Transcriptome analysis of the molting gland (Y-organ) from the blackback land crab,
Gecarcinus lateralis

Sunetra Das¹, Natalie L. Pitts¹, Megan R. Mudron¹; David S. Durica², and Donald L. Mykles^{1*}

¹Department of Biology, Colorado State University, Fort Collins, CO 80523

²Department of Biology, University of Oklahoma, Norman, OK 73019

Keywords: Blast2GO; Crustacea; Decapoda; ecdysteroid; *Gecarcinus*; molting; mTOR; transforming growth factor-beta; transcriptome; Y-organ; signal transduction; molt-inhibiting hormone

7 figures; 7 tables

Appendix: 2 figures; 2 tables

*Corresponding author:

Dr. Donald L. Mykles
Department of Biology
Colorado State University
Fort Collins, CO 80523 USA
Tel: 970-491-7616
Email: Donald.Mykles@colostate.edu

Abstract

In decapod crustaceans, arthropod steroid hormones or ecdysteroids, regulate molting. These hormones are synthesized and released from a pair of molting glands called the Y-organs (YO). Cyclic nucleotide, mTOR, and TGF β /Smad signaling pathways mediate molt cycle-dependent phase transitions in the YO. To further identify the genes involved in the regulation of molting, a YO transcriptome was generated from three biological replicates of intermolt blackback land crab, *Gecarcinus lateralis*. Illumina sequencing of cDNA libraries generated 227,811,829 100-base pair (bp) paired-end reads; following trimming, 90% of the reads were used for further analyses. The trimmed reads were assembled *de novo* using Trinity software to generate 288,673 contigs with a mean length of 872 bp and a median length of 1,842 bp. Redundancy among contig sequences was reduced by CD-HIT-EST and the output constituted the baseline transcriptome database. Using Bowtie2, 92% to 93% of the reads were mapped back to the transcriptome. Individual contigs were annotated using BLAST, HMMER, TMHMM, SignalP, and Trinotate, resulting in assignments of 20% of the contigs. Functional and pathway annotations were carried out via Gene Ontology (GO) and KEGG Orthology (KO) analyses; 58% and 44% of the contigs with BLASTx hits were assigned to GO and KO terms, respectively. The gene expression profile was similar to a crayfish YO transcriptome database and the relative abundance of each contig was highly correlated among the three *G. lateralis* replicates. Signal transduction pathway orthologs were well represented, including those in the mTOR, TGF β , cyclic nucleotide, MAP kinase, calcium, VEGF, phosphatidylinositol, ErbB, Wnt, Hedgehog, Jak-STAT, and Notch pathways.

1. Introduction

Decapod crustaceans possess a rigid exoskeleton that must be shed periodically for organismal growth, a process called ecdysis or molting. Molting is regulated by two endocrine organs: the X-organ/sinus gland (XO/SG) complex, located in the eyestalk ganglia, and a pair of molting glands, or Y-organs (YOs), located in the cephalothorax (Hopkins, 2012; Skinner, 1985). The interaction between molt-inhibiting hormone (MIH), an inhibitory neuropeptide produced by the XO/SG complex, and steroid molting hormones (ecdysteroids), produced by the YO, drives the progression through the molt cycle (Webster, 2015; Webster et al., 2012). Hemolymph ecdysteroid titers are low during the intermolt stage, increase during the premolt stage, drop precipitously near the end of premolt, and remain at low levels during postmolt (Chang and Mykles, 2011; Mykles, 2011). The increase in ecdysteroid titers initiate and coordinate the physiological processes required for molting, such as degradation and reformation of the exoskeleton (Skinner, 1985), claw muscle atrophy (Mykles, 1997; Mykles and Medler, 2015), and limb regeneration (Hopkins and Das, 2015; Mykles, 2001; Skinner, 1985).

The YO undergoes transitions in physiological properties at critical stages of the molt cycle. During intermolt, the YO is kept in the basal state by pulsatile releases of MIH to maintain low hemolymph ecdysteroid titers (Chung et al., 2010; Nakatsuji et al., 2009). Repression of ecdysteroid synthesis via MIH signaling in the YO is mediated by cyclic nucleotide second messengers (Covi et al., 2009). MIH binding to a putative G protein-coupled membrane receptor is hypothesized to initiate a cAMP-dependent triggering phase, which is followed by a NO/cGMP-dependent summation phase for prolonged inhibition of ecdysteroidogenesis between MIH pulses (Chang and Mykles, 2011; Covi et al., 2012; Webster, 2015). A reduction in MIH during intermolt de-represses the YO. The activated YO hypertrophies to increase ecdysteroid

synthesis; the hemolymph ecdysteroid titer increases and the animal transitions to the premolt stage (Chang and Mykles, 2011). The duration of early premolt depends on environmental and physiological conditions. Environmental stress, acting through the XO/SG complex, can prolong the premolt period, as the YO remains sensitive to MIH and a stress neuropeptide, crustacean hyperglycemic hormone (CHH) (Chang and Mykles, 2011; Shrivastava and Princy, 2014). Autotomy of a limb regenerate suspends premolt for a few weeks, which allows time for a new regenerate to form and grow, and the animal molts with a complete set of walking legs (Chang and Mykles, 2011; Mykles, 2001; Yu et al., 2002). A critical transition occurs at mid-premolt, when the animal becomes committed to molt. The YO increases ecdysteroid production further and becomes insensitive to MIH and CHH (Chang and Mykles, 2011; Covi et al., 2012). The animal progresses through to ecdysis without delay. If a limb regenerate is autotomized at this stage, there is no regeneration and the animal molts without a full complement of legs (Mykles, 2001; Yu et al., 2002).

Two highly conserved signaling pathways mediate the molt cycle-dependent transitions of the YO. Activation of the YO in early premolt requires mechanistic Target of Rapamycin (mTOR)-dependent protein synthesis. mTOR is a protein kinase that controls global translation of mRNA into protein in eukaryotic cells (Baretic and Williams, 2014). Its activity is controlled by a variety of signals that regulate energy allocation to protein synthesis, such as nutrients, cellular energy status, growth factors, and stress (Albert and Hall, 2015; Cetrullo et al., 2015; Shimobayashi and Hall, 2014). Cycloheximide, an inhibitor of mRNA translation, and rapamycin, an mTOR inhibitor, repress YO ecdysteroid secretion *in vitro* (Abuhagr et al., 2014b; Mattson and Spaziani, 1986, 1987). Moreover, increases in mRNA levels of mTOR signaling components *Gl-mTOR* and *Gl-Akt*, as well as *Gl-elongation factor 2 (Gl-EF2)* in mid-premolt

and late premolt stages coincide with the increase in ecdysteroid production in the committed YO (Abuhagr et al., 2014b). The transition of the YO from the activated to the committed state involves TGF β /Smad signaling via an Activin-like membrane receptor (Chang and Mykles, 2011). An Activin receptor inhibitor (SB431542) blocks YO commitment, but has no effect on YO activation in eyestalk-ablated *G. lateralis in vivo* (Abuhagr et al., 2012). Both signaling pathways are also important in insect molt regulation. mTOR activity controls the size and ecdysteroid synthetic capacity of the prothoracic gland (PG) by prothoracicotropic hormone (PTTH), bombyxin, and insulin-like peptides (ILPs) (Covi et al., 2012; Gu et al., 2015; Hatem et al., 2015; Smith et al., 2014; Teleman, 2010; Yamanaka et al., 2013). A recent report of an ortholog of *Drosophila* Ilp7 and eight insulin-like growth factor binding proteins in the rock lobster, *Sagmariasus verreauxi* (Chandler et al., 2015), lends support for the importance of ILP/mTOR signaling in crustaceans, and raises the possibility that the YO, like the insect PG, is regulated by insulin-like growth factors (Yamanaka et al., 2013). Moreover, TGF β signaling, which is mediated by Smad transcription factors (Macias et al., 2015), is necessary for the PG to respond to PTTH and insulin. Blocking Activin/Smad signaling in the *Drosophila* PG prevents the PTTH-triggered ecdysteroid peak that initiates metamorphosis (Gibbens et al., 2011; Pentek et al., 2009). Taken together, the data suggest that mTOR and TGF β /Smad pathways play essential roles in the regulation of crustacean and insect molting glands by neuropeptides.

RNA-Seq technology has quickly become a powerful tool in decapod crustacean physiology, as it provides a deeper and broader range of transcripts than other methods (Wang et al., 2009). As the field lacks a decapod species with a fully mapped and annotated genome, *de novo* assembly of RNA-Seq data allows the cataloging of all the genes expressed in a tissue, essentially leapfrogging the genome to the direct analysis of genes that define a specific function.

This methodology can identify gene ontologies and networks associated with a physiological process, as well as quantify levels of mRNA abundance for all genes transcriptionally activated within that physiological state. Transcriptomic approaches have revealed genes associated with reproduction (Gao et al., 2014), development (Chandler et al., 2015; Christiaens et al., 2015; Li et al., 2015; Wei et al., 2014a), chitin metabolism in integumentary tissues (Abehsera et al., 2015; Tom et al., 2014), digestion (Wei et al., 2014b), neuroendocrine regulation (Christie, 2014; Ventura et al., 2014), and molting and growth (Lv et al., 2014; Tom et al., 2013). In the present study, Illumina high-throughput sequencing and *de novo* assembly was used to create a YO transcriptome database of the blackback land crab, *Gecarcinus lateralis*. The species is an important model for the study of the endocrine regulation molting by neuropeptides and other factors (Chang and Mykles, 2011). A pipeline for the validation, analysis, and functional assignment of contigs is described. Genes encoding signal transduction pathways were characterized. As expected, the genes in the cyclic nucleotide, insulin/mTOR, and TGF β /Smad pathways were well represented in the annotated database. Genes in the MAP kinase, calcium, VEGF, phosphatidylinositol, ErbB, Wnt, Hedgehog, Jak-STAT, and Notch pathways were also expressed. The diversity of signaling pathways raises the possibility that the YO can integrate endocrine, paracrine, and autocrine signals in order to respond appropriately to environmental and physiological conditions that affect molting.

2. Methods

2.1. Animals

Adult male *Gecarcinus lateralis* were collected in the Dominican Republic, shipped to Colorado State University, and maintained as described (Covi et al., 2010). YOs were dissected

from the branchial chamber side of the anterior branchiostegite region of the cephalothorax and stored in 300 µL RNAlater (Life Technologies, Grand Island, NY, USA) at -20 °C until processing. Hemolymph ecdysteroid titers were quantified using a competitive Enzyme Linked Immunosorbent Assay (ELISA) (Abuhagr et al., 2014a; Kingan, 1989).

2.2 mRNA isolation, library preparation, and sequencing

Total RNA was isolated using RNeasy™ Mini Kits (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. RNA was quantified using Quant-iT™ RiboGreen® RNA Assay (Life Technologies, Carlsbad, CA, USA). The mRNA purification and cDNA synthesis was carried out with a TruSeq™ Stranded mRNA Library Prep Kit (Illumina). Three cDNA libraries, designated Im1, Im2, and Im3, were generated; each library was derived from mRNA from six YOs pooled from three intermolt animals. Paired-end sequencing of the cDNA libraries using an Illumina HiSeq™ 2000 instrument was performed at the Oklahoma Medical Research Foundation. All samples were run in a single sequencing lane with three adaptor tags.

2.3. Transcriptome assembly and annotation

The quality of paired-end raw reads in fastq format was assessed using the FASTQC program (Babraham Institute, Cambridge, UK). Quality reads with a minimum phred (nucleotide base call) score of 28 and length ranging from 36 bp -100 bp were extracted by trimming of low quality reads and adapter sequences via Trimmomatic software (version number: 0.32) (Bolger et al., 2014). The trimmed reads obtained from three different biological YO replicates were concatenated into two files containing forward and reverse sequences, respectively. Further, both paired and unpaired reads were used for downstream analyses. The

trimmed forward and reverse reads were then assembled via Trinity software with default settings (version number: r20130814) (Haas et al., 2013). The minimum contig length was set at 201 bp. Following assembly, the contigs were clustered based on a 90% sequence similarity threshold using the CD-HIT-EST program (version number: 4.6.1) (Li and Godzik, 2006). The output of CD-HIT-EST was used as the reference transcriptome to map the reads from individual libraries. We designated the assembled data as the YO baseline transcriptome.

For annotation, both nucleotide sequences and predicted protein sequences were used to run BLAST queries against NCBI Non-Redundant (NR), Swissprot (SP), TrEMBL (Uniprot), and Uniprot Uniref90 protein databases (<http://www.uniprot.org/downloads>; Figure 1) (Altschul et al., 1990; Bairoch and Apweiler, 2000). The NR and SP databases were downloaded on April 1, 2015, TrEMBL on October 19, 2015, and Uniref90 on June 26, 2015. Stand-alone software was used for running BLAST (version: 2.2.28) against the above-mentioned databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download). The output of BLASTx against the NR database was generated as an xml format (outfmt 5) and was parsed with a python script developed by the Palumbi laboratory at Stanford University (<http://sfg.stanford.edu/scripts.html>) (De Wit et al., 2012). The results from BLASTx against SP and Uniprot databases were in tabular format (outfmt 6). The cut-off for e-value was set for $1e^{-5}$ and 10 hits were assigned per contig.

The functional analysis of the transcriptome via Gene Ontology (GO) was generated via Blast2GO Basic (version number 3.0.8) using the BLASTx against the NR database as input (Conesa et al., 2005). The GO numbers obtained from Blast2GO were entered into GOSlimViewer (http://www.agbase.msstate.edu/cgi-bin/tools/goslimviewer_select.pl) using the “Generic” GOSlim Set option to create a summary of the GO terms associated with annotated

191 transcripts in the YO baseline transcriptome (McCarthy et al., 2006). For the purpose of pathway
192 annotation, we used the output from BLASTx against the NR database for assignments to Kyoto
193 Encyclopedia for Genes and Genomes (KEGG) pathways and KEGG orthology (KO).
194 MEGAN V5.6.6 software (Huson and Mitra, 2012; Kanehisa and Goto,
195 2000) was used for generating KEGG annotations. We used the auxiliary mapping files for KO
196 and pathway assignments rather than the out-of-date built-in KEGG pathways ([http://ab.inf.uni-](http://ab.inf.uni-tuebingen.de/data/software/megan5/download/welcome.html)
197 [tuebingen.de/data/software/megan5/download/welcome.html](http://ab.inf.uni-tuebingen.de/data/software/megan5/download/welcome.html)). In order to compare the pathway
198 annotations of the land crab baseline transcriptome to the previously published YO transcriptome
199 of crayfish, we downloaded the assembled *Pontastacus leptodactylus* transcriptome (TSA:
200 GAFS000000000 from NCBI database (Tom et al., 2013). The crayfish transcriptome was
201 processed in the same manner as the land crab data to generate comparable KEGG pathway and
202 KO annotations.

203 TransDecoder (<https://transdecoder.github.io/>) was used to predict coding peptide sequences
204 from the baseline transcriptome contig sequences. These peptide sequences were annotated via
205 BLASTp against known databases with a cut-off of $1e^{-5}$. In order to identify conserved protein
206 families among the predicted peptide sequences, HMMER hmmscan
207 (<https://svn.janelia.org/eddylib/eddys/src/hmmer/trunk/documentation/man/hmmscan.man>) was
208 used to search for sequences against a Pfam-A database (downloaded on June 26, 2015)
209 (<http://hmmer.janelia.org/>). In addition, transmembrane helical domains and cleavage sites for
210 signal peptides were identified using TMHMM (version number: 2.0c) and SignalP (version
211 number: 4.1) software, respectively. All the outputs obtained from the above-mentioned
212 resources like BLAST, HMMER, TMHMM, and SignalP were used to achieve a comprehensive
213 annotation for each contig. For this purpose we used Trinotate (version number: r201407708) to

generate a flat file report containing all annotation information for each contig (Haas et al., 2013).

2.4. Transcript abundance estimation and statistical analyses

The relative transcript abundance in the three cDNA libraries was determined via the following steps. The good quality reads from all three intermolt replicates were mapped back to the YO baseline transcriptome using Bowtie (version number 2.0.3) to generate a SAM (Sequence Alignment/Map) file (Langmead and Salzberg, 2012). SAMtools software (version number: 0.1.18) was used convert the SAM files to BAM (binary form of SAM file) files (Li et al., 2009). These BAM files were used to quantify abundances of the mapped transcripts for individual libraries using eXpress software (version number: 1.5.1) (Roberts and Pachter, 2013). Expression levels are defined in fragments per kilobase per million reads (FPKM). Further data manipulation and statistical analyses (correlation coefficients) were performed using R statistical software (R-Development-Core-Team, 2015).

2.5. Computational resources

The assembly of the baseline transcriptome was performed at DIAG (Data Intensive Academic Grid) via Secure Shell (SSH) (<http://diagcomputing.org>). Additional computation jobs (trimming, mapping, annotation, and abundance estimation) were performed at the Oklahoma State University High Performance Computing Center using SSH (<http://hpc.it.okstate.edu/>).

3. Results

3.1. Assembly and analysis of the baseline transcriptome

The YO baseline biological replicates (designated Im1, Im2, and Im3) were selected from intermolt (stage C₄) animals using standard molt staging criteria (Covi et al., 2010). Each replicate consisted of YOs from 3 animals. The mean hemolymph ecdysteroid titers were 0.91 ± 0.02 pg/μl for Im1, 1.32 ± 0.01 pg/μl for Im2, and 1.41 ± 0.02 pg/μl for Im3. Reads from Illumina sequencing were assembled to generate the YO baseline transcriptome (Fig. 1). Table 1 compares the outcomes of the sequence analysis for each library. A total of 227,811,829 reads were generated. Following trimming, 90% of the reads were retained for further analysis. The general features of the baseline transcriptome library are summarized in Table 2. The good quality reads were assembled via the Trinity program (Haas et al., 2013). This generated 288,673 contigs with a mean length of 871 bp and median length (N50) of 1,842 bp. The length distribution of the transcriptome indicated that 40% of the contigs ranged between 201 and 299 bp in length and 30% of all contigs were greater than 500 bp (Appendix Fig. A.1). By comparison, 60% of the contigs that were returned with a BLASTx hit were greater than 500 bp (Appendix Fig. A.2). The median length of the protein coding sequences was greater (740 bp) than the median length of the entire transcriptome (227 bp), suggesting the presence of small non-coding RNAs and/or assembly artifacts. The distribution of contig length from this study is comparable to published *de novo* transcriptomes (Ghaffari et al., 2014; Lenz et al., 2014).

In order to reduce the number of redundant sequences, the CD-HIT-EST program was used to cluster similar sequences at a threshold set at 90% nucleotide similarity (Li and Godzik, 2006). The output from the CD-HIT-EST program (231,579 contigs) constituted the YO baseline transcriptome. Using Bowtie, reads from individual libraries were mapped back to the baseline transcriptome. Read alignment was similar among the three replicates; about 92% to 93% of the reads mapped back to the transcriptome (Table1).

As there is no reference genome for *G. lateralis*, the fidelity of the assembly was assessed by comparing the sequences of contigs with those of cDNAs obtained by Sanger sequencing. DNA and encoded peptide sequences were obtained from the NCBI nucleotide and protein databases, respectively, for 13 cDNAs that were previously identified in *G. lateralis* tissues by RT-PCR cloning. Nucleotide identities were 97% to 100% between sequences obtained via Sanger and Illumina sequencing (Table 3). Over half of the 13 sequences had greater than 99% identity in protein alignment. The small discrepancy of 1% to 3% between the nucleotide and protein sequences was attributed to differences between the two technologies, the tissue sources used, novel isoforms, and sequencing errors (Durica et al., 2014). These comparisons confirmed the accuracy of the sequence generation and assembly protocols, validating the use of this data for further analysis.

3.2. Transcriptome annotation

The baseline transcriptome was annotated using multiple methods as shown in the annotation pipeline (Fig. 2). All contigs (231,597) were annotated via a BLASTx search against the NCBI non-redundant (NR), UniProtKB/Swiss-Prot (SP), and UniProtKB/TrEMBL (<http://www.uniprot.org/downloads>) databases. This resulted in 34,605, 23,504, and 25,723 significant hits, respectively, using $1e^{-5}$ as the e-value cutoff (Table 4). Each BLAST output generated top ten hits per annotated contig. About 20% of the contigs were identified. The relatively low hit rate was attributed to the lack of an annotated decapod crustacean genome and the restriction of the analysis to protein coding genes, which constitute less than 1% of the total RNA.

One problem associated with BLASTx analysis against the NR database was that the output for many sequences was “hypothetical protein” or “predicted protein”. In order to increase the identification of contigs with known genes, the BLASTx output was analyzed against the SP and TrEMBL databases (Fig. 2). Open reading frames (ORFs) were generated from the YO baseline transcriptome using TransDecoder. This resulted in 81,481 predicted peptides, of which 38,151 (47%) were complete (containing both start and stop codons); 19,475 (24%) were five-prime partial (containing a start codon); 6,933 (9%) were three-prime partial (ending with a stop codon); and 16,922 (21%) were internal (lacking both the start and stop codons). These ORFs were annotated via BLASTp against the NR, SP, and Uniprot-UniRef90 databases. These BLAST analyses resulted in 24,850, 19,707, and 24,817 predicted peptides/ORFs with significant hits, respectively (Fig. 2). In addition to the BLAST analyses, the ORF sequences were analyzed to identify predicted Pfam domains (HMMER/Pfam), trans-membrane helices (TMHMM), and signal peptide cleavage sites (SignalP). All the outputs from BLASTx, BLASTp, HMMER/Pfam, TMHMM, and SignalP were combined into a single annotation file using Trinotate (Fig. 2). The result was 18,772 contigs with predicted Pfam domains, 20,706 contigs with predicted transmembrane helices (1 to 16 transmembrane helices per contig), 5,745 contigs with a signal peptide cleavage site, and 9,566 contigs with GO annotation from Pfam-A hits (Table 4).

3.3. Functional annotation - Blast2GO

Blast2GO was used to perform functional annotation for the YO baseline transcriptome via Gene Ontology (GO) (Ashburner et al., 2000; Götz et al., 2008). Of the 34,605 BLASTx hits, 20,208 unique contigs were assigned to at least one of the three parent GO categories: Molecular

Function (MF), Cellular Component (CC), and Biological Process (BP). A Venn diagram illustrates the distribution of unique contigs annotated to one or combination of the three GO categories (Fig. 3). Many of the contigs were assigned to two or all three categories. For example, there was considerable overlap of contigs assigned to the CC and BP categories, with only 674 contigs assigned only to CC and 749 contigs assigned only to BP. One-third (6,718) of the contigs were assigned to all three categories. The overall distribution was 17,841 contigs (51%) in MF, 9,021 contigs (26%) in CC, and 14,266 contigs (41%) in BP.

A second level GO analysis assigned the contigs to 41, 33, and 71 GO Slims terms in the MF, CC, and BP categories, respectively (<http://geneontology.org/page/go-slim-and-subset-guide>). Within the MF category, the most represented GO Slims were ion binding (GO:0043167) and oxidoreductase activity (GO:0016491) (Fig. 4A). The most represented GO Slims were cell (GO:0005623) and intracellular (GO:0005622) in the CC category (Fig. 4B) and cellular nitrogen compound metabolic process (GO:0034641) and biosynthetic process (GO:0009058) in the BP category (Fig. 4C). Of the 20,208 GO assignments within the baseline transcriptome, 7,272 were assigned to unique GO terms; the distribution was 2,086 (29%) in MF, 814 (11%) in CC, and 4,372 (60%) in BP.

3.4. Pathway annotation - KEGG and comparative analysis of YO transcriptomes

The pathway annotation for the YO baseline transcriptome was conducted via KEGG. KEGG orthology (KO) assignment and pathway annotation was performed on the data from BLASTx hits (34,605 contigs) against the NR database; these are the same data used for the GO analysis (Fig. 2). The *G. lateralis* KO assignments were compared to those from a YO transcriptome generated from the crayfish, *Pontastacus leptodactylus* (TSA: GAFS000000000)

(Tom et al., 2013). The KO assignment analysis resulted in annotation of 241 and 232 pathways in land crab and crayfish, respectively. The total number of *G. lateralis* contig hits that were assigned to one or more of the five global KEGG pathways (Metabolism Genetic, Information Processing, Organismal Systems, Cellular Processes, and Environmental Information Processing) was 15,090, of which 6,222 had no specific pathway assignments (Fig. 5A). By comparison, 6,805 crayfish BLASTx hits were assigned to KEGG pathways, of which 3,099 were unclassified (Fig. 5B). The unclassified contigs in both land crab and crayfish were assigned with a KO number. However, these contigs were not incorporated into any particular KEGG pathway. As the KEGG pathways are manually curated, it is possible that these contigs are part of pathways that have not been assigned in the database. Although there were fewer numbers of assignments for crayfish compared to land crab, the percentage of BLASTx hits distributed among KEGG global pathways was similar in both species (Fig. 5).

The KEGG pathways are gene interaction pathways displayed in a hierarchical manner (Kanehisa et al., 2014). The *G. lateralis* and *P. leptodactylus* data were further assigned to the KEGG second and third tier in the pathway hierarchy. In land crab, the Metabolism global pathway constituted 32% of the KEGG assignments (Fig. 6A). The Signal Transduction pathway had the largest number of assignments of the second-tier pathways, followed by Translation, Carbohydrate Metabolism, and Transport and Catabolism pathways (Fig. 6A). The assignments of the crayfish contigs showed similar trends, with Signal Transduction and Translation pathways ranked first and second, respectively (Fig. 6A). Transport and Catabolism pathways were ranked third and Folding, Sorting, and Degradation pathway was ranked fourth in terms of number contigs assigned to KO (Fig. 6A).

Contigs assigned to the Signal Transduction second-tier hierarchy were examined further. There are a total of 28 gene interaction pathways in the Signal Transduction third-tier hierarchy (<http://www.genome.jp/kegg/pathway.html>). The 1,079 contigs in Signal Transduction were assigned to 16 signaling pathways; the 12 pathways with the largest number of contigs are shown in Figure 6B for both *G. lateralis* and *P. leptodactylus*. Pathway rankings based on number of KEGG assignments, were similar for both species; PI3K-Akt, calcium, MAPK, and Wnt signaling pathways were ranked first to fourth (Fig. 6B). The components of the mTOR and TGF β pathways were well represented in the transcriptome. Interaction maps for mTOR and TGF β pathways included comparisons of the KO assigned in land crab and crayfish (Fig. 7). All but one of the crayfish KO assignments (K04657; *chordin* or *short gastrulation*) was present in land crab (Fig. 7B, Appendix Table A.1 and B.1). However, although a *chordin/sog* assignment was not observed in the KEGG analysis, an ortholog for *chordin/sog* (contig: comp117728_c0_seq1; BLASTx output: gi|308197284|gb|ADO17754.1| short gastrulation protein [*Parhyale hawaiiensis*]; e-value = 0; score = 1817) was detected in the land crab transcriptome database. Seven KEGG identifiers not assigned in the crayfish YO transcriptome were detected in the land crab YO transcriptome; four in the mTOR (K03259, K07209, K04688, K07298) and three in the TGF β ((K04661, K03347, K03456) signaling pathways (Fig. 7 and Appendix Table A.1 and B.1). Additionally, the main components of the Notch, Hedgehog, Wnt, and MAPK signaling pathways were identified by KEGG assignments (Table 5).

3.5. Correlation of expression levels among replicates

The YOs used for the baseline transcriptomes were from intermolt (stage C₄) animals. Although the libraries were from the same molt stage, the validity of the biological replicates

was evaluated by comparing the digital expression levels of contigs in the three libraries using both count and FPKM values. Count data is used for downstream differential gene expression analysis using Bioconductor packages, such as DESeq and EdgeR (<http://www.bioconductor.org/packages/release/bioc/>) (Anders and Huber, 2010; Robinson et al., 2010). FPKM is a normalized value of relative transcript abundance that corresponds to the RNA concentration at the time of tissue collection. Over one-half (122,485 or 53%) of all the contigs had an FPKM value of zero in at least one library, suggesting that either (1) there was no expression of that particular transcript or (2) the transcript sequence was an artifact of *de novo* assembly. Pearson coefficients were calculated in a pairwise statistical analysis of count and FPKM values (Table 6). Both count and FPKM values were positively correlated ($P = 0.94$ to 0.99), which indicates the three libraries can be used as biological replicates to represent YO gene expression in intermolt animals.

4. Discussion

High throughput sequencing technologies has allowed “omics” studies to be more cost-effective for non-model organisms (Abehsera et al., 2015; Christie, 2014; Groh et al., 2014; Jung et al., 2011; Tom et al., 2014; Tom et al., 2013; Ventura et al., 2014). Using similar strategies, a *de novo* transcriptome was assembled to catalog genes that are expressed in the *G. lateralis* YO. An objective was to generate a comprehensive database of genes that are well annotated and could serve as a reference transcriptome for decapod crustaceans. The data built from Illumina technology ensures fidelity during sequencing, as indicated by the high sequence identities between the contigs and the same genes sequenced using Sanger technology (Table 3). Although only 20% of the contigs from the YO transcriptome were annotated, over 34K protein-coding sequences were identified. The number of annotated contigs in the *G. lateralis* YO transcriptome

is greater than other decapod crustacean transcriptomes (Li et al., 2015; Lv et al., 2014; Suwansa-Ard et al., 2015). This indicates that there is a need for using a standard method that incorporates all available tools for annotating *de novo*-assembled transcriptomes in species without genome sequence information. In addition to BLASTx and BLASTp, HMMER, TMHMM, and SignalP were used to identify protein family domains, transmembrane helices, and signal peptide sequences, respectively, in the *G. lateralis* YO transcriptome. Trinotate software was used to combine the analyses to generate a comprehensive annotation profile for each contig. This combined method provided a detailed analysis of the YO transcriptome that did not rely solely on BLAST.

Comparison of the land crab and crayfish YO KEGG pathway annotations revealed that both transcriptomes have similar gene expression profiles. However, certain contigs were not annotated via KEGG, even with the presence of a BLASTx hit assigned to the particular contig. These discrepancies occurred in both crayfish and land crab transcriptomes and can be overcome by pathway annotation comparison, followed by manually curating of the orthologs identified via BLAST into a specific pathway. The high similarity of percentage of KO assignments across global KEGG pathways between land crab and crayfish indicates that the YO transcriptomes share a common expressed-gene profile. This further validates the application of RNA-Seq technology, as well as the utilization of proper annotation methods, for the identification of putative orthologs in a non-model organism. Thus, the *G. lateralis* YO transcriptome database is a useful resource to identify and catalog gene networks controlling YO ecdysteroidogenesis during the molt cycle.

Activation of the molting gland, which is under the control of neuropeptide hormones, drives the transition from intermolt to premolt. In insects, the PG is activated by PTTH released

420 from the brain. In decapod crustaceans, the YO is activated by a decrease in MIH release from
421 the XO/SG complex. The signaling pathways regulating the molting glands by neuropeptides
422 differ between crustaceans and insects, most likely necessitated by the opposite actions of MIH
423 and PTTH, respectively, on ecdysteroidogenesis. PTTH binding to Torso, a receptor tyrosine
424 kinase, triggers an influx of Ca^{2+} , resulting in activation of cAMP and/or MAP kinase pathways,
425 depending on species (Covi et al., 2012; Yamanaka et al., 2013). By contrast, MIH signaling
426 involves a cAMP/ Ca^{2+} -dependent triggering phase and a NO/cGMP-dependent summation phase
427 linked by calmodulin (CaM) and calcineurin (CaN), a CaM-dependent protein phosphatase
428 (Chang and Mykles, 2011; Covi et al., 2012; Webster, 2015). Activation of NO synthase (NOS)
429 through the combined actions of Ca^{2+} /CaM binding and CaN-dependent dephosphorylation
430 increases the production of NO from L-arginine. NO activates an NO-dependent guanylyl
431 cyclase (GC-I), resulting in a prolonged increase in cGMP that represses YO
432 ecdysteroidogenesis by cGMP-dependent protein kinase (PKG) (Chang and Mykles, 2011;
433 Mykles et al., 2010; Webster, 2015). All the components of MIH signaling are present in the YO
434 baseline transcriptome. Calcium signaling pathway genes were present in the transcriptome (Fig.
435 6B). In addition to cDNAs encoding NOS and GC-I β obtained by RT-PCR cloning (Kim et al.,
436 2004; Lee et al., 2007), transcripts encoding adenylyl cyclase, protein kinase A, CaM, CaN, and
437 PKG were identified (Table 7). The MIH receptor has yet to be characterized (Webster, 2015).
438 The annotated transcriptome can be used to identify MIH receptor candidates, as well as
439 characterize signaling pathway components, such as cyclic nucleotide phosphodiesterases
440 (PDEs), that contribute to the changes in MIH sensitivity over the molt cycle (Nakatsuji and
441 Sonobe, 2004; Nakatsuji et al., 2006).

Hypertrophy of the activated molting gland increases ecdysteroid synthetic capacity in premolt. Despite the diametrically opposite control by neuropeptides between insects and crustaceans, both YO and PG activation involve mTOR-dependent protein synthesis. Rapamycin inhibits YO ecdysteroidogenesis *in vitro* (Abuhagr et al., 2014b). Contigs encoding mTOR signaling pathway genes are well represented in the land crab and crayfish YO transcriptomes (Figs. 6B, 7A). mRNA levels of *Gl-mTOR* and *Gl-Akt* in the land crab YO are increased during premolt (Abuhagr et al., 2014b). Genetic manipulation of mTOR signaling affects PG ecdysteroidogenesis in *Drosophila*. Overexpression of PI3 kinase stimulates PG growth, whereas overexpression of Rheb-GTPase activating protein inhibits PG growth (Colombani et al., 2005; Layalle et al., 2008; Mirth et al., 2005). In *Bombyx mori*, inhibitors of PI3 kinase and mTOR (rapamycin) block PTTH-stimulated PG ecdysteroid secretion (Gu et al., 2012; Gu et al., 2011), whereas in *Manduca sexta*, PI3 kinase inhibitors have no effect on PTTH-stimulated ecdysteroid secretion (Smith et al., 2014). In *Drosophila* and both lepidopteran species, insulin and ILPs stimulate mTOR-mediated PG growth and ecdysteroidogenesis (Gu et al., 2015; Gu et al., 2011; Hatem et al., 2015; Smith et al., 2014; Teleman, 2010; Walkiewicz and Stern, 2009). These data show that mTOR signaling mediates the activation of the arthropod molting gland, either in response to release of tropic peptide hormones in insects (e.g., PTTH and ILP) or a reduction of inhibitory peptide hormones in crustaceans (e.g., MIH and CHH). The large number of contigs assigned to the PI3K-Akt signaling pathway (Fig. 6B) suggests that the YO may respond to growth factors as well.

Activation of TGF β /Smad signaling is necessary for ecdysteroidogenesis in molting glands, but its role differs between insects and crustaceans. In insects, TGF β /Smad signaling renders the PG competent to respond to PTTH and ILPs (Rewitz et al., 2013). In *Drosophila*, loss of Activin

465 signaling by knockdown of Type I and II receptors (Babo and Punt, respectively), Co-Smad
466 (Medea), or R-Smad (dSmad2) blocks the PTTH-induced ecdysteroid peak that triggers
467 metamorphosis (Gibbens et al., 2011). Conversely, expression of Activin (Actb) or constitutively
468 active Babo causes precocious pupariation (Gibbens et al., 2011). Knockdown of dSmad2
469 decreases the mRNA levels of PTTH receptor (Torso), insulin receptor, and Halloween genes,
470 supporting the role of TGF- β /Smad signaling in PTTH-dependent stimulation of PG
471 ecdysteroidogenesis (Gibbens et al., 2011). In crustaceans, TGF β /Smad signaling appears to be
472 necessary for YO commitment, which occurs at the transition from early premolt (stage D₀) to
473 the mid-premolt (stage D₁) (Chang and Mykles, 2011). Contigs encoding Activin/Smad signaling
474 pathway genes are well represented in the land crab and crayfish YO transcriptomes (Fig. 6B),
475 including a myostatin-like factor related to Activin (Table A.2). SB431542, an antagonist of the
476 Activin RII receptor, prevents the transition from the activated to the committed state in
477 eyestalk-ablated animals (Abuhagr et al., 2012). Taken together, the data suggest that Activin
478 regulates the sensitivity of the arthropod molting gland to neuropeptides. The differences
479 between the two arthropod groups are consistent with their opposing actions. PTTH is a tropic
480 factor, while MIH is a static factor. In insects, Activin/Smad signaling prevents precocious
481 molting, as it assures that the PG is not activated by PTTH and ILPs until an animal reaches its
482 critical weight (Rewitz et al., 2013). In decapod crustaceans, Activin/Smad signaling appears to
483 be necessary for the transition of the YO to the committed state. The consequent insensitivity to
484 inhibitory neuropeptides (MIH and CHH) and perhaps other factors assures that molting
485 processes, such as exoskeleton synthesis and limb regenerate growth, proceed without
486 interruption (Chang and Mykles, 2011; Covi et al., 2012). In summary, an Activin-like TGF β

487 factor drives critical transitions in the molting cycle. In insects, it makes the PG sensitive to
488 tropic peptides, while in decapods it makes the YO insensitive to static peptides.

489 Other signaling pathways represented in the transcriptome may regulate YO function. The
490 MAP kinase pathway, via ERK phosphorylation, may be involved in inhibition of YO
491 ecdysteroid production during postmolt in the crab, *Scylla serrata* (Imayavaramban et al., 2007).
492 Activation of protein kinase C, which is a target of the phosphatidylinositol pathway, stimulates
493 YO ecdysteroidogenesis in the rock crab, *Cancer antennarius*, but inhibits ecdysteroidogenesis
494 in the crayfish, *Orconectes sp.* (Mattson and Spaziani, 1987; Spaziani et al., 2001). Calcium
495 signaling is complex, as it can interact with other pathways to control ecdysteroid synthesis and
496 secretion. Manipulation of intracellular Ca^{2+} concentration with pharmacological reagents yields
497 inconsistent results, suggesting that sustained ecdysteroidogenesis requires Ca^{2+} concentrations
498 within a limited range (Spaziani et al., 2001). YO ecdysteroid secretion increases with increasing
499 extracellular Ca^{2+} concentrations (Spaziani et al., 2001). Moreover, intracellular Ca^{2+} is
500 correlated with YO ecdysteroidogenesis in the blue crab, *Callinectes sapidus* (Chen et al., 2012).
501 It is thought that Ca^{2+} action is mediated by CaM-dependent PDEs, thus countering the effects of
502 MIH by keeping cyclic nucleotide levels low, and/or by activating PKC (Nakatsuji et al., 2009).
503 Transcriptomics using RNA-Seq technology makes it possible to explore the interactions
504 between cyclic nucleotide and Ca^{2+} second messenger pathways that control YO
505 ecdysteroidogenesis.

506 There are few studies of the Wnt, Hedgehog, and Notch signal transduction pathways in the
507 arthropod molting gland. Contigs encoding Wnt, Hedgehog, and Notch signaling genes are
508 reported in the land crab and crayfish YO transcriptomes (Fig. 6B and Table 4), as well as in the
509 transcriptomes from other decapod tissues (Hao et al., 2014; Song et al., 2015; Ventura et al.,

2013; Wei et al., 2014a). To our knowledge, these pathways have not been studied in the YO. The insect PG expresses Wnt and Hedgehog signaling genes, and genetic studies in *Drosophila* suggest these pathways have a role in PG function. Hedgehog signaling inhibits ecdysteroid biosynthesis in the PG (Rodenfels et al., 2014). Wnt/ β -catenin signaling controls gene expression in the PG via Wnt-dependent cis-regulatory modules (Archbold et al., 2014). In mammals, Hippo, Wnt, and Notch signaling pathways activate mTORC1 (Shimobayashi and Hall, 2014). These data suggest that mTOR-dependent ecdysteroidogenesis by the arthropod molting gland is controlled by multiple signals.

In summary, a comprehensive YO transcriptome was assembled and annotated from deep sequencing with an Illumina platform. Analyses indicated high fidelity of *de novo* assembled contigs with *G. lateralis* cDNA sequences in GenBank. The database is available on iPlant and is intended to serve as a resource for the identification of gene networks controlling YO function. A significant finding was the similarity in the representation of numerous signal transduction pathways characterized in the YO transcriptomes of land crab and crayfish (Fig. 6B). Signal transduction genes constituted the largest second-tier hierarchical category of KEGG assignments (Fig. 6A). The presence of genes in the mTOR, TGF β , MAPK, calcium, and phosphatidylinositol signal transduction pathways are consistent with physiological studies (Chang and Mykles, 2011; Covi et al., 2009; Nakatsuji et al., 2009; Spaziani et al., 2001; Webster et al., 2012). However, the discovery of genes in the Wnt, VEGF, ErbB, Hedgehog, Jak-STAT, and Notch pathways raises the possibility that YO regulation is more complex. As the YO initiates and sustains molting processes, it is reasonable to hypothesize that the YO is able to integrate a variety of extrinsic and intrinsic signals to effect an appropriate response. The decision to initiate molting is controlled by MIH, but molting can be interrupted or delayed by

533 adverse conditions in early premolt. The decision to complete molting is made when the animal
534 transitions to mid premolt, at which point molting processes continue and the animal molts
535 without delay. The transition of the YO from the activated to committed state appears to require
536 Activin signaling. Taken together, the YO may be more like the insect PG, which is controlled
537 by a variety of tropic and static factors (Covi et al., 2012; Marchal et al., 2010; Rewitz et al.,
538 2013; Yamanaka et al., 2013). mTOR and Activin are potential targets of these factors. One goal
539 is to use transcriptomic tools to identify genes that drive the transitions in the physiological states
540 of the YO through the molt cycle.

541

542 **Availability of supporting data**

543 The assembled baseline transcriptome is available for researchers to download from the

544 iPlant CollaborativeTM (<http://www.iplantcollaborative.org/>). The public sharing link is:

545 <http://de.iplantcollaborative.org/dl/d/9B72D2F1-3F85-43FD-9AB1->

546 [B80CE84F1108/Gecarcinus_lateralis_YO_Baseline_Transcriptome.fasta.zip](http://de.iplantcollaborative.org/dl/d/9B72D2F1-3F85-43FD-9AB1-B80CE84F1108/Gecarcinus_lateralis_YO_Baseline_Transcriptome.fasta.zip).

547

548

549 List of abbreviations

550	APC	Adenomatosis polyposis coli protein
551	BAM	Binary form of a SAM file
552	BLAST	Basic Local Alignment Search Tool
553	BP	Biological Process
554	CaM	Calmodulin
555	cAMP	Cyclic Adenosine Monophosphate
556	CaN	Calcineurin
557	CC	Cellular Component
558	CD-HIT-EST	Cluster Database at High Identity with Tolerance
559	Ci	Cubitus Interruptus
560	CK II	Casein Kinase II
561	cGMP	Cyclic Gunaosine Phosphate
562	CHH	Crustacean Hyperglycemic Hormone
563	CREB	Cyclic AMP-Dependent Transcription Factor ATF-4
564	DIAG	Data Intensive Academic Grid
565	Dvl	Dishevelled (segment polarity protein)
566	ELISA	Enzyme Linked Immunosorbent Assay
567	EGF	Epidermal Growth Factor
568	EGFR	Epidermal Growth Factor Receptor
569	ERK	Extracellular Regulated Kinase
570	FGFR	Fibroblast Growth Factor Receptor 2
571	FPKM	Fragments per kilobase per millions reads
572	GO	Gene Ontology
573	GRB	Growth Factor Receptor-Binding Protein 2
574	GSK	Glycogen Synthasee Kinase 2 beta
575	GC-1β	Guanylyl Cyclase Beta
576	HMMER	Biosequence analysis using profile hidden Markov models
577	KEGG	Kyoto Encyclopedia of Gene and Genomes
578	KO	KEGG Orthology
579	ILP	Insulin Like Peptides
580	Im1-3	YO baseline biological replicates
581	JAK-STAT	Janus Kinase-Signal Transducer and Activator of Transcription
582	MAPK	Mitogen Activated Protein Kinases
583	MAP	Mitogen Activated Protein
584	MEGAN	MEtaGenome Analyzer
585	MEK1	Mitogen-Activated Protein Kinase Kinase I
586	MF	Molecular Function
587	MIH	Molt Inhibiting Hormone
588	MNK1/2	MAP Kinase Interacting Serine/Threonine Kinase
589	mTOR	Mechanistic Target of Rapamycin
590	NCBI	National Center for Biotechnology Information
591	NO	Nitric oxide
592	NOS	Nitric Oxide Synthase
593	NR	Non-Redundant (protein database)

594	ORF	Open Reading Frame
595	PCR	Polymerase Chain Reaction
596	PDE	Phosphodiesterase
597	Pfam	Protein family database
598	PG	Prothoracic Gland
599	PKC	Protein Kinase C
600	PKG	Protein Kinase G
601	PTTH	Prothoracic Hormone
602	RBPSUH	Recombining Binding Protein Suppressor of Hairless
603	Rheb	Ras Homolog Enriched in Brain
604	RMBT	Reads Mapped Back to Transcriptome
605	RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
606	SAM	Sequence Alignment Map
607	SHH	Secure Shell
608	SOS	Son of Sevenless
609	SP	SwissPort
610	SRF	Serum Response Factor
611	Su(fu)	Suppressor of Fused
612	TF 7	Transcription Factor 7
613	TGFβ	Transforming Growth Factor Beta
614	TMH	Transmembrane Helices
615	TMHMM	Transmembrane helices Markov model
616	TSA	Transcriptome Shotgun Assembly
617	Wnt	Wingless
618	XO/SG	X-organ/Sinus Gland
619	VEGF	Vascular Endothelial Growth Factor
620	YO	Y-organ
621		

622

623

Acknowledgements

We thank Dr. Ernest S. Chang and Sharon A. Chang, UC Davis Bodega Marine Laboratory, for ELISA on hemolymph samples; Hector C. Horta for collecting *G. lateralis*; and Graham Wiley, Oklahoma Medical Research Facility, for Illumina sequencing. We acknowledge the assistance of Richard Casey and Joseph Allison, Colorado State University, for generating scripts to assemble and annotate the RNA-Seq data and Jesse Schafer, Oklahoma State University High Performance Computing Center, for technical support. We also thank undergraduate students Stephanie Oatman and Matthew Donovan for analyzing data from Illumina and Sanger sequencing and Clayton Hallman, Colorado State University, for his assistance with statistical analysis of the transcriptome data using R. We thank the staff of Consejo Dominicano de Pesca y Acuicultura, Dominican Republic, for expediting approval of permits and identifying suitable sites for collecting *G. lateralis*. This research was supported by the National Science Foundation (IOS-1257732).

References

- Abehsera, S., Glazer, L., Tynyakov, J., Plaschkes, I., Chalifa-Caspi, V., Khalaila, I., Aflalo, E.D., Sagi, A., 2015. Binary gene expression patterning of the molt cycle: The case of chitin metabolism Plos One 10, e0122602.
- Abuhagr, A.M., Blindert, J.L., Nimitkul, S., Zander, I.A., LaBere, S.M., Chang, S.A., MacLea, K.S., Chang, E.S., Mykles, D.L., 2014a. Molt regulation in green and red color morphs of the crab *Carcinus maenas*: Gene expression of molt-inhibiting hormone signaling components. J. Exp. Biol. 217, 796-808.
- Abuhagr, A.M., Chang, E.S., Mykles, D.L., 2012. Role of mTOR and TGF beta in Y-organ activation during the crustacean molting cycle. Integr. Comp. Biol. 52, E202.
- Abuhagr, A.M., MacLea, K.S., Chang, E.S., Mykles, D.L., 2014b. Mechanistic target of rapamycin (mTOR) signaling genes in decapod crustaceans: Cloning and tissue expression of mTOR, Akt, Rheb, and p70 S6 kinase in the green crab, *Carcinus maenas*, and blackback land crab, *Gecarcinus lateralis*. Comp. Biochem. Physiol. 168A, 25-39.
- Albert, V., Hall, M.N., 2015. Reduced C/EBP beta-LIP translation improves metabolic health. EMBO Rep. 16, 881-882.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Molec. Biol. 215, 403-410.
- Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11.
- Archbold, H.C., Broussard, C., Chang, M.V., Cadigan, K.M., 2014. Bipartite recognition of DNA by TCF/pangolin is remarkably flexible and contributes to transcriptional responsiveness and tissue specificity of wingless signaling. Plos Genetics 10.

661 Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P.,
 662 Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L.,
 663 Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M.,
 664 Sherlock, G., Gene Ontology, C., 2000. Gene Ontology: Tool for the unification of
 665 biology. *Nat. Genet.* 25, 25-29.

666 Bairoch, A., Apweiler, R., 2000. The SWISS-PROT protein sequence database and its
 667 supplement TrEMBL in 2000. *Nucleic Acids Res.* 28, 45-48.

668 Baretic, D., Williams, R.L., 2014. The structural basis for mTOR function. *Sem. Cell. Dev. Biol.*
 669 36, 91-101.

670 Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh,
 671 S., Sorensen, I., Lichtenstein, G., Fich, E.A., Conte, M., Keller, H., Schneeberger, K.,
 672 Schwacke, R., Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., Aflitos, S.A., Schijlen, E.,
 673 Jimenez-Gomez, J.M., Ryngajllo, M., Kimura, S., Kumar, R., Koenig, D., Headland,
 674 L.R., Maloof, J.N., Sinha, N., van Ham, R.C.H.J., Lankhorst, R.K., Mao, L., Vogel, A.,
 675 Arsova, B., Panstruga, R., Fei, Z., Rose, J.K.C., Zamir, D., Carrari, F.F., Giovannoni, J.J.,
 676 Weigel, D., Usadel, B., Fernie, A.R., 2014. The genome of the stress-tolerant wild tomato
 677 species *Solanum pennellii*. *Nat. Genet.* 46, 1034-1038.

678 Cetrullo, S., D'Adamo, S., Tantini, B., Borzi, R.M., Flamigni, F., 2015. mTOR, AMPK, and
 679 Sirt1: Key players in metabolic stress management. *Crit. Rev. Eukaryot. Gene Expr.* 25,
 680 59-75.

681 Chandler, J.C., Aizen, J., Elizur, A., Hollander-Cohen, L., Battaglione, S.C., Ventura, T., 2015.
 682 Discovery of a novel insulin-like peptide and insulin binding proteins in the Eastern rock
 683 lobster *Sagmariasus verreauxi*. *Gen. Comp. Endocrinol.* 215, 76-87.

684 Chang, E.S., Mykles, D.L., 2011. Regulation of crustacean molting: A review and our
685 perspectives. *Gen. Comp. Endocrinol.* 172, 323-330.

686 Chen, H.-Y., Dillaman, R.M., Roer, R.D., Watson, R.D., 2012. Stage-specific changes in
687 calcium concentration in crustacean (*Callinectes sapidus*) Y-organs during a natural
688 molting cycle, and their relation to the hemolymphatic ecdysteroid titer. *Comp. Biochem.*
689 *Physiol.* 163A, 170-173.

690 Christiaens, O., Delbare, D., Van Neste, C., Cappelle, K., Yu, N., De Wilde, R., Van
691 Nieuwerburgh, F., Deforce, D., Cooreman, K., Smagghe, G., 2015. Differential
692 transcriptome analysis of the common shrimp *Crangon crangon*: Special focus on the
693 nuclear receptors and RNAi-related genes. *Gen. Comp. Endocrinol.* 212, 163-177.

694 Christie, A.E., 2014. *In silico* characterization of the peptidome of the sea louse *Caligus*
695 *rogercresseyi* (Crustacea, Copepoda). *Gen. Comp. Endocrinol.* 204, 248-260.

696 Chung, J.S., Zmora, N., Katayama, H., Tsutsui, N., 2010. Crustacean hyperglycemic hormone
697 (CHH) neuropeptides family: Functions, titer, and binding to target tissues. *Gen. Comp.*
698 *Endocrinol.* 166, 447-454.

699 Colombani, J., Bianchini, L., Layalle, S., Pondeville, E., Dauphin-Villemant, C., Antoniewski,
700 C., Carre, C., Noselli, S., Leopold, P., 2005. Antagonistic actions of ecdysone and
701 insulins determine final size in *Drosophila*. *Science* 310, 667-670.

702 Conesa, A., Götz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: A
703 universal tool for annotation, visualization and analysis in functional genomics research.
704 *Bioinformatics* 21, 3674-3676.

705 Covi, J.A., Bader, B.D., Chang, E.S., Mykles, D.L., 2010. Molt cycle regulation of protein
706 synthesis in skeletal muscle of the blackback land crab, *Gecarcinus lateralis*, and the

707 differential expression of a myostatin-like factor during atrophy induced by molting or
 708 unweighting. J. Exp. Biol. 213, 172-183.

709 Covi, J.A., Chang, E.S., Mykles, D.L., 2009. Conserved role of cyclic nucleotides in the
 710 regulation of ecdysteroidogenesis by the crustacean molting gland. Comp. Biochem.
 711 Physiol. 152A, 470-477.

712 Covi, J.A., Chang, E.S., Mykles, D.L., 2012. Neuropeptide signaling mechanisms in crustacean
 713 and insect molting glands. Invert. Reprod. Dev. 56, 33-49.

714 De Wit, P., Pespeni, M.H., Ladner, J.T., Barshis, D.J., Seneca, F., Jaris, H., Therkildsen, N.O.,
 715 Morikawa, M., Palumbi, S.R., 2012. The simple fool's guide to population genomics via
 716 RNA-Seq: an introduction to high-throughput sequencing data analysis. Molec. Ecol.
 717 Resources 12, 1058-1067.

718 Durica, D.S., Das, S., Najar, F., Roe, B., Phillips, B., Kappalli, S., Anilkumar, G., 2014.
 719 Alternative splicing in the fiddler crab cognate ecdysteroid receptor: Variation in receptor
 720 isoform expression and DNA binding properties in response to hormone. Gen. Comp.
 721 Endocrinol. 206, 80-95.

722 Gao, J., Wang, X., Zou, Z., Jia, X., Wang, Y., Zhang, Z., 2014. Transcriptome analysis of the
 723 differences in gene expression between testis and ovary in green mud crab (*Scylla*
 724 *paramamosain*). BMC Genomics 15.

725 Ghaffari, N., Sanchez-Flores, A., Doan, R., Garcia-Orozco, K.D., Chen, P.L., Ochoa-Leyva, A.,
 726 Lopez-Zavala, A.A., Salvador Carrasco, J., Hong, C., Briebe, L.G., Rudino-Pinera, E.,
 727 Blood, P.D., Sawyer, J.E., Johnson, C.D., Dindot, S.V., Sotelo-Mundo, R.R., Criscitiello,
 728 M.F., 2014. Novel transcriptome assembly and improved annotation of the whiteleg

729 shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture.
 730 Scientific Rep. 4.

731 Gibbens, Y.Y., Warren, J.T., Gilbert, L.I., O'Connor, M.B., 2011. Neuroendocrine regulation of
 732 *Drosophila* metamorphosis requires TGF beta/Activin signaling. Development 138,
 733 2693-2703.

734 Götz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M.,
 735 Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data
 736 mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420-3435.

737 Groh, K.C., Vogel, H., Stensmyr, M.C., Grosse-Wilde, E., Hansson, B.S., 2014. The hermit
 738 crab's nose-antennal transcriptomics. Front Neurosci 7.

739 Gu, S.-H., Chen, C.-H., Hsieh, Y.-C., Lin, P.-L., Young, S.-C., 2015. Modulatory effects of
 740 bombyxin on ecdysteroidogenesis in *Bombyx mori* prothoracic glands. J. Insect Physiol.
 741 72, 61-69.

742 Gu, S.H., Yeh, W.L., Young, S.C., Lin, P.L., Li, S., 2012. TOR signaling is involved in PTTH-
 743 stimulated ecdysteroidogenesis by prothoracic glands in the silkworm, *Bombyx mori*.
 744 Insect Biochem. Molec. Biol. 42, 296-303.

745 Gu, S.H., Young, S.C., Lin, J.L., Lin, P.L., 2011. Involvement of PI3K/Akt signaling in PTTH-
 746 stimulated ecdysteroidogenesis by prothoracic glands of the silkworm, *Bombyx mori*.
 747 Insect Biochem. Molec. Biol. 41, 197-202.

748 Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger,
 749 M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N.,
 750 Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc,
 751 R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from

RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protocols
8, 1494-1512.

Hao, T., Zeng, Z., Wang, B., Zhang, Y., Liu, Y., Geng, X., Sun, J., 2014. The protein-protein
interaction network of eyestalk, Y-organ, and hepatopancreas in Chinese mitten crab
Eriocheir sinensis. BMC Systems Biol. 8.

Hatem, N.E., Wang, Z., Nave, K.B., Koyama, T., Suzuki, Y., 2015. The role of juvenile hormone
and insulin/TOR signaling in the growth of *Manduca sexta*. BMC Biol. 13.

Hopkins, P.M., 2012. The eyes have it: A brief history of crustacean neuroendocrinology. Gen.
Comp. Endocrinol. 175, 357-366.

Hopkins, P.M., Das, S., 2015. Regeneration in crustaceans, in: E.S. Chang, M. Thiel (Eds.), The
Biology of Crustacea: Physiology. Oxford University Press, Oxford, U.K., 168-198.

Huson, D.H., Mitra, S., 2012. Introduction to the analysis of enviromental sequences:
metagenomics with MEGAN. Methods Molec. Biol. 856, 415-429.

Imayavaramban, L., Dhayaparan, D., Devaraj, H., 2007. Molecular mechanism of molt-
inhibiting hormone (MIH) induced suppression of ecdysteroidogenesis in the Y-organ of
mud crab: *Scylla serrata*. FEBS Lett. 581, 5167-5172.

Jung, H., Lyons, R.E., Dinh, H., Hurwood, D.A., McWilliam, S., Mather, P.B., 2011.
Transcriptomics of a giant freshwater prawn (*macrobrachium rosenbergii*): *De novo*
assembly, annotation and marker discovery. Plos One 6.

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic
Acids Res. 28, 27-30.

773 Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2014. Data,
 774 information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res.*
 775 42, D199-D205.

776 Kim, H.W., Batista, L.A., Hoppes, J.L., Lee, K.J., Mykles, D.L., 2004. A crustacean nitric oxide
 777 synthase expressed in nerve ganglia, Y-organ, gill and gonad of the tropical land crab,
 778 *Gecarcinus lateralis*. *J. Exp. Biol.* 207, 2845-2857.

779 Kingan, T.G., 1989. A competitive enzyme-linked immunosorbent assay: application in the
 780 assay of peptides, steroids, and cyclic nucleotides. *Anal. Biochem.* 183, 283-289.

781 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* 9,
 782 357-359.

783 Layalle, S., Arquier, N., Leopold, P., 2008. The TOR pathway couples nutrition and
 784 developmental timing in *Drosophila*. *Dev. Cell* 15, 568-577.

785 Lee, K.J., Kim, H.-W., Gomez, A.M., Chang, E.S., Covi, J.A., MykleS, D.L., 2007. Molt-
 786 inhibiting hormone from the tropical land crab, *Gecarcinus lateralis*: Cloning, tissue
 787 expression, and expression of biologically active recombinant peptide in yeast. *Gen.*
 788 *Comp. Endocrinol.* 150, 505-513.

789 Lenz, P.H., Roncalli, V., Hassett, R.P., Wu, L.-S., Cieslak, M.C., Hartline, D.K., Christie, A.E.,
 790 2014. *De novo* assembly of a transcriptome for *Calanus finmarchicus* (Crustacea,
 791 copepoda): The dominant zooplankter of the north Atlantic ocean. *Plos One* 9.

792 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
 793 Durbin, R., Genome Project Data, P., 2009. The Sequence Alignment/Map format and
 794 SAMtools. *Bioinformatics* 25, 2078-2079.

795 Li, W., Godzik, A., 2006. Cd-hit: A fast program for clustering and comparing large sets of
796 protein or nucleotide sequences. *Bioinformatics* 22, 1658-1659.

797 Li, Y., Hui, M., Cui, Z., Liu, Y., Song, C., Shi, G., 2015. Comparative transcriptomic analysis
798 provides insights into the molecular basis of the metamorphosis and nutrition metabolism
799 change from zoeae to megalopae in *Eriocheir sinensis*. *Comp. Biochem. Physiol.* 13D, 1-
800 9.

801 Lv, J., Liu, P., Gao, B., Wang, Y., Wang, Z., Chen, P., Li, J., 2014. Transcriptome analysis of the
802 *Portunus trituberculatus*: *De novo* assembly, growth-related gene identification and
803 marker discovery. *Plos One* 9.

804 Macias, M.J., Martin-Malpartida, P., Massague, J., 2015. Structural determinants of Smad
805 function in TGF-beta signaling. *Trends Biochem. Sci.* 40, 296-308.

806 Marchal, E., Vandersmissen, H.P., Badisco, L., Van de Velde, S., Verlinden, H., Iga, M., Van
807 Wielendaele, P., Huybrechts, R., Simonet, G., Smagghe, G., Vanden Broeck, J., 2010.
808 Control of ecdysteroidogenesis in prothoracic glands of insects: A review. *Peptides* 31,
809 506-519.

810 Mattson, M.P., Spaziani, E., 1986. Regulation of crab Y-organ steroidogenesis in vitro: Evidence
811 that ecdysteroid production increases through activation of cAMP-phosphodiesterase by
812 calcium-calmodulin. *Molec. Cell. Endocrinol.* 48, 135-151.

813 Mattson, M.P., Spaziani, E., 1987. Demonstration of protein kinase C activity in crustacean Y-
814 organs, and partial definition of its role in regulation of ecdysteroidogenesis. *Molec. Cell.*
815 *Endocrinol.* 49, 159-171.

816 McCarthy, F.M., Wang, N., Magee, G.B., Nanduri, B., Lawrence, M.L., Camon, E.B., Barrell,
 817 D.G., Hill, D.P., Dolan, M.E., Williams, W.P., Luthe, D.S., Bridges, S.M., Burgess, S.C.,
 818 2006. AgBase: A functional genomics resource for agriculture. BMC Genomics 7.
 819 Mirth, C., Truman, J.W., Riddiford, L.M., 2005. The role of the prothoracic gland in determining
 820 critical weight to metamorphosis in *Drosophila melanogaster*. Curr. Biol. 15, 1796-1807.
 821 Mykles, D.L., 1997. Crustacean muscle plasticity: Molecular mechanisms determining mass and
 822 contractile properties. Comp. Biochem. Physiol. 117B, 367-378.
 823 Mykles, D.L., 2001. Interactions between limb regeneration and molting in decapod crustaceans.
 824 Am. Zool. 41, 399-406.
 825 Mykles, D.L., 2011. Ecdysteroid metabolism in crustaceans. J. Steroid Biochem. Molec. Biol.
 826 127, 196-203.
 827 Mykles, D.L., Adams, M.E., Gade, G., Lange, A.B., Marco, H.G., Orchard, I., 2010.
 828 Neuropeptide action in insects and crustaceans. Physiol. Biochem. Zool. 83, 836-846.
 829 Mykles, D.L., Medler, S., 2015. Skeletal muscle differentiation, growth, and plasticity, in: E.S.
 830 Chang, M. Thiel (Eds.), The Natural History of Crustacea: Physiology. Oxford University
 831 Press, Oxford, U.K., pp. 134-167.
 832 Nakatsuji, T., Lee, C.Y., Watson, R.D., 2009. Crustacean molt-inhibiting hormone: Structure,
 833 function, and cellular mode of action. Comp. Biochem. Physiol. 152A, 139-148.
 834 Nakatsuji, T., Sonobe, H., 2004. Regulation of ecdysteroid secretion from the Y-organ by molt-
 835 inhibiting hormone in the American crayfish, *Procambarus clarkii*. Gen. Comp.
 836 Endocrinol. 135, 358-364.
 837 Nakatsuji, T., Sonobe, H., Watson, R.D., 2006. Molt-inhibiting hormone-mediated regulation of
 838 ecdysteroid synthesis in Y-organs of the crayfish (*Procambarus clarkii*): Involvement of

839 cyclic GMP and cyclic neucleotide phosphodiesterase. *Molec. Cell. Endocrinol.* 253, 76-
840 82.

841 Pentek, J., Parker, L., Wu, A., Arora, K., 2009. Follistatin preferentially antagonizes activin
842 rather than BMP signaling in *Drosophila*. *Genesis* 47, 261-273.

843 R-Development-Core-Team, 2015. A language and environment for statistical computing. R
844 Foundation for Statistical Computing, Vienna, Austria.

845 Rewitz, K.F., Yamanaka, N., O'Connor, M.B., 2013. Developmental checkpoints and feedback
846 circuits time insect maturation, in: Y.B. Shi (Ed.), *Animal Metamorphosis*, pp. 1-33.

847 Roberts, A., Pachter, L., 2013. Streaming fragment assignment for real-time analysis of
848 sequencing experiments. *Nat. Meth.* 10, 71-73.

849 Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: A bioconductor package for
850 differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-
851 140.

852 Rodenfels, J., Lavrynenko, O., Ayciriex, S., Sampaio, J.L., Carvalho, M., Shevchenko, A.,
853 Eaton, S., 2014. Production of systemically circulating Hedgehog by the intestine couples
854 nutrition to growth and development. *Gene Dev.* 28, 2636-2651.

855 Shimobayashi, M., Hall, M.N., 2014. Making new contacts: the mTOR network in metabolism
856 and signalling crosstalk. *Nat. Rev. Molec. Cell Biol.* 15, 155-162.

857 Shrivastava, S., Princy, S.A., 2014. Pharmacophore based approach to design inhibitors in
858 Crustaceans: An insight into the molt inhibition response to the receptor guanylyl
859 cyclase. *Indian J. Exp. Biol.* 52, 375-382.

860 Skinner, D.M., 1985. Molting and regeneration, in: D.E. Bliss, L.H. Mantel (Eds.), *The Biology*
861 *of Crustacea*. Academic Press, New York, pp. 44-146.

862 Smith, W.A., Lamattina, A., Collins, M., 2014. Insulin signaling pathways in lepidopteran
863 ecdysone secretion. *Front. Physiol.* 5.

864 Song, C., Cui, Z., Hui, M., Liu, Y., Li, Y., Li, X., 2015. Comparative transcriptomic analysis
865 provides insights into the molecular basis of brachyurization and adaptation to benthic
866 lifestyle in *Eriocheir sinensis*. *Gene* 558, 88-98.

867 Spaziani, E., Jegla, T.C., Wang, W.L., Booth, J.A., Connolly, S.M., Conrad, C.C., Dewall, M.J.,
868 Sarno, C.M., Stone, D.K., Montgomery, R., 2001. Further studies on signaling pathways
869 for ecdysteroidogenesis in crustacean Y-organs. *Am. Zool.* 41, 418-429.

870 Suwansa-Ard, S., Thongbuakaew, T., Wang, T., Zhao, M., Elizur, A., Hanna, P.J., Sretarugsa, P.,
871 Cummins, S.F., Sobhon, P., 2015. *In silico* neuropeptidome of female *Macrobrachium*
872 *rosenbergii* based on transcriptome and peptide mining of eyestalk, central nervous
873 system, and ovary. *Plos One* 10.

874 Teleman, A.A., 2010. Molecular mechanisms of metabolic regulation by insulin in *Drosophila*.
875 *Biochem. J.* 425, 13-26.

876 Tom, M., Manfrin, C., Chung, S.J., Sagi, A., Gerdol, M., De Moro, G., Pallavicini, A.,
877 Giulianini, P.G., 2014. Expression of cytoskeletal and molt-related genes is temporally
878 scheduled in the hypodermis of the crayfish *Procambarus clarkii* during premolt. *J. Exp.*
879 *Biol.* 217, 4193-4202.

880 Tom, M., Manfrin, C., Giulianini, P.G., Pallavicini, A., 2013. Crustacean oxi-reductases protein
881 sequences derived from a functional genomic project potentially involved in ecdysteroid
882 hormones metabolism: A starting point for function examination. *Gen. Comp.*
883 *Endocrinol.* 194, 71-80.

884 Ventura, T., Cummins, S.F., Fitzgibbon, Q., Battaglione, S., Elizur, A., 2014. Analysis of the
885 central nervous system transcriptome of the eastern rock lobster *Sagmariasus verreauxi*
886 reveals its putative neuropeptidome. Plos One 9.

887 Ventura, T., Manor, R., Aflalo, E.D., Chalifa-Caspi, V., Weil, S., Sharabi, O., Sagi, A., 2013.
888 Post-embryonic transcriptomes of the prawn *Macrobrachium rosenbergii*: Multigenic
889 succession through metamorphosis. Plos One 8.

890 Walkiewicz, M.A., Stern, M., 2009. Increased insulin/insulin growth factor signaling advances
891 the onset of metamorphosis in *Drosophila*. Plos One 4.

892 Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: A revolutionary tool for transcriptomics.
893 Nat. Rev. Genet. 10, 57-63.

894 Webster, S.G., 2015. Endocrinology of Molting, in: E.S. Chang, M. Thiel (Eds.), The Natural
895 History of Crustacea: Physiology. Oxford University Press, Oxford, U.K., pp. 1-35.

896 Webster, S.G., Keller, R., Dirksen, H., 2012. The CHH-superfamily of multifunctional peptide
897 hormones controlling crustacean metabolism, osmoregulation, moulting, and
898 reproduction. Gen. Comp. Endocrinol. 175, 217-233.

899 Wei, J., Zhang, X., Yu, Y., Huang, H., Li, F., Xiang, J., 2014a. Comparative transcriptomic
900 characterization of the early development in pacific white shrimp *Litopenaeus vannamei*.
901 Plos One 9.

902 Wei, J., Zhang, X., Yu, Y., Li, F., Xiang, J., 2014b. RNA-Seq reveals the dynamic and diverse
903 features of digestive enzymes during early development of Pacific white shrimp
904 *Litopenaeus vannamei*. Comp. Biochem. Physiol. 11D, 37-44.

905 Yamanaka, N., Rewitz, K.F., O'Connor, M.B., 2013. Ecdysone control of developmental
906 transitions: Lessons from *Drosophila* research. Ann. Rev. Entomol. 58, 497-516.

907 Yu, X.L., Chang, E.S., Mykles, D.L., 2002. Characterization of limb autotomy factor-proecdysis
908 (LAF_{pro}), isolated from limb regenerates, that suspends molting in the land crab
909 *Gecarcinus lateralis*. Biol. Bull. 202, 204-212.

910

911

Table 1. Comparison of the three biological replicates of Y-organs from intermolt (Im) *G. lateralis* (land crab) after trimming and mapping results back to each library.

Library	Total raw reads from Illumina	Quality reads following trimming	RMBT following mapping via Bowtie
Im1	72,080,779	62,823,512	92.96%
Im2	81,221,231	74,977,808	91.70%
Im3	74,509,819	68,725,152	91.80%

917 **Table 2.** Statistics summary for land crab YO baseline transcriptome.

Total number of reads	227,811,829
Total number of contigs	288,673
Mean length (bp)	871
Median length (N50) (bp)	1,842
Number of clusters (CD-HIT-EST output)	231,579
Range of contig lengths (bp)	201 to 28,392
Number of clusters with one contig	203,954
Number of clusters with >1 contigs	27,625
Number of predicted peptides (Transdecoder)	81,481

918

919 **Table 3.** Comparison of alignments between sequences obtained from Sanger and Illumina
920 sequencing.

921

Gene (GenBank #)	Contig Identification	Contig length (bp)	ORF length (aa)	% Protein identity	% Protein positives	% Nucleotide identity
Akt (HM989974.3)	comp105797_c0_seq1	2,638	510	100	100	99
EcR (AY642975.1)	comp116905_c1_seq4	5,899	550	100	100	99
EF2 (AY552550.1)	comp101058_c2_seq1	3,018	846	100	100	99
Mstn (EU432218.1)	comp112193_c3_seq5	1,871	497	100	100	99
Ras (HM989971.1)	comp33885_c0_seq1	2,044	182	99	100	99
Calpain-B (AY639153.1)	comp111795_c1_seq91	4,664	1015	99	99	99
E75 (DQ058409.2)	comp62979_c0_seq1	3,759	721	99	99	100
mTOR (HM989973.3)	comp118786_c0_seq1	8,306	2475	98	98	99
NO-insensitive GC III (DQ355438.1)	comp119965_c0_seq2	2,511	326	99	98	99
NOS (AY552549.1)	comp119039_c0_seq1	4,835	1211	98	98	99
β-actin (L76530.1)	comp110624_c1_seq1	1,186	332	98	98	97
Calpain-T (AY639154.1)	comp112562_c0_seq3	2,853	642	97	97	99
S6K (HM989975.3)	comp95075_c0_seq1	2,532	492	97	97	99

922 **Table 4.** Annotation summary for land crab YO baseline transcriptome.

923

Annotation mode or tool	Number of contigs with hits or assignments
BLASTx against NR	34,605
BLASTx against SP	23,504
BLASTx against Trembl-Uniprot	25,723
BLASTp against NR	24,850
BLASTp against SP	19,707
BLASTp against Uniprot-Uniref90	24,817
Pfam domains	18,772
TMHMM	20,706
SignalP	5,745
GO annotation from BLASTx against NR	20,208
KEGG annotation from BLASTx against NR	15,090

924

925 **Table 5.** Components of selected KEGG signal transduction pathways in the land crab YO
 926 transcriptome.

Signaling Pathway	Total assigned contigs	Selected genes represented in the KEGG pathway
Notch (KO04330)	53	Dvl (K02353); Notch (K02599); Fringe (K05948); Delta (K06051); RBPSUH (K06053); γ -Secretase complex (K04505, K06170, K06171, K06172)
Hedgehog (KO04340)	59	Patched (K06225); Smoothened (K06226); Fused (K06228); Su(fu) (K06229); Ci (K16797)
Wnt (KO04310)	155	Wnt (K00182); APC (K02085); β -catenin (K02105); Axin (K02157); Dvl (K02353); Frizzled (K02354); GSK (K03083); CK II (K03115); TF 7 (K04491)
MAPK (KO04011)	213	EGF (K04357); EGFR (K04361); GRB (K04364); MEK1 (K04368); ERK (K0437); MNK1/2 (K04372); CREB (K04374); c-Myc (K04377); SRF (K04378); Ras (K07827); FGFR (K05093); SOS (K03099)

927

928 **Table 6.** Pearson correlation coefficients between the biological replicates.

929 **A. Using Count data**

Library	Im2	Im3
Im1	0.99	0.94
Im2	-	0.96

930

931 **B. Using FPKM data**

Library	Im2	Im3
Im1	0.99	0.96
Im2	-	0.95

932

933 **Table 7.** BLASTx hits for putative orthologs involved in MIH signaling.

Gene	Contig identification	Contig length	BLASTx hit (Subject)	Subject length (bp)	Score	e-value
Calmodulin	comp72270_c0_seq1	2838	gi 206597719 gb ACI15835.1 calmodulin [<i>Procambarus clarkii</i>]	149	759	7.59E-91
Adenylyl cyclase	comp113291_c0_seq5	9417	gi 398260007 emb CCF77365.1 Rutabaga adenylyl cyclase [<i>Calliphora vicina</i>]	2087	202	7.38E-12
Adenylyl cyclase	comp120618_c0_seq1	2132	gi 646714914 gb KDR18711.1 Adenylate cyclase type 5, partial [<i>Zootermopsis nevadensis</i>]	910	1316	2.82E-165
Protein kinase A	comp117998_c0_seq3	3311	gi 91093068 ref XP_968170.1 cAMP-dependent protein kinase catalytic subunit [<i>Tribolium castaneum</i>]	353	1495	0
Protein kinase A	comp120302_c0_seq5	3191	gi 702441760 gb AIW09155.1 cAMP-activated protein kinase γ subunit, partial [<i>Carcinus maenas</i>]	369	1514	0
Protein kinase G	comp171371_c0_seq1	544	gi 255349294 gb ACU09499.1 cGMP-dependent protein kinase [<i>Spodoptera exigua</i>]	744	675	2.22E-80
Calcineurin	comp113416_c1_seq2	2183	gi 501291888 dbj BAN20424.1 calcineurin β subunit [<i>Riptortus pedestris</i>]	189	770	1.26E-93
Calcineurin	comp105509_c3_seq1	1920	gi 357614525 gb EHJ69131.1 calcineurin B [<i>Danaus plexippus</i>]	293	772	2.94E-93

Figure legends

Figure 1. *De novo* assembly pipeline for the *G. lateralis* YO transcriptome. The steps in the assembly of the raw reads from the Illumina sequencing and for the estimation of transcript abundance are illustrated. The software used for each step is indicated by italics; the numbers indicate the output from each step. See Methods for the versions of the software used.

Figure 2. Annotation pipeline for the *G. lateralis* YO transcriptome. Contigs were analyzed using either nucleotide (BLASTx) or peptide (Transdecoder and BLASTp) sequences to search against non-redundant (NR), SwissProt (SP), and TrEMBL or Uniref90 databases. The output from BLASTx against the NR database was used to obtain functional and pathway annotations with Blast2GO for Gene Ontology (GO) and with MEGAN for KEGG pathway assignments, respectively. Trinotate was used to identify and aggregate protein sequences with transmembrane helix (TMH), SignalP, and Pfam domains into a single comprehensive file. The numbers indicate the total hits/assignments from the analyses. See Methods for the versions of the software used.

Figure 3. Venn diagram of contig numbers assigned to parent gene ontology (GO) terms.

The Venn diagram indicates number of contigs annotated by at least one or two or a combination of all three parent GO terms (Molecular Function, Cellular Component, and Biological Process). BLAST2GO was used for annotation and GOSlimViewer was used to summarize the hierarchical GO terms associated with each contig. A total of 20,208 contigs were annotated with least one parent GO term. The diagram is drawn to scale.

Figure 4. Number of YO transcriptome contigs distributed among second-tier GO terms.

The pie charts show the distribution of second-tier GO terms associated with three parent terms:

Molecular Function (A), Cellular Component (B) and Biological Process (C). The legends rank the terms from the highest to lowest number of contigs in each second-tier GO categories.

Figure 5. KEGG pathway analysis of land crab and crayfish YO transcriptomes. The figure depicts the distribution of total assignments for both land crab (A) and crayfish (B) BLASTx assigned contigs among global KEGG pathways. Land crab had more contigs in each category, but the proportions of the five global KEGG pathways were similar between the two species, as depicted by the percentage values.

Figure 6. Comparison of land crab and crayfish contigs associated with selected second- and third-tier KEGG hierarchical pathways. (A) Second-tier KEGG pathways associated with Metabolism, Genetic Information Processing, Environmental Information Processing, and cellular processes. The two pathways with the highest number of BLASTx assigned contigs were signal transduction and translation in the annotated transcriptomes of both species. (B) Third-tier KEGG pathways in the Signal Transduction category. The PI3K-Akt pathway had the highest number of assigned contigs in both species, while the TGF β , Jak-STAT, and Notch pathways had the least representation.

Figure 7. Identification genes assigned to the mTOR and TGF β signaling pathways in land crab and crayfish transcriptomes. MEGAN was used to identify genes annotated via BLAST and assigned to a KEGG pathway. The mTOR (A) and TGF β (B) signaling pathway components/genes were represented in both species.

Table A.1. The genes and KO associated with mTOR pathway and their corresponding number of contigs in land crab and crayfish transcriptome.

mTOR signaling pathway	Land crab	Crayfish
K07207 - tuberous sclerosis 2	8	1
K07198 - 5'-AMP-activated protein kinase, catalytic alpha subunit [EC:2.7.11.11]	8	1
K03259 - translation initiation factor 4E	6	0
K02991 - small subunit ribosomal protein S6e	6	1
K00922 - phosphatidylinositol-4,5-bisphosphate 3-kinase [EC:2.7.1.153]	6	2
K16172 - insulin receptor substrate 1	4	1
K07209 - inhibitor of nuclear factor kappa-B kinase subunit beta [EC:2.7.11.10]	3	0
K03258 - translation initiation factor 4B	3	1
K04456 - RAC serine/threonine-protein kinase [EC:2.7.11.1]	3	2
K04688 - p70 ribosomal S6 kinase [EC:2.7.11.1]	2	0
K02649 - phosphoinositide-3-kinase, regulatory subunit	2	1
K06276 - 3-phosphoinositide dependent protein kinase-1 [EC:2.7.11.1]	2	1
K16185 - Ras-related GTP-binding protein A/B	2	1
K16186 - Ras-related GTP-binding protein C/D	2	1
K07206 - tuberous sclerosis 1	2	2
K08269 - unc51-like kinase [EC:2.7.11.1]	2	3
K02677 - classical protein kinase C [EC:2.7.11.13]	2	3
K07298 - serine/threonine-protein kinase 11 [EC:2.7.11.9]	1	0
K01110 - phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN [EC:3.1.3.16 3.1.3.48 3.1.3.67]	1	1
K07208 - Ras homolog enriched in brain	1	1
K08266 - G protein beta subunit-like	1	1
K07203 - FKBP12-rapamycin complex-associated protein	1	1
K07204 - regulatory associated protein of mTOR	1	1
K08267 - rapamycin-insensitive companion of mTOR	1	1
K08268 - hypoxia-inducible factor 1 alpha	1	1
K04371 - extracellular signal-regulated kinase 1/2 [EC:2.7.11.24]	1	1
K04373 - p90 ribosomal S6 kinase [EC:2.7.11.1]	1	1
K08271 - protein kinase LYK5	1	1
K07205 - eukaryotic translation initiation factor 4E binding protein 1	1	2
K08272 - calcium binding protein 39	1	2

984 **Table B.1. The genes and KO associated with TGF β pathway and their corresponding**
985 **number of contigs in land crab and crayfish transcriptome.**

TGF-beta signaling pathway	Land crab	Crayfish
K04513 - Ras homolog gene family, member A	5	1
K03347 - cullin 1	4	0
K03456 - protein phosphatase 2 (formerly 2A), regulatory subunit A	4	0
K04679 - MAD, mothers against decapentaplegic interacting protein	4	1
K04678 - E3 ubiquitin ligase SMURF1/2 [EC:6.3.2.19]	4	4
K04676 - mothers against decapentaplegic homolog 1/5/8	3	1
K04382 - protein phosphatase 2 (formerly 2A), catalytic subunit [EC:3.1.3.16]	3	3
K04688 - p70 ribosomal S6 kinase [EC:2.7.11.1]	2	0
K04662 - bone morphogenetic protein 2/4	2	1
K04675 - activin receptor type-1 [EC:2.7.11.30]	2	1
K04501 - mothers against decapentaplegic homolog 4	2	1
K04681 - retinoblastoma-like protein 1	2	1
K03094 - S-phase kinase-associated protein 1	2	1
K04677 - mothers against decapentaplegic homolog 6/7	2	2
K04377 - Myc proto-oncogene protein	2	2
K04661 - follistatin	1	0
K04659 - thrombospondin	1	1
K04667 - inhibin, beta	1	1
K04671 - bone morphogenetic protein receptor type-2 [EC:2.7.11.30]	1	1
K13578 - bone morphogenetic protein receptor type-1B [EC:2.7.11.30]	1	1
K04674 - TGF-beta receptor type-1 [EC:2.7.11.30]	1	1
K04500 - mothers against decapentaplegic homolog 2/3	1	1
K04682 - E2F transcription factor 4/5	1	1
K03868 - RING-box protein 1	1	1
K04371 - extracellular signal-regulated kinase 1/2 [EC:2.7.11.24]	1	1
K04657 - chordin	0	1

986

FIGURES

FIG 1

Fig 2

Fig 3

Fig 4

A. Molecular Function

B. Cellular Component

C. Biological Process

Fig 5

Fig 6A

Fig 6B

Fig 7A

Fig 7B