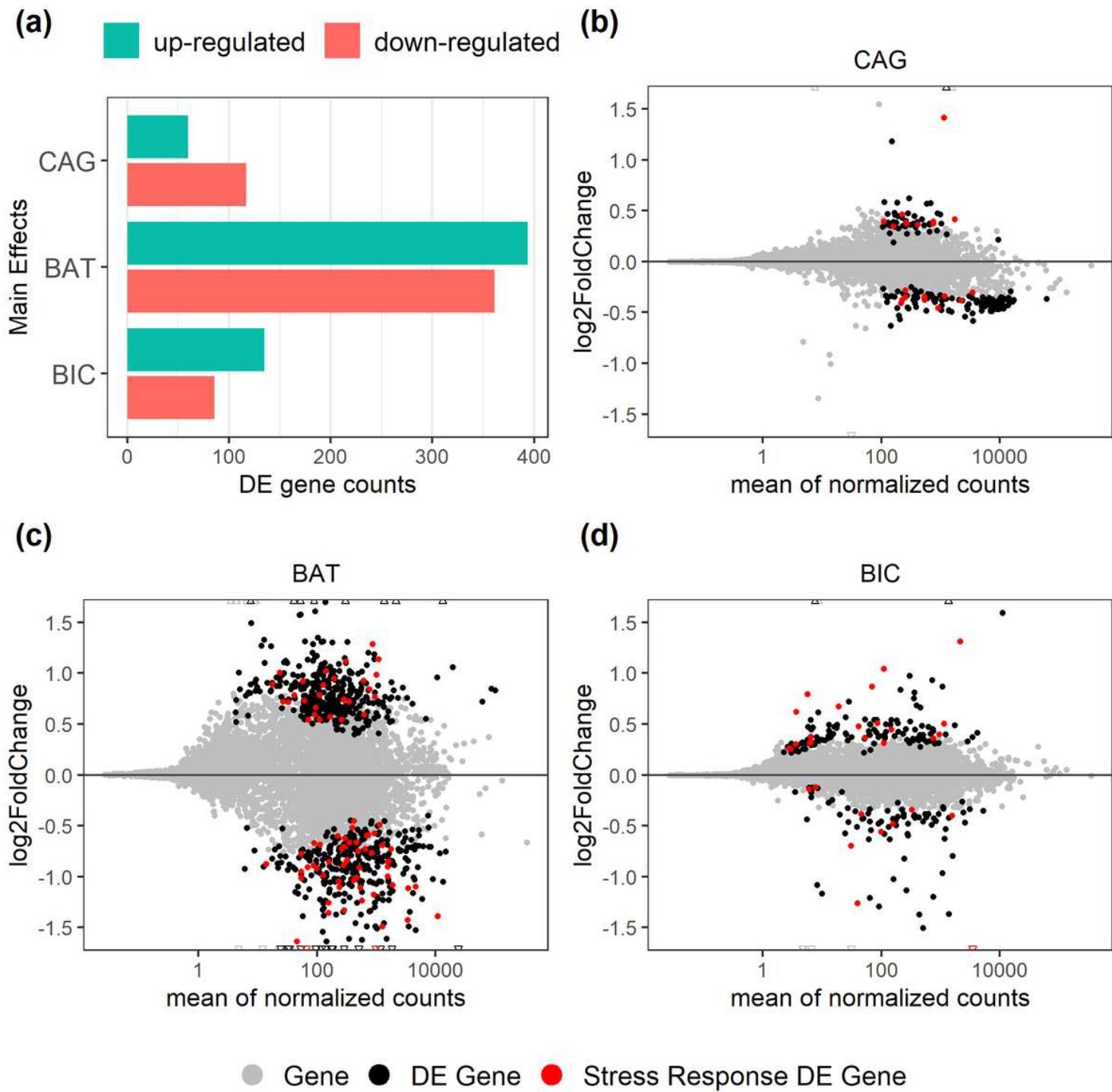


Figure 6

DE genes. Note: CAG = Cagayan, BAT = Bataan, BIC = Bicol (a) Bar graph showing number of DE genes classified according to the direction of regulation (b) Scatter plot of genes showing mean expression (measured as mean of normalized read counts) vs. log fold change for the 3 sites. Non-grey dots represent DE genes, red dots represent stress-related DE genes.

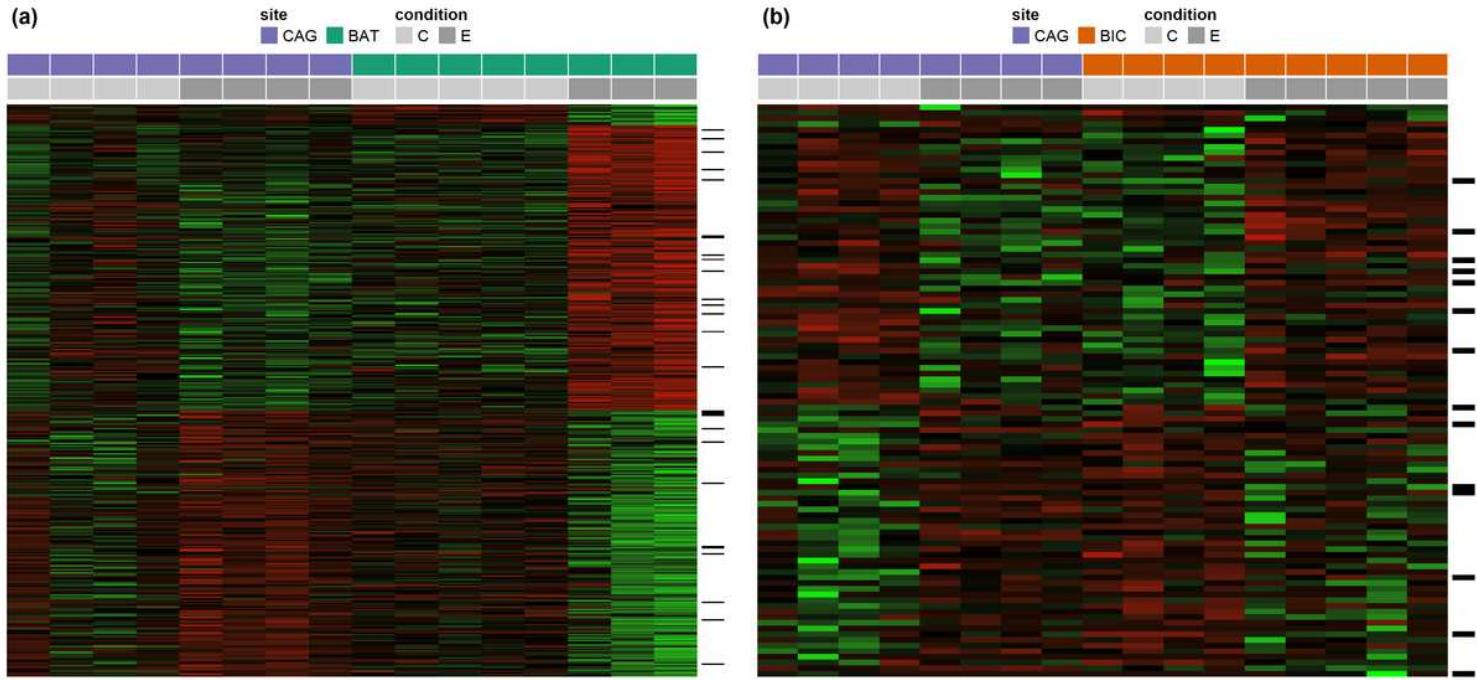


Figure 7

DDE genes. Note: CAG = Cagayan, BAT = Bataan, BIC = Bicol, C = Control, E = Heat-stressed. Heat map showing expression profile of the DDE genes. Each row corresponds to a DDE gene, each column to a site-treatment combination, and the color of each cell represents expression level – green for higher than row average, red for lower than row average, and the intensity of color represents the z-score of the normalized, variance-stabilized read counts.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinfo.pdf](#)