Title: Salinity-induced changes in gene expression from anterior and posterior gills of *Callinectes sapidus* (Crustacea: Portunidae) with implications for crustacean ecological genomics

Authors: Justin C. Havird^{a,b*}, Reed T. Mitchell^{c,d}, Raymond P. Henry^c, Scott R. Santos^a

^a Department of Biological Sciences and Molette Laboratory for Climate Change and Environmental Studies, Auburn University, 101 Rouse Life Sciences Bldg., Auburn, Alabama 36849

^b Dept. of Biology, Colorado State University, Room E106 Anatomy/Zoology Building Fort Collins, CO 80523, USA

^c Dept. of Biological Sciences, Auburn University, 101 Rouse Life Sciences Bldg., Auburn, AL 36849, USA

^d Walter Reed Biosystematics Unit, 4210 Silver Hill Rd, Suitland, MD, 20746

Keywords: crustacean transcriptomics, RNA-Seq, osmoregulation, read depth, gill physiology

*Corresponding author: Justin C. Havird
Dept. of Biology, Colorado State University,
Room E106 Anatomy/Zoology Building
Fort Collins, CO 80523, USA
Fax: 970-491-0649
Email: Justin.Havird@colostate.edu

27

28 **Running title:** Salinity-induced gene expression in blue crab

29

Abstract

Decapods represent one of the most ecologically diverse taxonomic groups within crustaceans, making them ideal to study physiological processes like osmoregulation. However, prior studies have failed to consider the entire transcriptomic response of the gill – the primary organ responsible for ion transport – to changing salinity. Moreover, the molecular genetic differences between non-osmoregulatory and osmoregulatory gill types, as well as the hormonal basis of osmoregulation, remain underexplored. Here, we identified and characterized differentially expressed genes (DEGs) via RNA-Seq in anterior (non-osmoregulatory) and posterior (osmoregulatory) gills during high to low salinity transfer in the blue crab *Callinectes sapidus*, a well-studied model for crustacean osmoregulation. Overall, we confirmed previous expression patterns for individual ion transport genes and identified novel ones with salinity-mediated expression. Notable, novel DEGs among salinities and gill types for *C. sapidus* included anterior gills having higher expression of structural genes such as actin and cuticle proteins while posterior gills exhibit elevated expression of ion transport and energy-related genes, with the latter likely linked to ion transport. Potential targets among recovered DEGs for hormonal regulation of ion transport between salinities and gill types included neuropeptide Y and a KCTD16-like protein. Using publically available sequence data, constituents for a “core” gill transcriptome among decapods are presented, comprising genes involved in ion transport and energy conversion and consistent with salinity transfer experiments. Lastly, rarefaction analyses lead us to recommend a modest number of sequence reads (~10–15 M), but with increased biological replication, be utilized in future DEG analyses of crustaceans.

1. Introduction

Decapod crustaceans occupy a multitude of environmental extremes, from the open ocean to the intertidal zone, estuaries, freshwater streams and lakes (Hobbs and Hart, 1982) as well as semi- and fully terrestrial habitats (Bliss, 1968). With regards to salinity gradients, crustaceans can be generally classified as being either stenohaline (i.e., having a narrow salinity tolerance) or euryhaline (i.e., having a broad salinity tolerance) (Mantel and Farmer, 1983). Ecologically, stenohaline species are restricted to the euhaline (i.e., 30-35‰) environments of bays, oceans or the lower reaches (polyhaline zone, 18-30‰) of estuaries, while euryhaline species also occur in the euhaline to limnetic (i.e., 0‰) zones (Henry, 2001). The primary physiological adaptation conferring euryhalinity, and thus an expanded ecological range, is the ability to osmoregulate. Specifically, while virtually all marine crustaceans are passive osmoconformers in full-strength seawater (i.e., the osmotic concentration of their hemolymph reflects that of ambient seawater), euryhaline species actively transition to hyper-osmoregulation at lower salinities (e.g., at a critical salinity of 26‰ in the blue crab *Callinectes sapidus*) (Henry, 2005). The physiological and genomic basis of this osmoregulatory transition lies in differential gene expression within the gill.

In the heavily studied decapods, the gill undergoes significant ultrastructural, biochemical, and physiological remodeling during salinity transfer, ultimately resulting in the activation of transport mechanisms that take up ions (primarily Na^+ and Cl^-) from brackish or fresh waters and concentrates them in the hemolymph (reviewed in Charmantier *et al.*, 2009; Freire *et al.*, 2008; Henry *et al.*, 2012; McNamara and Faria, 2012). These mechanisms involve coordination among transmembrane ion transporters (e.g., Na^+/K^+ -ATPase [NKA]) and accessory enzymes (e.g., carbonic anhydrase [CA]), with dramatic increases in both gene

expression and protein-specific activity during low-salinity transfers in brachyuran crabs (Havird *et al.*, 2013; Henry, 1988; Luquet *et al.*, 2005; Serrano *et al.*, 2007; Serrano and Henry, 2008; Towle *et al.*, 1976). Notably, only the posterior gills are involved in osmoregulation and undergo such changes, with anterior gills being specialized for respiration in marine species (Freire *et al.*, 2008; Henry, 1984; Henry and Cameron, 1982; Neufeld *et al.*, 1980; Taylor and Taylor, 1992).

Despite considerable knowledge of osmoregulation in decapods, a number of outstanding questions remain unanswered. For example, just a handful of genes have been thoroughly examined in response to salinity, namely those previously identified from physiological or biochemical studies as undergoing dramatic expression changes in high vs. low salinity (Havird *et al.*, 2013). Such gene-centric studies, however, fall short of identifying: 1) other candidates involved in ion transport; 2) targets within the large pool of potential regulatory mechanisms, and; 3) participants in metabolic pathways necessary to support the energetically demanding process of osmoregulation. These points are well-illustrated in a microarray examination of expression for 4462 genes in the green crab *Carcinus maenas* during salinity transfers (Towle *et al.*, 2011), where upregulation of well-characterized ion transporters such as NKA was confirmed while revealing: 1) novel ion transporters (e.g., a Na⁺/glucose transporter) with expression patterns similar to NKA and CA; 2) no differential expression (DE) in stress genes, suggesting responses to changing salinity are a routine physiological process; 3) up-regulated mitochondrial biogenesis pathways, implying increased ATP demand in brackish waters, and; 4) upregulation of numerous genes lacking annotation. Recently, studies have employed RNA-Seq to examine salinity-mediated gene expression in the gills of decapods, often recovering thousands of DE genes, many belonging to functional groups not previously implicated in osmoregulation (e.g., chitin and amino acid metabolism) (Li *et al.*, 2014; Lv *et al.*, 2013). In

spite of these advancements, the molecular basis underlying functional differences between the anterior and posterior gills of decapods remain poorly understood and have yet to be investigated at the transcriptomic level.

Here, we utilized an RNA-Seq approach to examine the transcriptomic responses of anterior and posterior gills to salinity transfer in *Callinectes sapidus*, a model species in crustacean osmoregulation studies (Cameron, 1978; Cameron and Batterton, 1978; Copeland and Fitzjarrell, 1968; Engel, 1977; Henry, 1988; Henry and Watts, 2001; Lovett *et al.*, 2006a; Piller *et al.*, 1995; Serrano *et al.*, 2007; Towle and Kays, 1986; Towle *et al.*, 1976). In this context, we predicted that prior patterns of differential expression in ion transporters and accessory enzymes should be confirmed while also identifying previously undocumented genes involved in osmoregulation. Specifically, we hypothesized the molecular basis underlying differing salinity responses in anterior vs. posterior gills involves not only changes in the baseline expression of ion transporters and accessory enzymes, but also structural and metabolic-related genes associated with the salinity-mediated remodeling and high energetic costs, respectively, of osmoregulation in the posterior gills. Along with this, a subsampling scheme and rarefaction analysis was conducted in order to suggest sequencing depth requirements for future RNA-Seq studies of environmentally-mediated gene expression. Lastly, we compared our *C. sapidus* transcriptomic data to publically available data from other species to identify and characterize a “core” transcriptome among gills of the decapods.

2. Methods

2.1. Animals, holding conditions, and RNA extraction

Adult, intermolt *Callinectes sapidus* were held in the laboratory as described previously (Mitchell and Henry, 2014). Briefly, animals obtained from commercial fishers in East Point, FL, USA were shipped to Auburn University, AL, USA and held in 570 L recirculating tanks of full-strength seawater (35‰) at 24 °C for three weeks, with salinity being monitored daily. Following this acclimation period, one anterior (G3) and one posterior (G7) gill was removed from each of three individuals. Another set of anterior and posterior gills was subsequently sampled from individuals transferred to 10‰ for 1 week after being maintained at 35‰ for three weeks. In both cases, gills were transferred immediately to 3 mL of RNAGents Denaturing Solution (Promega, Madison, WI, USA) and kept on ice until homogenization. Total RNA was extracted from the 12 *C. sapidus* gills ($n = 3$ for four treatments: anterior and posterior gills at 35‰ and 10‰) using the RNAGents Total RNA Isolation System (Promega) as described previously (Havird *et al.*, 2014b; Mitchell and Henry, 2014).

2.2. Transcriptomics: cDNA library preparation, sequencing, assembly, and annotation

Protocols largely followed those used for previously published crustacean transcriptomes generated in our laboratories (Havird *et al.*, 2014a). Briefly, the SMART cDNA construction kit (ClonTech Laboratories Inc., Mountain View, CA, USA), designed for priming efficiency and enrichment of full-length cDNAs from low RNA starting amounts (i.e., ~1-2 ng total), was used as described in Havird *et al.* (2014a), with libraries being sent to the HudsonAlpha Institute for Biotechnology's Genomic Services Lab (Huntsville, AL, USA). From these cDNA libraries, 100 bp paired-end (PE) reads were generated on an Illumina HiSeq 2000 (Illumina, San Diego, CA, USA), with each sample being sequenced on 1/6th of a lane. Raw reads in FASTQ format were deposited in NCBI's Sequence Read Archive (SRA) under BioProject PRJNA315576.

A reference, composite gill transcriptome was first assembled for *C. sapidus* using the single biological replicate with the greatest number of sequence reads from each treatment, for a total of four samples from four different treatments being used to assemble the composite gill transcriptome. All reads for each of the four samples were combined into a single FASTQ file and no quality filtering or trimming was performed given the overall high quality of reads (mean quality score = 35) and the fact that overzealous preprocessing can negatively impact transcriptome assembly (MacManes, 2014). Following digital normalization using the `normalize-by-median.py` script (Brown *et al.*, 2012), the *C. sapidus* composite gill transcriptome was assembled using Trinity version Trinityrnaseq_r20131110 (Grabherr *et al.*, 2011) under default parameters and 6 CPUs. Separate transcriptomes were also assembled for each of the four treatments by randomly subsampling 10 million (M) reads from each of the three biological replicates in a treatment with seqtk (<https://github.com/lh3/seqtk>) and combining them into a single FASTQ file (i.e., of 30 M reads per treatment). Assemblies were then generated using the same scheme as for the composite transcriptome. Each assembly generally took < 24 hours, and five unique contig sets (i.e., five transcriptomes) were generated overall.

Transcriptome annotation was done by extracting putative open reading frames (ORFs) from the five contig sets using Transdecoder (version TransDecoder_r20131117; <https://transdecoder.github.io/>) followed by submission to a local version of Trinotate (version Trinotate_r20131110; <https://trinotate.github.io/>). Several different annotation methods and databases are employed in Trinotate, including: similarity searches with NCBI-BLAST v.2.2.29+ (Altschul *et al.*, 1997), protein domain identification to the PFAM database (Punta *et al.*, 2012) using HMMER v.3.1b1 (Finn *et al.*, 2011), and queries to the UniProt, EggNOG, and Gene Ontology (GO) Pathways databases (Apweiler *et al.*, 2012; Ashburner *et al.*, 2000; Powell *et al.*,

2012). For each of the five transcriptomes, a tab-delimited table summarizing the annotation results at a BLAST Expect (E) value of $1e^{-5}$ was generated for parsing as described previously (Havird *et al.*, 2014a).

2.3. Differential gene expression and rarefaction analyses

Differentially expressed genes (DEGs) between all pairwise comparisons of the four treatments were identified using the utilities implemented in Trinity version Trinityrnaseq_r20131110, following the protocol outlined in Haas *et al.*, (2013). Briefly, DEG analyses were accomplished in three main steps: 1) reads from each sample were mapped to the *C. sapidus* composite gill transcriptome using Bowtie version 1.1.1 (Langmead and Salzberg, 2012); 2) Trimmed Mean of M-values (TMM) normalization (Dillies *et al.*, 2012; Li and Dewey, 2011) and estimates of fragments per kilobase of transcript per million mapped reads (FPKM; Trapnell *et al.*, 2010) were generated for each treatment/contig set using RSEM version 1.2.12, and; 3) DEGs were identified between each pairwise comparison of treatments using either edgeR (Robinson *et al.*, 2010) or DESeq (Anders and Huber, 2010), with $P < 0.05$, FDR (i.e., false discovery rate) < 0.1 , and C (i.e., minimal $\log_2$ fold-change) $= 2$. To test for gene-set enrichments of particular DEGs, Uniprot accession numbers were extracted from their annotation and submitted to the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.7 (Huang *et al.*, 2007a; Huang *et al.*, 2007b). In addition to the above analyses based on all reads (averaging 28.4 M reads per biological replicate), we examined how a subsampling of reads might influence the number, as well as identity, of DEGs between treatments. Here, a random subsample of either 1, 2, 5, 10, 15, or 20 M reads was extracted from each biological

replicate ($n = 12$ in total) using seqtk and DEGs identified as above for each pairwise comparison between treatments, with the exception of using a $FDR < 0.05$.

2.4. Towards identifying a “core” gill transcriptome for decapods

To identify and characterize a potential “core” gill transcriptome for the decapods (i.e., a set of contigs shared across species in the group and likely critical for gill function), publicly available transcriptomic data in NCBI’s Sequence Read Archive (SRA) from gill tissue of four other species (Table 1) were compared to the composite gill transcriptome described here for *C. sapidus*. Specifically, the four gill data sets were from the crabs *Portunus trituberculatus* (Lv *et al.*, 2013), *Hyas araneus* (Harms *et al.*, 2014; Harms *et al.*, 2013), and *Carcinus aestuarii* (Romiguier *et al.*, 2014), and the giant river prawn *Macrobrachium rosenbergii* (Mohd-Shamsudin *et al.*, 2013).

Raw sequence reads were downloaded for each gill data set by selecting SRA accessions possessing approximately ~30 M reads per species (Table 1) to be comparable to *C. sapidus*. Following digital normalization, transcriptomes were assembled with Trinity for the four species using the same parameters as previously described for *C. sapidus*. Given that data from multiple salinity treatments could be obtained for *P. trituberculatus*, a composite transcriptome was generated for this species by randomly subsampling 10 M reads from each of the three available salinity treatments (i.e., 5‰, 33‰ and 50‰) as described above for *C. sapidus*.

To recover those contigs in the four transcriptomes that were highly expressed, sequence reads were mapped to their respective assemblies using Bowtie version 1.1.1 (Langmead, 2010) and the numbers of mapped reads quantified using RSEM version 1.2.12, as implemented in Trinity with the provided `align_and_estimate_abundance.pl` script.

Contigs were then selected based on a minimum expression level of FPKM ≥ 10 , as this generally encompassed the top 20% of expressed contigs. Annotation of these contigs was done with Trinotate as described above, with ones lacking annotation excluded from further analyses. This filtering scheme was performed in order to obtain a set of high expression contigs for each transcriptome that could be functionally interpreted.

Shared, putative orthologs representing a potential “core” gill transcriptome for the five species were identified and characterized following the methodology described in Havird *et al.* (2014a). Briefly, the peptide sequences outputted by Transdecoder for each of the five transcriptomes/species were submitted to CD-HIT version 4.6.1-2012-08-27 (Fu *et al.*, 2012) to generate clusters at the 90% similarity level, with the single-longest representative of each cluster retained. These were then utilized as input to the OrthoMCL version 2.0.9 (Li *et al.*, 2003) protocol (<http://orthomcl.org/common/downloads/software/v2.0/UserGuide.txt>) and putative ortholog sets inferred using mcl version 12-068 with the $-I \ 1.5$ option (Salichos and Rokas, 2011). Only putative orthologs having a representative from each of the five species were included in the “core” gill transcriptome. Clustal Omega version 1.2.1 (Sievers *et al.*, 2011) was then used to align each set of putative orthologs, with consensus sequences from the alignments being annotated via BLASTP against NCBI’s nr database. To further characterize the functional groups found in the “core” gill transcriptome, DAVID was utilized as described above and its output visualized as a gene-set enrichment network generated via the Enrichment Map plugin version 0.1 to Cytoscape version 2.8.2 (Shannon *et al.*, 2003), following the protocol outlined from the Bader Lab (<http://www.baderlab.org/Software/EnrichmentMap/DavidTutorial>).

2.5. Data availability

All data are publicly available from <http://www.auburn.edu/~santosr/sequencedatasets.htm>. These include the transcriptomes from the five decapod species, including gill and treatment-specific as well as composite assemblies from *C. sapidus* (totaling nine .fasta files) along with their corresponding annotation files (e.g., .cds, .pep, .gff3, .bed, and tab-delimited summary files from Trinotate). Summarized output files for all pairwise treatment comparisons are also included, such as tables of expression statistics and annotations for all DEGs. Additionally, alignments representing each putative ortholog in the “core” gill transcriptome are provided, along with the functional gene-set enrichment files of these contigs from DAVID. Any intermediate files not publically available will be provided upon request. Lastly, contigs from the *C. sapidus* composite gill transcriptome were uploaded to NCBI’s Transcriptome Shotgun Assembly (TSA) database under accession GEID.

3. Results

3.1. *Callinectes sapidus* and other decapod gill transcriptomes

We first assembled a composite gill transcriptome for *C. sapidus* to serve as a reference using sequence reads taken from a single sample (i.e., the one with the highest number of reads) of each treatment. In total, 131,926,487 100 bp PE reads were utilized and the composite gill transcriptome assembled into 239,158 contigs (Table 1). However, only 35.4% of these contigs were able to be annotated via Trinotate (Table 1). Treatment-specific transcriptomes for *C. sapidus* were also assembled using a random subsampling of 10 M reads from each of the three biological replicates in a treatment. Assembly statistics were similar for these treatment-specific

transcriptomes, although they had generally fewer contigs, as well as lower annotation rates, than the composite reference transcriptome (Table 1).

We also assembled transcriptomes from gill tissues of four additional decapods based on publicly available RNA-Seq data, and some of these assemblies differed substantially from those described previously. For example, ~3 times more contigs were recovered from *Hyas araneus* than originally described (Harms *et al.*, 2013), likely due to Illumina, rather than Roche 454, data becoming available for this species. However, contig numbers for *Portunus trituberculatus* (123,015 vs. 94,511) and *Macrobrachium rosenbergii* (95,321 vs. 102,230) were similar relative to previous transcriptome assemblies, likely because they were generated from the same sequence data with comparable assembly methods (Lv *et al.*, 2013; Mohd-Shamsudin *et al.*, 2013). Lastly, to the best of our knowledge, this is the first report of a gill transcriptome for *Carcinus aestuarii*, which possessed 48,355 contigs and a similar annotation rate to that of *C. sapidus* (Table 1).

3.2. Identification of DEGs

Here, we focus on the DEGs identified using edgeR since those most closely approximated numbers reported in previous studies (Li *et al.*, 2014; Lv *et al.*, 2013; Towle *et al.*, 2011). Overall, a total of 477 DEGs were identified among the pairwise comparisons, relating to a wide variety of physiological processes (Fig. 1). However, while hundreds of DEGs were identified in most pairwise comparisons (Table 2), our ability to interpret the biological significance from many of them is unfortunately hampered due to an extremely low overall annotation rate of ~24.6%. Moreover, while a few previously-studied genes known to be involved in ion transport, as well as some other genes of interest, were not identified as DEGs in

our study (Fig. 2), this may be due to: 1) relatively low power given our limited sample sizes ($n = 3$ per treatment); 2) high levels of inter-individual variation in gene expression in the current study (Fig. 1), and/or; 3) methodological differences between previous qPCR studies and the work presented here.

At 35‰, *C. sapidus* acts as an osmoconformer, so no osmoregulatory processes should be operating in the gills. In spite of this, 109 genes were identified as differentially expressed between anterior and posterior gills at 35‰, including ion transporters previously known to be upregulated in response to low (i.e., 10‰) salinity. Specific examples are the β subunit of NKA, a NKA-interacting protein, a Na^+/H^+ exchanger, and a degenerin-like/ Na^+ channel expressed 19-, 128-, 6-, and 113-fold higher, respectively, in posterior compared to anterior gills. Other genes highly expressed in posterior gills included RNA replication proteins, a transportin, a NADH-cytochrome b5 reductase, a glutamate synthase, and a neuropeptide Y receptor. In contrast, cuticle proteins, actin, and a retinol dehydrogenase were expressed at higher levels in anterior gills. The DAVID analyses of these DEGs recovered statistically significant gene-set enrichment for GO terms related to cuticle structure, actin, and membrane integrity between the anterior and posterior gills at 35‰.

Following seven days at 10‰, when *C. sapidus* acts as a strong osmoregulator, just 13 DEGs were identified between anterior and posterior gills. One of these was again the β subunit of NKA, being expressed 22-fold higher in posterior gills. While an ADIPOR-like receptor and a neuregulin gene were also expressed higher in posterior than in anterior gills, the other 10 DEGs from this comparison lacked annotation. Given this, the DAVID analyses did not identify any significant gene-set enrichment in GO terms between the anterior and posterior gills of *C. sapidus* at 10‰.

When anterior gills were compared between salinity treatments, DEGs that were upregulated at 10‰ were ones involved in virus RNA replication and a potassium channel, while upregulated genes at 35‰ included an angiotensin-converting enzyme and actin. In this case, DAVID analyses included significant gene-set enrichment for GO terms related to RNA transcription, viral transcription, and potassium/ion channels.

Comparing salinity treatments of the posterior gills identified only a single DEG previously known to be important for osmoregulation (Henry, 2001): the cytoplasmic isoform of carbonic anhydrase (i.e., CAC, CAII), which was upregulated 28-fold at 10‰ compared to 35‰. Along with CAC, other annotated, upregulated DEGs recovered at 10‰ were an alcohol dehydrogenase, a degenerin-like/ Na^+ channel protein, and a chloride channel. On the other hand, genes upregulated at 35‰ included a NADH-cytochrome b5 reductase and a venom carboxylesterase. Here, DAVID analyses found significant gene-set enrichment for GO terms related to mitochondrial and organelle structures, oxidative reduction, and various metabolic processes.

Lastly, while it may be difficult to interpret DEGs identified in comparisons where gill type and salinity are confounding factors, it is notable that many of the recovered genes were the same ones mentioned above. Generally, ion transporters, such as NKA, and mitochondrial metabolism proteins were upregulated in posterior gills at 10‰ while cuticle proteins and actin-like proteins were upregulated in anterior gills at 35‰. For example, NKA β and CAC were upregulated 50-fold and 66-fold, respectively, along with 5-6-fold upregulation of six OXPHOS subunits, in posterior gills at 10‰ compared to anterior gills at 35‰. NKA β was also expressed at levels 8-fold higher in posterior gills at 35‰ relative to anterior gills at 10‰, implying

particular ion transporters in the posterior gills may play central roles and be expressed at higher levels than in their anterior counterparts regardless of salinity.

3.3. Rarefaction analyses of salinity-induced differential expression

Based on rarefaction analyses, numbers of recovered DEGs between specific pairwise comparisons generally increased as more sequence reads were utilized in their estimate (Fig. 3). In most comparisons, however, the bulk (>50%) of DEGs were recovered with just 10 M reads while the vast majority (~80%) were identified using 15 M reads (Fig. 3). For the anterior gills at 35‰ vs. posterior gills at 10‰ comparison, which yielded the largest number of DEGs, 11.4 DEGs were further recovered for each additional 1 M reads analyzed. However, this pattern reached an asymptote, subsequently dropping to a 2.3 increase in DEGs per additional M reads when >10 M reads were included in analyses. Notably, a different trend was observed for comparisons with more moderate numbers of DEGs (i.e., when comparing anterior gills at 10‰ vs. 35‰), with DEG numbers generally increasing linearly with number of reads analyzed (i.e., no asymptote was reached). Finally, for the comparison with the fewest DEGs (i.e., anterior gills at 10‰ vs. posterior gills at 10‰), highly similar numbers of DEGs were recovered regardless of how many reads were included in analyses. In all cases, the annotated identities of DEGs, and biological interpretation of the results, were essentially identical between subsampling schemes.

3.4. A “core” gill transcriptome for decapods

The “core” gill transcriptome identified here for decapods consisted of 476 annotated transcripts, not including ones also shared between the species but lacking annotation, which were excluded from further analyses. Based on these 476 transcripts, 158 nodes having

functional annotation were identified (Fig. 4). By using the “Word Cloud” feature of the Enrichment Map plugin for Cytoscape, enriched GO terms associated with functional sub-networks in the “core” gill transcriptome included those attributable to cellular maintenance such as transcription and translation, as well as ones associated with known gill physiology functions. As an example of the latter, 59 nodes related to ion transport were identified as a sub-network in the “core” gill transcriptome of these decapods (Fig. 4).

5. Discussion

5.1. Distinct transcriptome signatures among anterior and posterior gills of *C. sapidus*

In *C. sapidus*, posterior gills are solely responsible for osmoregulation while anterior ones are considered to function primarily in passive respiratory gas transport (Lovett *et al.*, 2006a; Lovett *et al.*, 2006b). This is consistent with previous gene-targeted studies of osmoregulatory genes, such as NKA α , reporting higher expression levels in posterior gills, even when individuals were chronically acclimated to high salinity (Serrano *et al.*, 2007). In this context, our transcriptome data extends this trend to previously unexplored ion transporting genes, which had higher expression levels in the posterior relative to the anterior gills during chronic high salinity acclimation (the physiological baseline condition), including most that were DE due to salinity: NKA β , a NKA β -interacting protein, the degenerin/Na⁺ channel, and NHE2. In sharp contrast, fewer genes were DE between gill types under low salinity, with NKA β being the only ion transporter having statistically higher expression in posterior gills regardless of treatment type. A potential reason for this result is that a single biological replicate from the anterior gill at 10‰ possessed expression patterns similar to the posterior gills. Thus, increasing biological replication may yield the anticipated trend of increased ion transporter expression

being salinity-induced in the posterior gills only. Along with ion transporters, numerous other genes were categorized as DE between gill types due to salinity. Of these, ones relating to energy metabolism (e.g., NADH-cytochrome b5 reductase) and structural remodeling (e.g., actin and cuticle proteins) further support our initial hypotheses, and likely reflect the energetic requirements and ultrastructural modifications, respectively, of osmoregulation in the posterior gills of decapods like *C. sapidus*.

Notably, other DE genes according to gill type provide insight into possible regulatory pathways of osmoregulation, thus illuminating the still largely unexplored hormonal basis of ion uptake in crustaceans (Henry *et al.*, 2012). For example, a neuropeptide Y receptor, which is a neurotransmitter conserved across nearly all animals that functions in G-protein signaling and regulates a variety of physiological functions (Alfalsh and Michel, 2004), was upregulated in posterior gills. Another possible target upregulated in posterior gills was the BTB/POZ domain-containing protein KCTD16, which is also involved in G-protein signaling as well as K⁺ channel function (Liu *et al.*, 2007). Because an as-of-yet unidentified hormone found in the X-organ of the crustacean eyestalk is likely responsible for regulating CAC induction under low salinity (Henry and Campoverde, 2006), examining the acute expression as well as activity of these proteins specifically in crustacean eyestalks during salinity transfer could be a logical next step towards identifying the hormonal basis of osmoregulatory mechanisms in crustaceans. Additionally, an ADIPOR-like receptor was upregulated in posterior gills regardless of salinity, implying lipid metabolism may play an important role in providing the energy needed for posterior gill functions like osmoregulation. Supporting this, lipids have been suggested as an energy source during salinity acclimation in crustaceans (Goolish and Burton, 1989; Luvizotto-Santos *et al.*, 2003).

5.2. Salinity-induced gene expression in the gills of *C. sapidus*

By examining salinity-induced gene expression across the entire gill transcriptome of *C. sapidus*, the results presented here generally confirmed patterns of differential expression previously identified in gene-centric studies. Although post-transcriptional and post-translational modifications also likely play a role in osmoregulation (Henry *et al.*, 2012; McNamara and Faria, 2012; Piller *et al.*, 1995; Towle *et al.*, 2001) for crustaceans like *C. sapidus*, focus has been primarily on characterizing expression patterns of genes such as Na⁺/K⁺-ATPase (NKA; Lucu and Towle, 2003; Serrano *et al.*, 2007) and carbonic anhydrase (CA; Mitchell and Henry, 2014; Serrano *et al.*, 2007) that act as ion transporters or accessory enzymes, respectively. The up-regulation of osmoregulatory genes that result from decreases in ambient salinity typically take place over a period of hours, with the resulting changes in protein activity occurring over a period of days (e.g., Serrano *et al.*, 2007; Serrano and Henry, 2008). This pattern, which is common among osmoregulating species, has been interpreted as evidence for a common mechanism of recruitment of osmoregulatory genes for long-term acclimation. Ion transport also occurs independently of osmoregulation and plays a role in the mechanism of cell volume regulation and acid-base regulation (reviewed by Henry *et al.*, 2012), processes that take place over a short time span and are more transient in duration. It is in these processes that post-translational mechanisms would be most likely be involved, and which represent a fertile area for systematic investigation.

While some methodological differences exist between ours and previous studies (e.g., which exact gill pairs were examined), direct comparison can be made between them in order to extrapolate overall themes. For example, the cytoplasmic isoform of CA (CAc) generally tends

to be upregulated 10- to 100-fold in the posterior gills of *C. sapidus* at 7 days following transfer from seawater to brackish water (Mitchell and Henry, 2014; Serrano *et al.*, 2007), which is typical of most previously studied crustaceans (Havird *et al.*, 2013). Our results are consistent with this, as CAC was significantly upregulated 28-fold at 10‰ in posterior gills relative to 35‰. Although the previously studied α subunit of NKA was not identified as a DEG here, this is likely due to its modest increase of 2.2-fold in our study (Fig. 2), despite the fact that all three samples had elevated NKA α expression in the posterior gills at 10‰ compared to 35‰. This increase, although not statistically significant, is similar to the 2.0-, 2.5-, and 3.4-fold differences previously reported for *C. sapidus* (Lovett *et al.*, 2006b; Lucu and Towle, 2003; Serrano *et al.*, 2007). Similarly, two other previously examined genes in *C. sapidus*, arginine kinase and the membrane-associated isoform of CA (CAG), were not DE in our study, as was also reported for CAG (Serrano *et al.*, 2007) as well as one study of AK (Lovett *et al.*, 2006b), although this same gene was identified as being upregulated 3.3-fold elsewhere (Serrano *et al.*, 2007).

The difference in upregulation observed for NKA compared to CA may be attributed to the subcellular localization of these two proteins. The NKA, which was weakly upregulated, is a trans-membrane protein that transports ions from either the outer or inner boundary layers (Henry *et al.*, 2012; Towle and Weihrauch, 2001), which is functionally a separate fluid compartment within the cell due to its low, and essentially stagnant, volume. As such, a small increase in NKA protein abundance could have a large effect on membrane transport. Although NKA drives Na^+ and Cl^- uptake, this process requires the counterions H^+ and HCO_3^- , which come from the hydration of CO_2 by CAC (e.g., Henry and Cameron, 1983; Henry *et al.*, 2003). The supply of these counterions from the general cytoplasm to the interior boundary layer, however, is believed to be diffusion-limited. Because of this, CAC must maintain the

instantaneous equilibrium between CO_2 and $\text{H}^+/\text{HCO}_3^-$ throughout the entire intracellular volume while ensuring the supply of these products (i.e., diffusion from cytoplasm to boundary layer) does not become the rate-limiting step in ion uptake (Gutknecht *et al.*, 1977). Given this, a high degree of CAC gene induction may be necessary to meet the increased rates of ion uptake across the gill under low salinity. Other salinity-induced DE ion transporters are also likely membrane-integrated proteins, possibly explaining the similar levels of upregulation among these genes. On the other hand, DEGs showing CAC-like patterns may be involved in accessory functions for ion transport related to osmoregulation, like dynein, which underwent a 33-fold increase.

In addition to the well-characterized genes discussed above, we recovered ion transporters upregulated at low salinity, but which have not been previously identified as such in prior studies of crustacean osmoregulation. These included upregulation of 1) a chloride channel (7-fold in posterior gills); 2) a degenerin-like protein, which is a type of sodium channel, (34-fold in posterior gills), and; 3) a potassium channel (6-fold in the anterior gills). While such channels have been proposed in models of ion regulation (Fehsenfeld and Weihrauch, 2016; Henry *et al.*, 2012; McNamara and Faria, 2012; Onken *et al.*, 2003; Towle and Weihrauch, 2001), salinity-induced DE of such genes has not been shown previously for many crustaceans. Additional ion transporters of interest include those being DE in other crustaceans but not identified as such in our study (Fig. 2). For instance, the $\text{Na}^+/\text{K}^+/\text{2Cl}^-$ -cotransporter (NKCC) was not identified as being DE in spite of a 2-fold upregulation in posterior gills at 10‰ compared to 35‰ (Fig. 2); in contrast, it was significantly upregulated ~10-fold in posterior gills 8 days after transfer from 32‰ to 2‰ in *Chasmagnathus granulatus* (Luquet *et al.*, 2005). Such genes are typically upregulated more so during extreme salinity transfers (Havird *et al.*, 2013), which may explain this discrepancy with *C. granulatus*. However, it is also likely that detecting small

changes in DE for ion transporters such as NKA and NKCC may require more liberal statistical parameters in future RNA-Seq studies (e.g., potentially lowering *C*, the threshold for fold-changes).

Along with ion transporters and their accessory enzymes, additional genes of potential physiological importance were upregulated under low salinity in both posterior and anterior gills. An example is carbohydrate sulfotransferase 11, which catalyzes the transfer of sulfate onto chondroitin, the major structural component of the extracellular matrix. In this case, its upregulation in posterior gills might play a role in the structural remodeling that occurs during low salinity transfer (Freire *et al.*, 2008; Taylor and Taylor, 1992). However, the reason for upregulation of these, and other structurally-related, genes in anterior gills is less readily explainable and somewhat puzzling as no major remodeling is known to take place in anterior gills during salinity transfer. One possibility is that both anterior and posterior gills contend with cell swelling, and subsequent volume regulation, under low salinity, with structural changes possibly being involved in that process. Future experiments to test this hypothesis could involve monitoring such genes in the gills of purely osmoconforming decapod species. Significant DE was also observed in virus-related replication genes in the anterior gills due to low salinity, raising the interesting scenario that viruses may propagate preferentially under such circumstances. Moreover, because low-salinity conditions require increased energy expenditure, such conditions could put individuals into energetic deficiency and possibly lead to greater susceptibility of viral or bacterial infection in sites such as the gills. Support for such a scenario comes from the upregulation of mitochondrial and energy-related genes in low salinity reported here, along with findings showing growth retardation due to viral infection is exacerbated by salinity in the shrimp *Penaeus vannamei* (Bray *et al.*, 1994).

Interestingly, no stress-related genes were upregulated at 10‰ in *C. sapidus*, including heat shock proteins (HSPs) or oxidative stress enzymes (Fig. 2). This mirrors results from microarray analyses of the green crab *Carcinus maenas* (Towle *et al.*, 2011) and supports the hypothesis that this degree of salinity transfer does not represent a significant stressor to at least some euryhaline crustaceans (de la Vega *et al.*, 2007; Towle *et al.*, 2011). As a general caveat to our study, however, DEGs were investigated seven days following salinity transfer. Thus, we may have missed some acute changes since the strongest DE response in crustaceans tends to occur 1-3 days following salinity transfer (Havird *et al.*, 2013). Lastly, previous RNA-Seq based studies of salinity acclimation in crabs which identified thousands of DEGs (Li *et al.*, 2014; Lv *et al.*, 2013) likely suffered from many false positives due to a lack of replication (see below), and although our level of biological replication was modest ($n = 3$), we feel the DEGs identified here represent a more realistic response to salinity at the transcriptomic level.

5.3. Suggested read depth for RNA-seq analyses of crustacean gills and other tissues

It is clear from rarefaction analyses (Fig. 3) that a subset of the ~28 M sequence reads per sample generated here would have produced qualitatively similar results across most DE analyses. For the pairwise comparison between anterior gills at 35‰ vs. posterior gills at 10‰, which produced the highest number of DEGs, a clear pattern of diminishing returns was evident when utilizing greater than 10 M reads per sample. Essentially, 15 M reads per sample would have been more than adequate to identify similar numbers of DEGs as when the complete dataset of ~28 M reads per sample were utilized.

Such benchmark datasets and recommendations are generally lacking in RNA-Seq studies and will likely need to be explored and established on a case-by-case basis (Sims *et al.*,

2014). For example, the Encyclopedia of DNA Elements Consortium (ENCODE) used similar rarefaction analyses to demonstrate that ~36 M reads were required to accurately quantify expression levels for ~80% of genes with an FPKM of at least 10, when based on an overall dataset of 214 M reads, from human embryonic stem cells (ENCODE Project Consortium, 2004, 2011; Trapnell *et al.*, 2009). Similarly, we recommend sequencing ~10-15 M reads per sample as a reasonable starting point when estimating DE and quantifying DEGs from the gills, and potentially other tissues, of crustaceans.

Given that a lack of biological replication is a problem in crustacean transcriptomic studies (Havird and Santos, 2016b) and can lead to erroneous interpretations of RNA-Seq studies (Havird and Santos, 2016a), recommending a more moderate number of sequence reads will hopefully lead to diverting resources towards increased biological replication in this field. For popular parametric based analyses of DE such as those used here (i.e, edgeR and DESeq), it is clear that limiting biological replication results in increased numbers of false positive DEGs (Tarazona *et al.*, 2011), which likely explains why previous, non-replicated RNA-Seq analyses yielded more DEGs than we identified (Li *et al.*, 2014; Lv *et al.*, 2013). For example, the number of DEGs for anterior vs. posterior gills of *C. sapidus* increased by more than eight-fold when a single biological replicate was randomly utilized in the same analytical pipeline employed here (data not shown).

What constitutes the “appropriate” number of biological replicates for inclusion in any given RNA-Seq study is less clear. Here, we utilized three biological replicates per treatment, which could have limited our ability to detect DEGs due to high inter-individual variation. Supporting this possibility, an RNA-Seq power analysis in Scotty (Busby *et al.*, 2013), with the data presented here suggested 5-7 biological replicates would be needed to identify > 50% of the

DEGs having a 5-fold difference in expression between our various treatments. Although additional biological replicates was not a possibility in the current study, it implies increased biological replication may yield higher numbers of DEGs. Notably, the Scotty analyses predicted that only ~13 M reads would need to be generated, supporting our rarefaction analysis (Fig. 3). Taken together, the moderately replicated study presented here serves as an example for generally emphasizing biological replication over read depth in RNA-Seq studies.

5.4. Towards a “core” gill transcriptome for decapods

Utilizing publicly available sequence data, we identified 476 well-annotated transcripts shared and expressed at appreciable levels among the gills of five decapod species. While many of these were associated with general cellular processes such as transcription and translation, which are likely highly expressed in most tissues, a number of the salinity-induced DEGs discussed above were components of this “core” gill transcriptome, including CAC, NKA α , and other ion transporters. This implies ion transport is a critical component of gill function, even in the stenohaline osmoconformers examined here such as *Hyas araneus* (Anger, 1985; Pfaff, 1997; Torres *et al.*, 2002). Also recovered were several members of the mitochondrial electron transport chain, further supporting the idea that osmoregulatory processes in the gills impose high energetic demands. Notably, genes not normally attributed to osmoregulation were also identified in the “core” gill transcriptome, including several stress-related genes such as superoxide dismutases (SODs), ubiquitin-related machinery, glutathione peroxidases, and HSPs. One interesting hypothesis for these results is that while such genes are not DE due to salinity, they are instead constitutively expressed at appreciable levels in order to cope with the environmental challenges experienced by the gills. Another possibility is that the high energetic

demands of gills induces elevated expression of such genes to cope with an abundance of reactive oxygen species (ROS) produced as a physiological byproduct.

Transcripts lacking annotation were abundant in the *C. sapidus* composite gill transcriptome, as was likely the case for the “core” gill transcriptome. Thus, numerous genes that may play critical roles in the gills of crustaceans like *C. sapidus* could not be attributed to specific biological functions. Unfortunately, annotation rates for crustacean transcriptomes are commonly low (Havird and Santos, 2016b), and the *Daphnia pulex* genome, although distantly related to the economically and ecologically important decapods, is itself meagerly annotated (Colbourne *et al.*, 2011). Given this, a key advancement to the future of crustacean transcriptomics studies will be to investigate and implement alternative annotation methods beyond the utilization of BLAST against curated databases, which is currently the most common practice (Havird and Santos, 2016b). Methods making use of Hidden Markov models or structural homology offer a promising alternative in this area.

6. Conclusions

Here, we examined patterns of differential gene expression related to gill type following salinity transfer in the blue crab *C. sapidus*, a well-studied model species for osmoregulation in crustaceans. Novel candidate genes underlying differences between functionally distinct gills were identified, including ion transporters and ones related to cellular structure, energy metabolism, and immune function. Potential candidates for the regulation of ion transport included neuropeptide Y, a KCTD16-like protein, and an ADIPOR-like receptor. Furthermore, expression patterns for previously identified genes known to play a role in osmoregulation were also generally supported. To utilize transcriptomic approaches to investigate further time points,

salinity transfers, gill pairs, or species in osmoregulatory physiology, our rarefaction analyses lead us to recommend a modest number of sequence reads (~10–15 M) and increased biological replication to ensure a robust experimental design. Finally, a “core” gill transcriptome for decapods was identified using comparative approaches among species with publically available data, suggesting elevated expression of osmoregulatory and stress-response proteins may be a common phenotype in crustacean gills. We hope the results presented here will encourage future studies of crustacean ecological genomics and ecophysiology via RNA-Seq.

Acknowledgements

We thank two anonymous reviewers for comments that improved the manuscript. We are also grateful to Donald Mykles, Karen Burnett, David Durica, and Jonathon Stillman for organizing the “Tapping the Power of Crustacean Transcriptomes to Address Grand Challenges in Comparative Biology” symposium at the 2016 meeting of the Society for Integrative and Comparative Biology and inviting us to present components of this work. This work was supported in part by the National Science Foundation (NSF) [DEB #0949855 and ANT #1043745 to SRS, and EPS #11-58862 to RPH]. JCH was supported in part by a National Institutes of Health Postdoctoral Fellowship (F32GM116361). This represents contributions #145 and #53 to the Auburn University (AU) Marine Biology Program and Molette Biology Laboratory for Environmental and Climate Change Studies, respectively.

Literature cited

Alfalah, M., Michel, M.C., 2004. Neuropeptide Y and related peptides. Springer, Berlin; New York.

600 Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.,
601 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.
602 Nucleic Acids Res 25, 3389-3402.

603 Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome
604 Biol 11, R106.

605 Anger, K., 1985. Influence of salinity on larval development of the spider crab, *Hyas araneus*,
606 reared in the laboratory, in: J.S. Gray, M.E. Christiansen (Eds.), Marine biology of polar regions
607 and effects of stress on marine organisms. Wiley, Chichester.

608 Apweiler, R., Jesus Martin, M., O'onovan, C., Magrane, M., 2012. Reorganizing the protein
609 space at the Universal Protein Resource (UniProt). Nucleic Acids Res 40, D71-75.

610 Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P.,
611 Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A.,
612 Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene
613 ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25, 25-
614 29.

615 Bliss, D.E., 1968. Transition from water to land in decapod crustaceans. Amer Zool 8, 355-392.

616 Bray, W.A., Lawrence, A.L., Leungtrujillo, J.R., 1994. The effect of salinity on growth and
617 survival of *Penaeus vannamei*, with observations on the interaction of Iahn virus and salinity.
618 Aquaculture 122, 133-146.

619 Brown, C.T., Howe, A., Zhang, Q., Pyrkosz, A.B., Brom, T.H., 2012. A reference-free
620 algorithm for computational normalization of shotgun sequencing data. BioArXiv:1203:4802.

- 621 Busby, M.A., Stewart, C., Miller, C.A., Grzeda, K.R., Marth G.T., 2013. Scotty: a web tool for
 622 designing RNA-Seq experiments to measure differential gene expression. *Bioinformatics* 29,
 623 656–657.
- 624 Cameron, J.N., 1978. NaCl balance in blue crabs, *Callinectes sapidus*, in fresh water. *J Comp*
 625 *Physiol B* 123, 127-135.
- 626 Cameron, J.N., Batterton, C.V., 1978. Antennal gland function in the freshwater blue crab,
 627 *Callinectes sapidus*: water, electrolyte, acid–base and ammonia excretion. *J Comp Physiol B*
 628 123.
- 629 Charmantier, G., Charmantier-Daures, M., Towle, D., 2009. Osmotic and ionic regulation in
 630 aquatic arthropods, in: D.H. Evans (Ed.), *Osmotic and ionic regulation: cells and animals*. CRC
 631 Press, Boca Raton, 165-230.
- 632 Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H.,
 633 Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., Bauer, D.J., Caceres, C.E., Carmel, L., Casola,
 634 C., Choi, J.H., Detter, J.C., Dong, Q., Dusheyko, S., Eads, B.D., Frohlich, T., Geiler-Samerotte,
 635 K.A., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E.V., Kultz, D., Laforsch,
 636 C., Lindquist, E., Lopez, J., Manak, J.R., Muller, J., Pangilinan, J., Patwardhan, R.P., Pitluck, S.,
 637 Pritham, E.J., Rechtsteiner, A., Rho, M., Rogozin, I.B., Sakarya, O., Salamov, A., Schaack, S.,
 638 Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D.,
 639 Tu, H., Vos, H., Wang, M., Wolf, Y.I., Yamagata, H., Yamada, T., Ye, Y., Shaw, J.R., Andrews,
 640 J., Crease, T.J., Tang, H., Lucas, S.M., Robertson, H.M., Bork, P., Koonin, E.V., Zdobnov, E.M.,
 641 Grigoriev, I.V., Lynch, M., Boore, J.L., 2011. The ecoresponsive genome of *Daphnia pulex*.
 642 *Science* 331, 555-561.

- 643 Copeland, D.E., Fitzjarrell, A.T., 1968. The salt absorbing cells in the gills of the blue crab
 644 (*Callinectes sapidus* Rathbun) with notes on modified mitochondria. *Z Zellforsch Mikrosk Anat*
 645 92, 1-22.
- 646 de la Vega, E., Hall, M.R., Wilson, K.J., Reverter, A., Woods, R.G., Degnan, B.M., 2007. Stress-
 647 induced gene expression profiling in the black tiger shrimp *Penaeus monodon*. *Physiol*
 648 *Genomics* 31, 126-138.
- 649 Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime,
 650 C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloe, D., Le Gall, C.,
 651 Schaeffer, B., Le Crom, S., Guedj, M., Jaffrezic, F., 2012. A comprehensive evaluation of
 652 normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief*
 653 *Bioinform* 14, 671-683.
- 654 ENCODE Project Consortium, 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project.
 655 *Science* 306, 636-640.
- 656 ENCODE Project Consortium, 2011. A user's guide to the encyclopedia of DNA elements
 657 (ENCODE). *PLoS Biol* 9, e1001046.
- 658 Engel, D.W., 1977. Comparison of the osmoregulatory capabilities of two portunid crabs,
 659 *Callinectes sapidus* and *C. similis*. *Mar Biol* 41, 275-279.
- 660 Fehsenfeld, S., Weihrauch, D., 2016. The role of an ancestral hyperpolarization activated cyclic
 661 nucleotide-gated K⁺-channel in branchial acid-base regulation in the green crab, *Carcinus*
 662 *maenas* (L.). *J Exp Biol* 219, 887-896.
- 663 Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity
 664 searching. *Nucleic Acids Res* 39, W29-37.

- 665 Freire, C.A., Onken, H., McNamara, J.C., 2008. A structure-function analysis of ion transport in
666 crustacean gills and excretory organs. *Comp Biochem Physiol A Mol Integr Physiol* 151, 272-
667 304.
- 668 Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-
669 generation sequencing data. *Bioinformatics* 28, 3150-3152.
- 670 Goolish, E.M., Burton, R.S., 1989. Energetics of osmoregulation in an intertidal copepod:
671 Effects of anoxia and lipid reserves on the pattern of free amino accumulation. *Funct Ecol* 3, 81-
672 89.
- 673 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X.,
674 Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N.,
675 di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011.
676 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat*
677 *Biotechnol* 29, 644-652.
- 678 Gutknecht, J., Bisson, M.A., Tosteson, F.C., 1977. Diffusion of carbon-dioxide through lipid
679 bilayer membranes - effects of carbonic-anhydrase, bicarbonate, and unstirred layers. *J Gen*
680 *Physiol* 69, 779-794.
- 681 Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger,
682 M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F.,
683 Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N.,
684 Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity
685 platform for reference generation and analysis. *Nat Protoc* 8, 1494-1512.

686 Harms, L., Frickenhaus, S., Schiffer, M., Mark, F.C., Storch, D., Held, C., Portner, H.O.,
 687 Lucassen, M., 2014. Gene expression profiling in gills of the great spider crab *Hyas araneus* in
 688 response to ocean acidification and warming. BMC Genomics 15:789.

689 Harms, L., Frickenhaus, S., Schiffer, M., Mark, F.C., Storch, D., Portner, H.O., Held, C.,
 690 Lucassen, M., 2013. Characterization and analysis of a transcriptome from the boreal spider crab
 691 *Hyas araneus*. Comp Biochem Physiol Part D Genomics Proteomics 8, 344-351.

692 Havird, J.C., Henry, R.P., Wilson, A.E., 2013. Altered expression of Na⁺/K⁺-ATPase and other
 693 osmoregulatory genes in the gills of euryhaline animals in response to salinity transfer: A meta-
 694 analysis of 59 quantitative PCR studies over 10 years. Comp Biochem Physiol Part D Genomics
 695 Proteomics 8, 131-140.

696 Havird, J.C., Santos, S.R., 2016a. Developmental transcriptomics of the Hawaiian anchialine
 697 shrimp *Halocaridina rubra* Holthuis, 1963 (Crustacea: Atyidae). Integr Comp Biol. In press.

698 Havird, J.C., Santos, S.R., 2016b. Here we are, but where do we go? A systematic review of
 699 crustacean transcriptomic studies from 2014-2015. Integr Comp Biol. In press.

700 Havird, J.C., Santos, S.R., Consortium, G.R.D., 2014a. Genomic Resources Notes accepted 1
 701 June 2014-31 July 2014. Molecular Ecology Resources 14, 1322-1322.

702 Havird, J.C., Santos, S.R., Henry, R.P., 2014b. Osmoregulation in the Hawaiian anchialine
 703 shrimp *Halocaridina rubra* (Crustacea: Atyidae): expression of ion transporters, mitochondria-
 704 rich cell proliferation and hemolymph osmolality during salinity transfers. J Exp Biol 217, 2309-
 705 2320.

706 Henry, R.P., 1984. The role of carbonic anhydrase in blood ion and acid-base regulation. Amer
 707 Zool 24, 241-253.

- 708 Henry, R.P., 1988. Subcellular distribution of carbonic anhydrase activity in the gills of the blue
709 crab, *Callinectes sapidus*. J Exp Zool 245, 1-8.
- 710 Henry, R.P., 2001. Environmentally mediated carbonic anhydrase induction in the gills of
711 euryhaline crustaceans. J Exp Biol 204, 991-1002.
- 712 Henry, R.P., 2005. Critical salinity, sensitivity, and commitment of salinity-mediated carbonic
713 anhydrase induction in the gills of two euryhaline species of decapod crustaceans. J Exp Zool A
714 Comp Exp Biol 303, 45-56.
- 715 Henry, R.P., Cameron, J.N., 1982. The distribution and partial characterization of carbonic
716 anhydrase in selected aquatic and terrestrial decapod crustaceans. J Exp Zool 221, 309-321.
- 717 Henry, R.P., Cameron, J.N., 1983. The role of branchial carbonic anhydrase in respiration, ion
718 regulation and acid-base balance in the aquatic crab, *Callinectes sapidus* and the terrestrial crab,
719 *Gecarcinus lateralis*. J Exp Biol 103, 205-223.
- 720 Henry, R.P., Campoverde, M., 2006. Neuroendocrine regulation of carbonic anhydrase
721 expression in the gills of the euryhaline green crab, *Carcinus maenas*. J Exp Zool A Comp Exp
722 Biol 305, 663-668.
- 723 Henry, R.P., Gehnrich, S., Weihrauch, D., Towle, D.W., 2003. Salinity-mediated carbonic
724 anhydrase induction in the gills of the euryhaline green crab, *Carcinus maenas*. Comp Biochem
725 Physiol A Mol Integr Physiol 136, 243-258.
- 726 Henry, R.P., Lucu, C., Onken, H., Weihrauch, D., 2012. Multiple functions of the crustacean gill:
727 osmotic/ionic regulation, acid-base balance, ammonia excretion, and bioaccumulation of toxic
728 metals. Front Physiol 3:431.

- 729 Henry, R.P., Watts, S.A., 2001. Early carbonic anhydrase induction in the gills of the blue crab,
730 *Callinectes sapidus*, during low salinity acclimation is independent of ornithine decarboxylase
731 activity. J Exp Zool 289, 350-358.
- 732 Hobbs, H.H., Hart, C.W., 1982. The shrimp genus *Atya* (Decapoda: Atyidae). Smithsonian
733 Institution Press, Washington.
- 734 Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R.,
735 Baseler, M.W., Lane, H.C., Lempicki, R.A., 2007a. The DAVID Gene Functional Classification
736 Tool: a novel biological module-centric algorithm to functionally analyze large gene lists.
737 Genome Biol 8, R183.
- 738 Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R.,
739 Baseler, M.W., Lane, H.C., Lempicki, R.A., 2007b. DAVID Bioinformatics Resources:
740 expanded annotation database and novel algorithms to better extract biology from large gene
741 lists. Nucleic Acids Res 35, W169-175.
- 742 Langmead, B., 2010. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics
743 Chapter 11, Unit 11, 17.
- 744 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9,
745 357-359.
- 746 Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or
747 without a reference genome. BMC Bioinformatics 12, 323.
- 748 Li, E., Wang, S., Li, C., Wang, X., Chen, K., Chen, L., 2014. Transcriptome sequencing revealed
749 the genes and pathways involved in salinity stress of Chinese mitten crab, *Eriocheir sinensis*.
750 Physiol Genomics 46, 177-190.

- 751 Li, L., Stoeckert, C.J., Jr., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for
752 eukaryotic genomes. *Genome Res* 13, 2178-2189.
- 753 Liu, Z., Xiang, Y., Sun, G., 2007. The KCTD family of proteins: structure, function, disease
754 relevance. *Cell Biosci* 3, 45.
- 755 Lovett, D.L., Colella, T., Cannon, A.C., Lee, D.H., Evangelisto, A., Muller, E.M., Towle, D.W.,
756 2006a. Effect of salinity on osmoregulatory patch epithelia in gills of the blue crab *Callinectes*
757 *sapidus*. *Biol Bull* 210, 132-139.
- 758 Lovett, D.L., Verzi, M.P., Burgents, J.E., Tanner, C.A., Glomski, K., Lee, J.J., Towle, D.W.,
759 2006b. Expression profiles of Na⁺,K⁺-ATPase during acute and chronic hypo-osmotic stress in
760 the blue crab *Callinectes sapidus*. *Biol Bull* 211, 58-65.
- 761 Lucu, C., Towle, D.W., 2003. Na(+)+K(+)-ATPase in gills of aquatic crustacea. *Comp Biochem*
762 *Physiol A Mol Integr Physiol* 135, 195-214.
- 763 Luquet, C.M., Weihrauch, D., Senek, M., Towle, D.W., 2005. Induction of branchial ion
764 transporter mRNA expression during acclimation to salinity change in the euryhaline crab
765 *Chasmagnathus granulatus*. *J Exp Biol* 208, 3627-3636.
- 766 Luvizotto-Santos, R., Lee, J., Branco, Z., Bianchini, A., Nery, L., 2003. Lipids as energy source
767 during salinity acclimation in the euryhaline crab *Chasmagnathus granulata* Dana, 1851
768 (crustacea-grapsidae). *J Exp Zool A Comp Exp Biol* 295, 200-205.
- 769 Lv, J., Liu, P., Wang, Y., Gao, B., Chen, P., Li, J., 2013. Transcriptome analysis of *Portunus*
770 *trituberculatus* in response to salinity stress provides insights into the molecular basis of
771 osmoregulation. *PLoS One* 8, e82155.

- MacManes, M.D., 2014. On the optimal trimming of high-throughput mRNA sequence data. Front Genet 5, 13.
- Mantel, L.H., Farmer, L.L., 1983. Osmotic and ionic regulation, in: L.H. Mantel (Ed.), The Biology of the Crustacea, vil. V, Internal Anatomy and Physiological Regulation Academic Press, New York, 53-161.
- McNamara, J.C., Faria, S.C., 2012. Evolution of osmoregulatory patterns and gill ion transport mechanisms in the decapod Crustacea: a review. J Comp Physiol B Biochem Systemic Environment Physiol 182, 997-1014.
- Mitchell, R.T., Henry, R.P., 2014. Carbonic anhydrase induction in euryhaline crustaceans is rate-limited at the post-transcriptional level. Comp Biochem Physiol A Mol Integr Physiol 169, 15-23.
- Mohd-Shamsudin, M.I., Kang, Y., Lili, Z., Tan, T.T., Kwong, Q.B., Liu, H., Zhang, G., Othman, R.Y., Bhassu, S., 2013. In-depth transcriptomic analysis on giant freshwater prawns. PLoS One 8, e60839.
- Neufeld, G.J., Holliday, C.W., Pritchard, J.B., 1980. Salinity adaptation of gill Na,K-ATPase in the blue crab *Callinectes sapidus*. J Exp Zool 211, 215-224.
- Onken, H., Tresguerres, M., Luquet, C.M., 2003. Active NaCl absorption across posterior gills of hyperosmoregulating *Chasmagnathus granulatus*. J Exp Biol 206, 1017-1023.
- Pfaff, K., 1997. Einfluß der Salinität auf den Stoffbestand der Larvenstadien einer marinen Dekapodenart. MSc thesis. University of Darmstadt, Darmstadt.
- Piller, S., Henry, R., Doeller, J., Kraus, D., 1995. A comparison of the gill physiology of two euryhaline crab species, *Callinectes sapidus* and *Callinectes similis*: energy production,

794 transport-related enzymes and osmoregulation as a function of acclimation salinity. J Exp Biol
 795 198, 349-358.

796 Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T.,
 797 Letunic, I., Doerks, T., Jensen, L.J., von Mering, C., Bork, P., 2012. eggNOG v3.0: orthologous
 798 groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res 40, D284-
 799 289.

800 Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund,
 801 K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A.,
 802 Finn, R.D., 2012. The Pfam protein families database. Nucleic Acids Res 40, D290-301.

803 Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for
 804 differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140.

805 Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y.,
 806 Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J.M., Nabholz, B., Roux, C.,
 807 Tsagkogeorga, G., Weber, A.A., Weinert, L.A., Belkhir, K., Bierne, N., Glemin, S., Galtier, N.,
 808 2014. Comparative population genomics in animals uncovers the determinants of genetic
 809 diversity. Nature 515, 261-263.

810 Salichos, L., Rokas, A., 2011. Evaluating ortholog prediction algorithms in a yeast model clade.
 811 PLoS One 6, e18755.

812 Serrano, L., Halanych, K.M., Henry, R.P., 2007. Salinity-stimulated changes in expression and
 813 activity of two carbonic anhydrase isoforms in the blue crab *Callinectes sapidus*. J Exp Biol 210,
 814 2320-2332.

- 815 Serrano, L., Henry, R.P., 2008. Differential expression and induction of two carbonic anhydrase
816 isoforms in the gills of the euryhaline green crab, *Carcinus maenas*, in response to low salinity.
817 Comp Biochem Physiol Part D Genomics Proteomics 3, 186-193.
- 818 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N.,
819 Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of
820 biomolecular interaction networks. Genome Res 13, 2498-2504.
- 821 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H.,
822 Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of
823 high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7, 539.
- 824 Sims, D., Sudbery, I., Illott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and
825 coverage: key considerations in genomic analyses. Nat Rev Genet 15, 121-132.
- 826 Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., Conesa, A., 2011. Differential
827 expression in RNA-seq: a matter of depth. Genome Res 21, 2213-2223.
- 828 Taylor, H.H., Taylor, E.W., 1992. Gills and lungs: the exchange of gases and ions, in: F.
829 Harrison, A.G. Humes (Eds.), Microscopic Anatomy of Invertebrates, Vol. 10, Decapoda,
830 Crustacea John Wiley & Sons, New York, 203-293.
- 831 Torres, G., Gimenez, L., Anger, K., 2002. Effects of reduced salinity on the biochemical
832 composition (lipid, protein) of zoea 1 decapod crustacean larvae. J Exp Mar Biol Ecol 277, 43-
833 60.
- 834 Towle, D., Weihrauch, D., 2001. Osmoregulation by gills of euryhaline crabs: Molecular
835 analysis of transporters. Amer Zool 41, 770-780.

836 Towle, D.W., Paulsen, R.S., Weihrauch, D., Kordylewski, M., Salvador, C., Lignot, J.-H.,
837 Spaning-Pierrot, C., 2001. $\text{Na}^+\text{--K}^+$ -ATPase in gills of the blue crab *Callinectes sapidus*: cDNA
838 sequencing and salinity-related expression of α -subunit mRNA and protein. J Exp Biol 204,
839 4005-4012.

840 Towle, D.W., Henry, R.P., Terwilliger, N.B., 2011. Microarray-detected changes in gene
841 expression in gills of green crabs (*Carcinus maenas*) upon dilution of environmental salinity.
842 Comp Biochem Physiol Part D Genomics Proteomics 6, 115-125.

843 Towle, D.W., Kays, W.T., 1986. Basolateral localization of Na^+/K^+ -ATPase in gill epithelium of
844 two osmoregulating crabs, *Callinectes sapidus* and *Carcinus maenas*. J Exp Zool 239, 311-318.

845 Towle, D.W., Palmer, G., Harris, J., 1976. Role of gill $\text{Na}^+ + \text{K}^+$ -dependent ATPase in acclimation
846 of blue crabs (*Callinectes sapidus*) to low salinity. J Exp Zool 196, 315-322.

847 Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-
848 Seq. Bioinformatics 25, 1105-1111.

849 Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg,
850 S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals
851 unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28,
852 511-515.

853

854

855 **Table 1.** Assembly statistics for crustacean transcriptomes examined in this study. For *Callinectes sapidus*, abbreviations represent
 856 anterior (AG) or posterior (PG) gills and low (10‰) or high (35‰) salinity, respectively.

Species	SRA Accession	# Reads (x 10 ⁶)	# Contigs	% Contigs Annotated via BLASTx	Av. Contig Length (bp)	Max Length (bp)	N50 (bp)	# Contigs Following RSEM Filter ¹
<i>Callinectes sapidus</i>	PRJNA315576							
Composite		131.9	239158	35.4	1165.1	19475	2480	7380
AG High		30.0	90183	27.1	940.3	22183	1928	NA
AG Low		30.0	91703	26.3	893.7	22179	1819	NA
PG High		30.0	85386	28.4	824.9	17073	1581	NA
PG Low		30.0	80968	22.0	727.6	16398	1297	NA
<i>Carcinus aestuarii</i>	SRR1324847	37.1	48355	35.6	750.8	16936	1334	3920
<i>Hyas araneus</i>	ERR220388, ERR220390, ERR220389	26.5	59429	23.1	494.2	11101	592	14209
<i>Macrobrachium rosenbergii</i>	SRR345608	17.7	95321	20.4	643.5	13611	1000	18636
<i>Portunus trituberculatus</i>	SRR1013696, SRR1013694, SRR1013695	30.0	123015	27.8	1186.6	36358	2817	9388

¹#Contigs with FPKM ≥ 10

Table 2. Numbers of differentially expressed (DE) genes, their annotation rates, and general categories of upregulation and downregulation for all pairwise comparisons of anterior (AG) or posterior (PG) gills and low (10‰) or high (35‰) salinity in *Callinectes sapidus*. Up- vs downregulation is in reference to the first treatment in the pair.

Comparison	#DE Genes	# Annotated	Upregulated categories	Downregulated categories
PG High vs. PG Low	94	16	Ion transport	Energy metabolism, venom carboxyl esterase NA
AG Low vs. PG Low	13	4	ADIPOR-like receptor, NKA β	
AG High vs. PG High	109	38	Virus replication, ion transport	Structural/cuticle genes
AGHigh vs. AG Low	78	13	Virus replication	Structural/cuticle genes
AGHigh vs. PGLow	380	97	Ion transport, energy/mitochondrial genes	Structural/cuticle genes
AGLow vs. PGHigh	96	22	Virus replication	Collagen

Fig. 1. Heat map summarizing differentially expressed genes between anterior (AG) or posterior (PG) gills and low (10‰) or high (35‰) salinity treatments in *Callinectes sapidus*. Results of a cluster analysis are presented to the left of the heat map and show groups with similar expression patterns. A cluster key is provided which shows manual interpretation of the conditions in which a particular gene cluster was upregulated. Genes of interest that are discussed in the text are highlighted with red branches and identified to the right of the heat map based on their annotation. Heat map and cluster analysis was generated with the Trinity software suite (Haas *et al.*, 2013).

Fig. 2. Expression levels for genes not differentially expressed at statistically significant levels but of general interest and discussed in the text. Abbreviations: NKA α , Na⁺/K⁺-ATPase α ; NKCC, Na⁺/K⁺/2Cl⁻-cotransporter; NHE2, Na⁺/H⁺ exchanger 2; HSP, heat shock protein; SOD, superoxide dismutase; GST, glutathione S-transferase; GPx, glutathione peroxidase.

Fig. 3. Rarefaction analysis of the number of differentially expressed genes recovered from all pairwise comparisons of anterior (AG) or posterior (PG) gills and low (10‰) or high (35‰) salinity in *Callinectes sapidus* as a function of subsampling different numbers of sequence reads from each biological sample.

Fig. 4. Gene-set enrichment network depicting functional clusters of genes identified in the “core” gill transcriptome of the five decapod species examined in this study. Only 158 of the 476 transcripts included in the “core” gill transcriptome were able to be placed in the network (shown as nodes) based on functional annotation analyses using the Database for Annotation, Visualization and Integrated Discovery (DAVID) and default visualization parameters in Enrichment Map. The “Word Cloud” plugin to Enrichment Map was utilized in identifying common terms found in the functional annotations of the genes of each cluster (color-coded by their associated functional annotation). Each node represents a transcript that is shared and expressed at relatively high levels in all five decapods, and edges represent functional connections between transcripts. See text for additional details.