

Construction and characterization of EST libraries from the porcelain crab, *Petrolisthes cinctipes*

Jonathon H. Stillman,^{1,*} Kristen S. Teranishi,[†] Abderrhamane Tagmount,^{*} Erika A. Lindquist,[‡] and Peter B. Brokstein[‡]

^{*}Romberg Tiburon Center, San Francisco State University, 3152 Paradise Drive, Tiburon, CA 94920, USA;

[†]Department of Zoology, University of Hawaii at Manoa, Honolulu, HI 96822, USA; [‡]Joint Genome Institute, DOE, 2800 Mitchell Drive, Bldg 400-467, Walnut Creek, CA 94598, USA

Synopsis The thermal phenotype of an organism (heat and cold tolerance, thermal range, and thermal plasticity) is an essential feature of how the organism performs across thermal environments and in response to thermal stress. Porcelain crabs are of interest in addressing questions of thermal phenotype because of their high species diversity and the large variation in thermal phenotype among species, as well as the biogeographic patterning of these crabs along environmental stress gradients. We are studying the cellular bases of thermal phenotype and physiological responses to environmental stress using a functional genomics cDNA microarray approach. To do this, we have isolated total RNA from a range of tissues from 1 species of porcelain crab (*Petrolisthes cinctipes*) exposed to a suite of thermal conditions, and have used this RNA to construct a 13 824-clone EST library. Here, we describe construction, EST sequencing, assembly and clustering, and results of BLASTx homology search for our initial 13 824-clone library. From 12 060 usable ESTs, 6717 consensus sequences were identified, and roughly 50% of these have homology to known proteins. At present, an additional 50 000–75 000-clone library of *P. cinctipes* ESTs is being generated, with the aim of developing a library with near-complete coverage of the transcriptome. The libraries and sequence information that will be generated as a result of this project should be of value for crustacean biologists working across a broad range of scientific disciplines (for example, physiology, developmental biology, biological rhythms, ecology, fisheries biology), as well as in studies of molecular evolution and phylogeography.

Introduction

Genomics-based investigative approaches have assumed a major role in advances across a wide range of disciplines, from the biomedical field to ecology, evolution, and comparative physiology (Alizadeh and others 2000; Feder and Mitchell-Olds 2003; Gracey and Cossins 2003; Hofmann and others 2005). Sequencing genomes from model organisms such as *Caenorhabditis elegans* and *Drosophila melanogaster* set the stage for rapid progress in characterizing gene function and regulation (*C. elegans* Sequencing Consortium 1998; Adams and others 2000). It is not, however, the general case that these model organisms are necessarily the select study systems for the comparative physiologist (Gracey and Cossins 2003).

The marine environment provides an interesting setting for addressing ecological and comparative physiological questions such as how ecosystems or organisms and their distributional limits might be affected by global environmental change (Stillman 2003; Hofmann and others 2005). At present, full genome sequence

information for marine organisms like *Fugu rubripes* (Japanese pufferfish) (Aparicio and others 2002) and *Ciona intestinalis* (sea squirt) (Dehal and others 2002) are available. Additionally, many laboratories have opted to construct expressed sequence tag (EST) libraries from marine study systems for use in addressing problems spanning ecotoxicology (for example, mussels [Vernier and others 2003] and copepods as bioindicators of pollution [Lee and others 2005]), immunology (disease resistance in shrimp [Supungul and others 2002], immune response in oysters [Gueguen and others 2003]), muscle physiology (for example, muscle growth in scallops [Roberts and Goetz 2003]), and pharmacology (for example, cone snail conotoxins [Pi and others 2006]). The range of organisms being studied at the gene expression level provides fruitful grounds for comparative work as well.

Recently, a hub for EST information and microarray data specific to marine organisms was developed by the Marine Genomics project (McKillen and others 2005). Databases for over 19 marine species are accessible or

From the symposium “Genomic and Proteomic Approaches in Crustacean Biology” presented at the annual meeting of the Society for Integrative and Comparative Biology, January 4–8, 2006, at Orlando, Florida.

¹E-mail: stillmaj@sfsu.edu

Integrative and Comparative Biology, pp. 1–12

doi:10.1093/icb/icl007

© The Author 2006. Published by Oxford University Press on behalf of the Society for Integrative and Comparative Biology. All rights reserved.

For permissions please email: journals.permissions@oxfordjournals.org.

underway (<http://www.marinegenomics.org>) and include those for 7 crustaceans *Callinectes sapidus* (Blue crab, 1742 ESTs), *Homarus americanus* (American Atlantic lobster, 5043 ESTs), *Litopenaeus setiferus* (White shrimp, 1041 ESTs), *Litopenaeus stylirostris* (Blue shrimp, 227 ESTs), *Litopenaeus vannamei* (White shrimp, 13 704 ESTs), *Palaemonetes pugio* (Daggerblade grass shrimp, 8821 ESTs), and *Calanus finmarchicus* (North Atlantic copepod, 309 ESTs) (McKillen and others 2005).

Here, we present our efforts for developing an EST database for the porcelain crab, *Petrolisthes cinctipes*, for use in comparative functional genomic analyses of mechanistic bases of thermal adaptation as well as patterns of thermal stress responses across ecological gradients. Development of this porcelain crab genomics resource allows us to take advantage of the great biological and ecological diversity of the porcelain crabs, and builds on a wealth of data on comparative thermal physiology for these organisms. Porcelain crabs of the genus *Petrolisthes* (Decapoda: Anomura: Porcellanidae) comprise over 100 species spanning a wide latitudinal range and includes species inhabiting discrete vertical zones within geographical ranges including the North Temperate, Northern Gulf of California, Tropical, and South Temperate regions of the Eastern Pacific coast (Haig 1960; Stillman 2002). A phylogeny of porcelain crabs from the Eastern Pacific has been conducted allowing for comparative analyses to be made in an evolutionary context (Stillman and Reeb 2001). *P. cinctipes* has been an effective model for research in thermal physiology (Stillman 2002, 2003). This species inhabits the mid to upper intertidal zone in the northeastern Pacific and experiences temperatures as high as 31°C and as low as -1°C during low tide emersion in the summer and winter, respectively (Stillman 2002, 2004). We have constructed from *P. cinctipes* a cDNA library for use in studying the mechanisms, at the gene expression level (using microarray analyses), that underlie organismal thermal phenotypes.

Here, we present the methodology used, and results of our initial library construction, sequencing, and annotation. Then, we describe our present efforts, in conjunction with the Joint Genome Institute (JGI), to expand the *P. cinctipes* EST library to cover as much of the transcriptome as possible.

Methods

Specimen collection and handling

Our cDNA libraries were constructed from specimens of 1 species of porcelain crab (*P. cinctipes*) that was subjected to a wide array of experimental conditions

before sacrifice. In our experiments, we collected crabs from across latitudinal and seasonal gradients and exposed these animals to acute heat and cold shocks, generally to ~30 and ~0°C, respectively. We also acclimated crabs to a wide range of thermal conditions that elicit changes in whole organism thermal phenotype, and exposed these animals to acute heat and cold shocks. Acclimation temperatures ranged from 8 to 25°C, and acclimation duration ranged from 2 days to 2 months. From hundreds of individual crabs used across 2 years of experiments, we preserved RNA by snap-freezing whole crabs by freeze-clamp, or by dissecting fresh tissues into 1 ml Trizol reagent (Invitrogen) and then storing frozen until RNA extraction. RNA degradation was not generally observed in Trizol preserved samples.

RNA extraction

Total RNA was extracted by either powdering tissue under liquid N₂ and then thawing the powder in Trizol (whole crabs), or by homogenizing dissected tissues (heart, gill, nerve axons from walking legs (including both sensory and motor neurons), hepatopancreas, and claw muscle tissues) in 1 ml Trizol using a powered rotor-stator homogenizer. Homogenates were allowed to sit at room temperature for 15 min to ensure dissociation of nucleoprotein complexes, and centrifuged to remove cellular debris. Supernatants (up to 1 ml volume) were removed to fresh microcentrifuge tubes, mixed with 200 µl chloroform by vortex for 30 s, and centrifuged for 15 min at 16 000 g. The aqueous (top) layer was removed, mixed with 250 µl isopropanol and 250 µl high salt precipitation solution (0.8 M Na citrate, 1.2 M NaCl), and incubated at -20°C overnight. Precipitated RNA was pelleted by centrifugation at 16 000 g for 45 min at 4°C. Pellets were washed in 60% EtOH and resuspended in 1 mM Na citrate, pH 6.4. RNA concentration and purity were determined spectrophotometrically. Generally, A₂₆₀/A₂₈₀ ratios were ≥1.9, and concentrations ranged from 1 to 1000 ng/µl, depending on tissue type and starting tissue quantity. Following quantification of total RNA, we mixed equal amounts of RNA from each tissue type of each individual together to make 7 pooled RNA samples: 2 from heart (from different sets of RNA extracts), and 1 each from gill, nerve, hepatopancreas, muscle, and whole crabs.

cDNA library construction

From each of the 7 pooled RNA samples, we constructed a cDNA library using the BD Clontech SMART cDNA library construction kit. First strand synthesis of cDNA, long-distance PCR (LD-PCR) for

synthesis of full-length ds cDNA, and *Sfi*I digestion were conducted according to BD Clontech Protocol#PT3000-1, version#PR15738 except for the following: (1) PCR cleanup using Qiagen's QIAquick PCR Purification Kit (QIAquick Spin Handbook, 7/2002) replaced the proteinase K digestion step and (2) 1% xylene cyanol dye was not added immediately after *Sfi*I digestion. LD-PCR was generally stopped at 14–16 cycles, ~2 cycles before saturation of PCR product amplification (as analyzed by agarose gel electrophoresis). Following *Sfi*I digestion, samples were ethanol precipitated (2 volumes 100% ethanol, 1/10 volume 3 M Na acetate pH 5.2), resuspended in 25 μ l Qiagen buffer EB (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), and run out on a 0.75% agarose gel. cDNAs ranging from 500 to 5000 bp (most PCR products were 0.5–2 kb) were extracted using Qiagen's QIAquick Gel Extraction Kit (QIAquick Spin Handbook 7/2002). Purified cDNAs were directionally cloned into the *Sfi*I sites of either pTriplEx vector (Clontech, GenBank Accession #U39779), or a modified pTrueBlue vector, pTB (Genomics One TBP0527, a gift from Dr Andrew Gracey) in a ligation reaction using 0.5 μ l T4 DNA ligase, 0.5 μ l T4 DNA ligase buffer (New England BioLabs M0202S), 15–40 ng cDNA, and 0.5 μ l empty vector in 5 μ l at 16°C overnight. The ligation reaction was precipitated with 200 μ l isobutanol, placed on ice for 2 h, washed with 70% ethanol, and resuspended in 5 μ l water.

Transformation of 1 μ l of the ligation reaction into ElectroMAX DH10B *Escherichia coli* competent cells (Invitrogen, Cat No. 18290-015) was accomplished by electroporation. Following 1 h growth in SOC medium at 37°C, cells were plated on selective blue-white screening LB-agar medium with X-gal, IPTG, and carbenicillin overnight at 37°C. Individual colonies were handpicked with sterile wooden toothpicks into wells of Nunc 384-well plates each containing 50 μ l of selective LB with ampicillin. A total of 36 384-well plates (13 824 total colonies) were picked, and this number was divided by library as in Table 1.

Characterization of cDNA libraries

PCR analysis of cloned cDNAs

PCR was used to amplify cDNAs from each colony using vector specific primers to confirm that a single EST was cloned and for purposes of microarray printing. Primers used for pTriplEx were: 5'-CTCGGGAAGCGCGCCATTGTGTTGGT (forward) and 5'-ATACGTCTCACTATAGGGCGAATTGGCC (reverse), and for pTB were: 5'-ACAGGAGCAAAAAC-CATGGTCG (forward) and 5'-CGGGCTCTAGATCC-GGAGT (reverse). Overnight bacterial cultures were

Table 1 Library composition by tissue type

Tissue	Number of clones	% Total of library
Heart (2) ¹	4992	36.1
Nerve	768	5.6
Whole crabs	2304	16.7
Hepatopancreas	2304	16.7
Gill	2304	16.7
Claw muscle	1152	8.3

¹A total of 2 libraries were made from different sets of heart tissue RNA extracts.

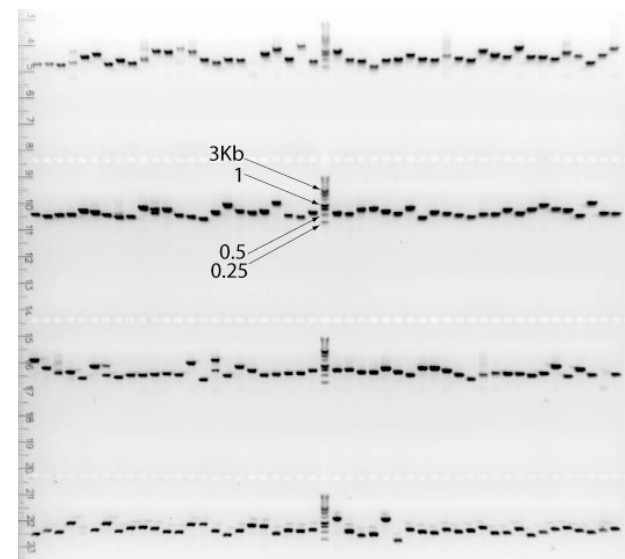


Fig. 1 PCR products stained by ethidium bromide and separated by 1% agarose gel electrophoresis shown by 1 representative gel. On this gel there are 192 PCR products, representing half of 1 of the heart plates. Each row on the gel contains a single lane of 1 Kb ladder (sizes indicated in the second row on the gel). This is 1 of 72 such gels that were run to characterize the library.

used to seed 30 μ l PCRs in 384-well PCR plates using an MJ-research DNA engine thermal cycler. PCR conditions were 94°C for 5 min, followed by 40 cycles of 92°C for 15 s, 54°C for 30 s, and 71.5°C for 1 min, with an additional 7 min at 71.5°C at the end of the program for any incomplete PCR products to be finished. The 13 824 PCR products were analyzed by 1% agarose gel electrophoresis (Fig. 1).

Sequencing of cloned cDNAs

Each cloned EST was sequenced by the JGI 2006 Community Sequencing Program (CSP). Sequences were generated from overnight cultures using rolling circle amplification with TempliPhi and sequenced on an ABI3730xl following standard JGI internal protocols (<http://www.jgi.doe.gov/>)

sequencing/protocols/protos_production.html). Every clone was sequenced from both the 3' and 5'-ends using 1 of the appropriate PCR primers (above). Some clones were sequenced more than once. Raw sequence traces automatically enter the JGI EST Pipeline, as described in the following 3 paragraphs.

The JGI EST Pipeline begins with the cleanup of DNA sequences derived from the 5' and 3'-end reads from a library of cDNA clones. The Phred software (Ewing and Green 1998; Ewing and others 1998) is used to call the bases and generate quality scores. Vector, linker, adapter, poly-A/T, and other artifact sequences are removed using the Cross_match software (Ewing and Green 1998; Ewing and others 1998), as well as a short pattern-finder developed internally at JGI. Low quality regions of the read are identified using JGI software that masks regions with a combined quality score of <15. The longest high quality region of each read is used as the EST. ESTs <150 bp, or containing common contaminants (for example, rRNA, mitochondrial DNA, *E. coli*, common vectors, and sequencing standards) are also removed from the dataset.

EST clustering is performed *ab initio*, based on alignments between each pair of ESTs. Pairwise EST alignments are generated using the Malign software (J. Chapman and others, unpublished), a modified version of the Smith–Waterman algorithm (Smith and Waterman 1981), which was developed at the JGI for use in whole-genome shotgun assembly. ESTs sharing an alignment of at least 98% identity, and 150 bp overlap are assigned to the same cluster. All alignments generated by Malign are restricted such that they will always extend to within a few bases of the ends of both ESTs. These are relatively strict clustering cutoffs, and are intended to avoid placing divergent members of gene families in the same cluster. However, these clustering cutoffs could also have the effect of separating splice variants into different clusters. Optionally, ESTs that do not share alignments can be assigned to the same cluster, if they are derived from the same cDNA clone.

For each cluster of EST sequences, cluster consensus sequences are generated by running the Phrap software (Ewing and Green 1998; Ewing and others 1998) on the ESTs comprising each cluster. This matches well with the directed sequencing assumptions underlying the Phrap algorithm, as each cluster comprises a clean “tiling path”, which can be easily assembled. Additional improvements were made to the Phrap assemblies by using the “forcelevel 4” option, which decreases the chances of generating multiple consensi for a single cluster, where the consensi differ only by sequencing errors (P. Brokstein, personal observation).

Protein homology for each consensus sequence was determined using BLASTx to search against 3 different databases: GenBank non-redundant (nr), Swiss-Prot, and the Gene Ontology (GO) database. BLASTx was run using NCBI Blast 2.2.6. In order to ensure consistency of BLASTx e-value between BLASTx runs on databases of varying size, the $-Y$ parameter was set to $1.75e12$. Data used for nr and Swiss-Prot BLAST were downloaded from GenBank on September 27, 2005, and the GO representative sequence file (go_200510-seqdb.fasta) was downloaded from ftp://ftp.geneontology.org/pub/go/godatabase/archive/full/latest. Classification of GO terms (Fig. 2) was done using basic MS Excel search functions.

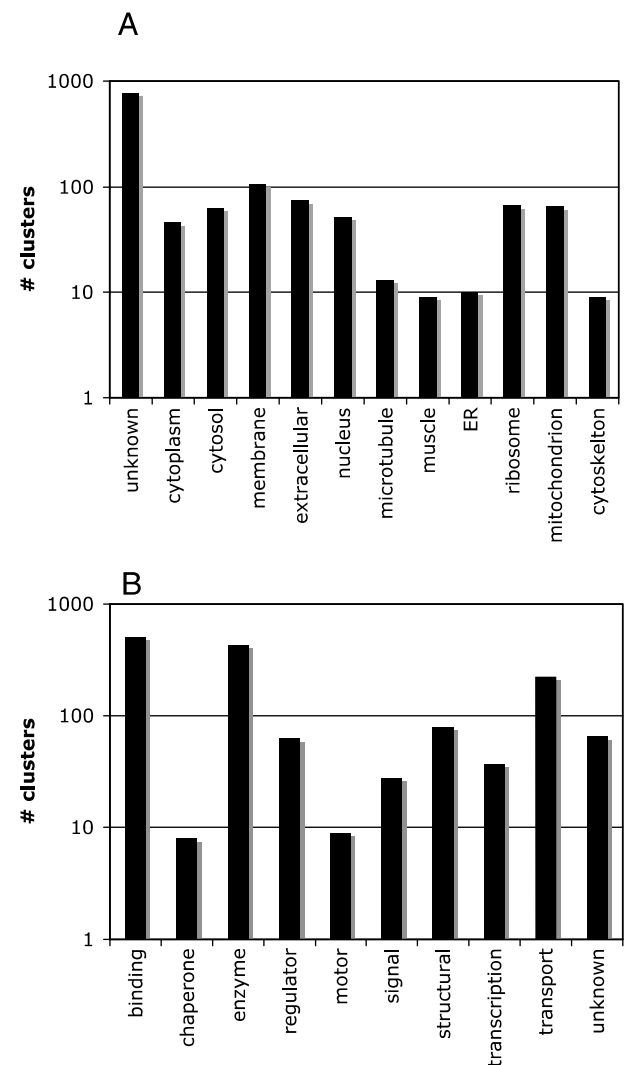


Fig. 2 Categorization of gene ontology (GO) terms for cellular compartment (A) and molecular function (B) based on 1289 strong matches of clusters (Table 3). Each, “unknown” category includes clusters that had GO matches that stated “unknown” as well as clusters for which there was no GO term stated for compartment or function.

Results and discussion

Insert size characterization

Overall, the smallest PCR products were ~ 250 bp, the largest were ~ 2000 bp, and the average size was 1000 bp (Fig. 1). Approximately 13 271 (96%) of the clones yielded a single PCR product, ~ 300 of the 13 824 clones did not produce a PCR product, and ~ 235 produced multiple PCR products.

Sequence analysis

From these clones picked from the 7 libraries that were constructed (Table 1), a total of 35 232 sequencing reactions were performed and run on the ABI sequencer. Following quality control assessment, 22 463 sequences of high quality, representing 12 062 clones, were obtained (Fig. 1). For 1762 clones, sequence quality was either poor (for example, failed sequencing reactions, or the 235 cases where no or multiple PCR bands were observed) (Fig. 1), or represented empty vector or a cloned bacterial gene. There was no further processing of sequence data from these clones.

The 22 463 high quality sequences were found to represent 6717 consensus sequences. Thus 48.6% (6717/13 824) of the cDNA library represented unique cDNAs. This was moderately comparable with EST libraries for other crustaceans including the intertidal harpacticoid copepod *Tigriopus japonicus* (262/686 = 38.2% non-overlapping ESTs) and the green shore crab *Carcinus maenas* (1928/5362 = 36.0% unique cDNAs from a normalized cDNA library), and was middle range when compared with EST libraries for the water-flea *Daphnia pulex* (12 600/71 000 = 17.7% unique genes in a unidirectional cDNA library) and the American lobster *H. americanus* (3773/4604 = 82.0% unique cDNAs from a normalized cDNA library) (Lee and others 2005; Colbourne and others 2005; Towle 2005).

The 6717 consensus sequences fell into 5078 different clusters (Table 2). There were 4024 clusters that were only represented by a single clone (singlets), and there were 2 clusters that represented the most redundant transcripts, 1 with 473 clones and the other with 2137 clones (Table 2). The largest cluster contained 194 different consensus sequences, representing 102 different cDNAs.

Sequence similarity of consensus sequences or cluster sequences using a translated query (BLASTx) resulted in the greatest percentage of matches from the nr database (Table 3) and the lowest percentage of matches from the Swiss-Prot database (Table 3). BLASTx hits were considered a strong match if the expect score was $< 1e-4$, a weak match if $1e-1 \geq \text{expect} \geq 1e-4$ and no match if $\text{expect} > 1e-1$.

Table 2 Summary of phrap assembly and clustering

	Number of cDNAs	Number of clusters	Number of clones
	1	4024	4024
	2–5	840	2138
	6–10	114	883
	11–20	61	851
	21–30	18	439
	31–68	13	538
	70–185	6	593
	473	1	473
	2137	1	2137
Total	12 062	5078	12 062

Table 3 Results of BLASTx analysis of 12 060 ESTs

	Database		
	Nr ¹	Swiss-Prot ²	GO ³
By cluster (5078 total)			
Strong match ⁴	1597 (31.5%)	1187 (23.4%)	1289 (25.4%)
Weak match ⁵	1055 (20.8%)	1195 (23.5%)	1178 (23.2%)
No match ⁶	2426 (47.7%)	2696 (53.1%)	2611 (51.4%)
By consensus sequence (6717 total)			
Strong match ⁴	2023 (30.1%)	1523 (22.7%)	1638 (24.4%)
Weak match ⁵	1321 (19.7%)	1482 (22.0%)	1443 (21.5%)
No match ⁶	3373 (50.2%)	3712 (55.3%)	3636 (54.1%)

¹nr: All non-redundant GenBank CDS translations + PDB + Swiss-Prot + PIR + PRF.

²Swiss-Prot: the last major release of the SWISS-PROT protein sequence database.

³GO: the Gene Ontology database.

⁴Strong match: $\text{expect} < 1e-4$.

⁵Weak match: $1e-1 \geq \text{expect} \geq 1e-4$.

⁶No match: $\text{expect} > 1e-1$.

In general, ~ 23 – 30 % of the consensus sequences or clusters had strong matches, 20 – 22 % had weak matches, and 50 – 55 % did not match any known sequence (Table 3).

Of the strong matches, different transcripts were predominant in the different tissue libraries (Table 4). Anti-lipopolysaccharide factor (anti-LPS) was most abundant in heart and whole crab libraries. Anti-LPS is an anticoagulating agent and would be expected to be highly expressed in the heart; its predominance in the whole crab library might be explained by expression in hemocytes. Genes for mitochondrial proteins were most abundant in the nerve library, which likely reflects high nerve ATP generation, and the fact that most of the nerve tissue used was axon and not cell bodies. Proteases were most common in the hepatopancreas library, which was anticipated

Table 4 Most represented cDNAs¹ in each tissue library

Heart	
Anti-lipoplysaccharide factor gi113657	<i>Tachypleus tridentatus</i> 4e–17 (159) ²
ATP lipid-binding protein gi 18700491	<i>Marsupenaeus japonicus</i> 2e–30 (42)
ATP synthase epsilon chain gi 21297384	<i>Anopheles gambiae</i> 6e–12 (12)
Caspase recruitment domain family 6 gi 28481366	<i>Mus musculus</i> 3e–06, (19)
Carcinin-like protein gi 51890388	<i>Carcinus maenas</i> 1e–08 (13)
Cytochrome c oxidase subunit 1 gi 63003725	<i>M. japonicus</i> 1e–160 (71)
Cytochrome c oxidase subunit 2 gi 7542363	<i>Pagurus longicarpus</i> 5e–94 (20)
Cytochrome c oxidase subunit 3 gi 15150786	<i>P. longicarpus</i> 2e–90 (12)
Cytochrome c oxidase subunit 7c gi 51011612	<i>Ixodes pacificus</i> 3e–15 (21)
Glutathione S-transferase gi 2842718	<i>A. gambiae</i> 3e–61 (15)
Putative antimicrobial peptide gi 17223036	<i>Litopenaeus vannamei</i> 2e–12 (19)
Zinc finger protein gi 55649977	<i>Pan troglodytes</i> 3e–10 (24)
Nerve	
Cytochrome b gi 62161334	<i>Pseudocarcinus gigas</i> 1e–148 (32)
Cytochrome oxidase subunit 1 gi 63003725	<i>M. japonicus</i> 1e–160 (74)
Cytochrome oxidase subunit 2 gi 7542363	<i>P. longicarpus</i> 5e–94 (38)
Whole crab	
Anti-LPS factor gi 113657	<i>T. tridentatus</i> 4e–17 (73)
Slow tropomyosin isoform gi 2660866	<i>Homarus americanus</i> 1e–124 (27)
Putative antimicrobial peptide gi 17223036	<i>L. vannamei</i> 2e–12 (27).
Hepatopancreas	
Trypsin gi 785035	<i>L. vannamei</i> 1e–132 (703)
Chitinase gi 55239200	<i>A. gambiae</i> 3e–05 (63)
Collagenolytic serine protease gi 18266071	<i>Paralithodes camtschaticus</i> 1e–127 (72)
Zinc proteinase Mpc1 gi 19774211	<i>P. camtschaticus</i> 1e–111 (42)
Gills	
Arthrodiol cuticle protein AMP8.1 gi 54042606	<i>Callinectes sapidus</i> 2e–17 (23)
60S ribosomal protein L40 gi 56417572	<i>Aedes albopictus</i> 6e–64 (21)
ATP lipid-binding protein like protein gi 18700491	<i>M. japonicus</i> 2e–30 (19)
Alpha actin gi 15148888	<i>H. americanus</i> 8e–41 (19)
Metallothionein-1 gi 102749	<i>H. americanus</i> (6e–21) (16)

Table 4 Continued

Translationally controlled tumor protein gi 56199605	<i>Fenneropenaeus merguensis</i> (7e–67) (14)
Ubiquitin/ribosomal protein S30e fusion protein gi 69608642	<i>Hister</i> sp. 4e–23 (12)
Claw muscle	
Myosin 1 light chain gi 48141152	<i>Apis mellifera</i> 1e–49 (105)
Putative muscle actin gi 53830696	<i>Oncometopia nigricans</i> (0.0) (50)
Alpha actin gi 15148888	<i>H. americanus</i> 8e–41 (23)
Troponin I fast skeletal muscle gi 102756	<i>Astacus astacus</i> 4e–34, (20)
Slow tropomyosin isoform gi 2660866	<i>H. americanus</i> 1e–124 (19)
Sarcoplasmic calcium-binding protein I gi 134309	<i>Pontastacus leptodactylus</i> 8e–92 (17)
LIM protein gi 50982101	<i>Apriona germari</i> 4e–21 (15)
Slow muscle myosin S1 heavy chain gi 37925239	<i>H. americanus</i> 1e–133 (14)

¹From strong matches (Table 3).

²Organization: **name of gene** GenBank accession no. # species e–value (**# clones**) (www.ncbi.nlm.nih.gov/BLAST/).

because of this tissue's digestive role. Cuticle protein was an abundant transcript in the gill library, and reflects the large cuticular surface of this tissue. Lastly, as anticipated, muscle proteins were most abundant in the claw muscle library (Table 4).

A large number of transcripts occurred in only 1 or 2 of the tissue libraries (Appendix Table A1), although we should not necessarily attempt to confer function from this fact as our library is far from comprehensive for each tissue. Many transcripts that were observed in specific tissues have interesting biological function and warrant further discussion. For example, in heart libraries an EST encoding the iron-binding protein transferrin was cloned (Appendix Table A1). Further analysis, however, is required to determine whether this EST represents *transferrin* or 1 of the mRNAs that encode the large proteinase inhibitor, pacifastin (Liang and others 1997).

Examination of the 2 largest clusters (Table 2) reveals that the clustering algorithm has grouped many different genes together. The 102 different cDNAs from the largest cluster (above, Table 2) included 4 separate cDNAs each for 2 proteins (anti-LPS and trypsin) and 2 separate cDNAs each for 3 proteins (tropomyosin, cytochrome b, cytochrome oxidase subunit I); the remaining 88 cDNAs in the largest cluster represented different proteins. The second largest cluster (above, Table 2) contained 44 different consensus sequences that encoded a total of 10 different proteins, including 4 cDNAs for

Table 5 RNA samples from *Petrolisthes cinctipes* used in phase II library construction

Tissues	Treatment (sampling points)
Heart	Heat 30°C, 4 h (2, 4, and 6 h 15°C recovery)
Gill	Cold 2°C, 4 h (2, 4, and 6 h 15°C recovery)
Crab "remains" (after heart, gill, and hepatopancreas removed)	H ₂ O ₂ , 0.5 mM (18 h) CdCl ₂ , 50 μM (24 h) Selenate, 50 μM (24 h) Selenite, 50 μM (24 h) Hypersalinity, 54‰ (18 h) Hyposalinity, 13‰ (18 h) Desiccation (24 h) Hypoxia, 2 h (20 min normoxia recovery) Starvation, 15 days (2 h postprandial) Insecticide, 1 spray Pyrethrin/200 ml, 5 min (4 h recovery)
Larvae, freshly molted whole crabs	Acclimated for 1–7 days in San Francisco Bay water (salinity 25–32‰)
Whole crabs	Acclimated for 1 month to 8, 15, 18, and 25°C
Heart, hepatopancreas	Field acclimatized, north–south, winter–summer
Heart, gill, claw	Acclimated for 1 month to 7, 19°C.
Heart, nerve, gill, claw	Acclimated to 1 month in a thermally fluctuating condition (8:18°C, 12 h:12 h)

arginine kinase, 3 cDNAs for alpha actin, and 2 cDNAs for beta actin. In both of these large clusters, there was also a number of consensus sequences that did not match any known proteins. This result is probably reflective of the challenges of developing clustering algorithms that work every time; these challenges are being actively pursued at JGI, and refinements to clustering algorithms are forthcoming.

GO terms for each cluster were categorized by cellular compartment (Fig. 2A) and molecular function (Fig. 2B). Information for biological process was not available for most of the clusters, and thus these data are not shown here. Some clusters had GO terms for more than 1 cellular compartment or more than 1 molecular function.

Construction of the second EST library for *P. cinctipes* will be performed at the JGI using RNA derived from crabs exposed to a wide array of stressors, thermal conditions, developmental states, and physiological states (Table 5). For this library, we have omitted

most of the hepatopancreas tissues because of the very high redundancy of trypsin in this tissue (Table 4). Stressors include temperature, chemicals, heavy metals, high and low salinity (osmotic stressors), and hypoxia. Thermal conditions include warm and cold acclimation, as well as acclimation to warm–cold fluctuating temperatures, and crabs acclimatized to the species' range of natural habitat conditions. By sampling larval and freshly molted crabs, we captured transcripts unique to processes occurring in those stages. Total RNA has been pooled from all of the tissue samples indicated in Table 5, and will be used by JGI to develop and sequence a library of 50 000–75 000 clones. At the completion of that project, we hope to have captured as much of the *P. cinctipes* transcriptome as possible. The final analyzed set of ESTs will be submitted to GenBank dbEST, and a "unique" gene set will be generated for use in printing cDNA microarrays to be used for functional genomics analyses.

Acknowledgments

This work was supported by NSF-IOB 0533920 to J.H.S., NSF graduate research fellowship to K.S.T. DNA sequencing was performed as a part of a 2006 Community Sequencing Project, under the auspices of the US Department of Energy's Office of Science, Biological, and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under Contract No. W-7405-ENG-36.

Conflict of interest: None declared.

References

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, and others. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–95.
- Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, and others. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503–11.
- Aparicio S, Chapman J, Stupka E, Putnam N, Chia J, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, and others. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301–10.
- C. elegans Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282:2012–18.

- Colbourne JK, Lindquist E, Bauer D, Brokstein P, Thomas WK, Andrews J. 2005. Gene discovery in *Daphnia* by expressed sequence tag sequencing. Society for Integrative and Comparative Biology 2006. Integr Comp Biol 45(6):979. Annual Meeting Abstracts.
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, and others. 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298:2157–67.
- Ewing B, Green P. 1998. Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–94.
- Ewing B, Hillier L, Wendl M, Green P. 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–85.
- Feder ME, Mitchell-Olds T. 2003. Evolutionary and ecological functional genomics. Nat Rev Genet 4:649–55.
- Gracey AY, Cossins AR. 2003. Application of microarray technology in environmental and comparative physiology. Annu Rev Physiol 65:231–59.
- Gueguen Y, Cadoret J, Flament D, Barreau-Roumiguier C, Girardot A, Garnier J, Hoareau A, Bachere E, Escoubas J. 2003. Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, *Crassostrea gigas*. Gene 303:139–45.
- Haig J. 1960. The Porcellanidae (Crustacea: Anomura) of the eastern Pacific. Allan Hancock Pacific Expeditions 24:1–440.
- Hofmann GE, Burnaford JL, Fielman KT. 2005. Genomics-fueled approaches to current challenges in marine ecology. Trends Ecol Evol 20(6):303–11.
- Lee Y, Kim I, Jung S, Lee J. 2005. Analysis of 686 expressed sequence tags (ESTs) from the intertidal harpacticoid copepod *Tigriopus japonicus* (Crustacea, Copepoda). Marine Pollution Bulletin 51:757–68.
- Liang Z, Sottrup-Jensen L, Aspan A, Hall M, Soderhall K. 1997. Pacifastin, a novel 155-kDa heterodimeric proteinase inhibitor containing a unique transferring chain. Proc Natl Acad Sci USA 94:6682–7.
- McKillen DJ, Chen YA, Chen C, Jenny MJ, Trent HF, Robalino J, McLean DC, Gross PS, Chapman RW, Warr GW, and others. 2005. Marine genomics: a clearing-house for genomic and transcriptomic data of marine organisms. BMC Genomics 6:34.
- Pi C, Liu Y, Peng C, Jiang X, Liu J, Xu B, Yu X, Yu Y, Jiang X, Wang L, and others. 2006. Analysis of expressed sequence tags from the venom ducts of *Conus striatus*: focusing on the expression profile of conotoxins. Biochimie 88:131–40.
- Roberts SB, Goetz FW. 2003. Expressed sequence tag analysis of genes expressed in the Bay Scallop, *Argopecten irradians*. Biol Bull 205:227–8.
- Smith TF, Waterman MS. 1981. Identification of common molecular substances. J Mol Biol 147:195–7.
- Stillman J. 2004. A comparative analysis of plasticity of thermal limits in porcelain crabs across latitudinal and intertidal zone clines. Int Congr Ser 1275:267–74.
- Stillman JH. 2002. Causes and consequences of thermal tolerance limits in rocky intertidal porcelain crabs, genus *Petroliastes*. Integr Comp Biol 42(4):790–6.
- Stillman JH. 2003. Acclimation capacity underlies susceptibility to climate change. Science 301:65.
- Stillman JH, Reeb CA. 2001. Molecular phylogeny of Eastern Pacific porcelain crabs, genera *Petroliastes* and *Pachycheles*, based on the mtDNA 16S rDNA sequence: phylogeographic and systematic implications. Mol Phylogenet Evol 19(2):236–45.
- Supungul P, Klinbunga S, Pichyangkura R, Jitrapakdee S, Hirono I, Aoki T, Tassanakajon A. 2002. Identification of immune-related genes in hemocytes of Black Tiger Shrimp (*Penaeus monodon*). Mar Biotechnol 4:487–94.
- Towle DW. 2005. Functional genomics of environmental adaptations in marine crustaceans. Society for Integrative and Comparative Biology 2006 Annual Meeting Abstracts. Integr Comp Biol 45(6):1085.
- Venier P, Pallavicini A, Nardi BD, Lanfranchi G. 2003. Towards a catalogue of genes transcribed in multiple tissues of *Mytilus galloprovincialis*. Gene 314:29–40.

Appendix

Table A1 Tissue specific transcripts (nr database)

Gene name	gi#	Species	e-value
Heart			
Carboxypeptidase, vitellogenic-like	62857515	<i>Xenopus tropicalis</i>	2e-67
Cytochrome P450 CYP4	18032259	<i>Cherax quadricarinatus</i>	8e-14
NADH dehydrogenase subunit 4	46575798	<i>Euphausia superba</i>	3e-71
Phospholipase/carboxylesterase	50737613	<i>Apis mellifera</i>	2e-21
Zinc finger, CCHC domain	27754058	<i>Mus musculus</i>	5e-13
NADH dehydrogenase subunit 1	15150779	<i>Marsupenaeus japonicus</i>	5e-98
Collagen alpha 2(IV) chain precursor	115347	<i>Ascaris suum</i>	1e-112
Transferrin	22597202	<i>Mastotermes darwiniensis</i>	9e-46
Galactokinase	1730187	<i>Homo sapiens</i>	3e-93
Vacuolar protein sorting protein	73920458	<i>Danio rerio</i>	6e-48
Protein disulfide isomerase	12025459	<i>Bombyx mori</i>	2e-15
Eukaryotic translation initiation factor 3 subunit 2	50759828	<i>Gallus gallus</i>	1e-48
Ring finger protein 26	27661650	<i>Rattus norvegicus</i>	4e-13
Nuclear transport factor 2-like export factor 2	57530164	<i>Homo sapiens</i>	2e-25
Enolase	1311141	<i>Anopheles gambiae</i>	6e-11
Ficolin 4	14349161	<i>Halocynthia roretzi</i>	7e-25
Succinyl-CoA:3-ketoacid-coenzyme A transferase 1	2492998	<i>Homo sapiens</i>	6e-42
Outer membrane receptor proteins (Fe transport)	42628903	<i>Haemophilus influenzae</i>	7e-6
Serine kinase SRPK2-alternatively spliced form	3406051	<i>Homo sapiens</i>	1e-80
X-box binding protein	13898897	<i>Xenopus laevis</i>	2e-16
Small zinc finger-like	5107178	<i>Ciona intestinalis</i>	1e-24
Cytohesin 1-like protein	46310219	<i>Danio rerio</i>	3e-82
26S protease regulatory subunit 7	51262008	<i>Xenopus tropicalis</i>	8e-43
Ubiquitin activating enzyme	2706522	<i>Drosophila melanogaster</i>	1e-41
Transglutaminase	14579327	<i>Pacifastacus leniusculus</i>	4e-26
Kinesin light chain	1170680	<i>Loligo pealei</i>	1e-66
Elongation factor 1 delta	12328436	<i>Bombyx mori</i>	6e-42
Pontin	12004636	<i>Xenopus laevis</i>	1e-106
SH3 domain-binding glutamic acid-rich protein	66513595	<i>Apis mellifera</i>	6e-26
Sphingomyelin phosphodiesterase, acid-like 3B	50759633	<i>Gallus gallus</i>	2e-45
Glyceraldehyde-3-phosphate dehydrogenase	31338868	<i>Procambarus clarkii</i>	1e-175
Extracellular superoxide dismutase precursor	4585366	<i>Pacifastacus leniusculus</i>	7e-57
Serine (or cysteine) proteinase inhibitor	61873714	<i>Bos taurus</i>	3e-13
Iron-sulfur cluster binding protein	67932841	<i>Solibacter usitatus</i>	3e-74
Ribulose-5-phosphate-3-epimerase	37747988	<i>Danio rerio</i>	7e-25
Endothelial cell growth factor 1	19923857	<i>Mus musculus</i>	6e-45
Proteasome subunit alpha type 7	41351173	<i>Danio rerio</i>	2e-94
26S protease regulatory subunit 8	1709799	<i>Manduca sexta</i>	1e-115
ATP-dependent 26S proteasome regulatory subunit	29825445	<i>Apis mellifera</i>	2e-69
Thioredoxin-dependent peroxide reductase	2507170	<i>Bos taurus</i>	3e-76
Heat shock cognate 70	23193450	<i>Chironomus tentans</i>	3e-25
Aminotransferase	33286231	<i>D. melanogaster</i>	7e-67
GTPase	23171521	<i>D. melanogaster</i>	5e-11

Table A1 Continued

Gene name	gi#	Species	e-value
Isocitrate dehydrogenase	48476117	<i>Crassostrea gigas</i>	2e-83
Ubiquinone biosynthesis protein COQ7-like protein	47157064	<i>Apis mellifera</i>	3e-69
26S protease regulatory subunit 4	30581054	<i>D. melanogaster</i>	1e-124
UDP-glucose ceramide glucosyltransferase-like 1	50751921	<i>Gallus gallus</i>	6e-75
Cold shock domain protein A long isoform	47059495	<i>Mus musculus</i>	1e-15
Protein phosphatase 1M regulatory subunit	45382609	<i>Gallus gallus</i>	6e-06
S-adenosylmethionine synthetase	452838	<i>D. melanogaster</i>	1e-128
Putative γ -aminobutyric acid receptor beta subunit	386138	<i>D. melanogaster</i>	1e-114
Splicing factor, arginine/serine-rich 7	22122585	<i>Mus musculus</i>	2e-41
Glyceraldehyde-3-phosphate dehydrogenase	6016080	<i>Procambarus clarkii</i>	1e-175
Proteasome (prosome, macropain) subunit, α type, 4	28279709	<i>Danio rerio</i>	1e-101
Leukotriene A-4 hydrolase	68358508	<i>Danio rerio</i>	1e-18
NADH dehydrogenase subunit 1	15150779	<i>M. japonicus</i>	5e-98
Proteasome 26S subunit, non-ATPase	37681913	<i>Danio rerio</i>	1e-42
Nerve			
Similar to apoptosis-linked gene 2	48094929	<i>Apis mellifera</i>	4e-69
Putative phosphoglycerate mutase	52630953	<i>Toxoptera citricida</i>	1e-113
U3 small nucleolar ribonucleoprotein	13528759	<i>Homo sapiens</i>	7e-42
20-beta-hydroxysteroid dehydrogenase	66517758	<i>Apis mellifera</i>	4e-33
Methylmalonate semialdehyde dehydrogenase	50748470	<i>Gallus gallus</i>	2e-76
3-hydroxyacyl-CoA dehydrogenase	2078327	<i>Homo sapiens</i>	1e-101
Cathepsin L-like cysteine protease precursor	16304178	<i>Delia radicum</i>	2e-13
Translation initiation factor eIF-2B α subunit	72014480	<i>Strongylocentrotus purpuratus</i>	9e-68
Mitochondrial carrier protein	72010168	<i>Homo sapiens</i>	6e-73
Gill			
Thioredoxin reductase 1	51704106	<i>Xenopus laevis</i>	1e-64
Aldehyde dehydrogenase (mitochondrial)	55638955	<i>Pan troglodytes</i>	2e-38
Proteasome activator subunit 3 isoform 1	30410794	<i>Homo sapiens</i>	6e-69
Enoyl Coenzyme A hydratase, short- chain 1	40555865	<i>Rattus norvegicus</i>	5e-67
Glyceraldehyde-3-phosphate dehydrogenase	31338868	<i>Procambarus clarkii</i>	1e-175
Chaperonin containing TCP1, subunit 5 (epsilon)	39850245	<i>Xenopus tropicalis</i>	1e-126
Defender against cell death 1 (DAD-1)	20138077	<i>D. melanogaster</i>	3e-43
Glutamate dehydrogenase, short peptide	458803	<i>D. melanogaster</i>	2e-58
Phosphoserine aminotransferase 1	39795813	<i>Danio rerio</i>	1e-44
26S proteasome regulatory chain 4	345717	<i>D. melanogaster</i>	1e-124
NADH dehydrogenase (ubiquinone) Fe-S protein 4	4505369	<i>Homo sapiens</i>	3e-44
Acyl-CoA-binding protein homolog	1168274	<i>Rana ridibunda</i>	7e-30
ADP-sugar pyrophosphatase (Nudix motif 5)	72005173	<i>Strongylocentrotus purpuratus</i>	3e-24
Methionine-R-sulfoxide reductase (Selenoprotein R)	34922549	<i>D. melanogaster</i>	8e-44
Sodium-dependent dicarboxylate transporter 2	68369674	<i>Danio rerio</i>	3e-41
Zinc finger, matrix type 5	57524615	<i>Danio rerio</i>	1e-12
Putative salivary sulfotransferase	67083857	<i>Ixodes scapularis</i>	0.27
NADH-ubiquinone oxidoreductase acyl-carrier subunit	1653987	<i>D. melanogaster</i>	4e-23
Glutamate-ammonia ligase	543538	<i>Panulirus argus</i>	1e-109
DNA methyltransferase 1-associated protein 1	72022174	<i>Strongylocentrotus purpuratus</i>	6e-15
X-box binding protein	13898897	<i>Xenopus laevis</i>	2e-16

Table A1 Continued

Gene name	gi#	Species	e-value
Vacuolar ATP synthase subunit E, putative	15222641	<i>Arabidopsis thaliana</i>	2e-31
Transient receptor potential (channel nanchung)	31745595	<i>D. melanogaster</i>	1e-10
Na ⁺ /K ⁺ -exchanging ATPase beta chain	84610	<i>Artemia franciscana</i>	7e-22
Metabotropic glutamate receptor	1834427	<i>D. melanogaster</i>	3e-36
Putative epidermal growth factor receptor	13445276	<i>Anopheles gambiae</i>	7e-64
Ecdysteroid receptor	13677226	<i>Celuca pugilator</i>	1e-106
Hepatopancreas			
Trypsin	785035	<i>Litopenaeus vannamei</i>	1e-132
Trypsin	27373057	<i>Aplysina fistularis</i>	1e-21
Collagenolytic serine protease	18266071	<i>Paralithodes camtschaticus</i>	3e-84
Duplex-specific nuclease	26892281	<i>Paralithodes camtschaticus</i>	1e-102
Transcriptional coactivator tubedown-100	22035307	<i>Homo sapiens</i>	1e-13
Carboxylesterase	66560187	<i>Apis mellifera</i>	1e-10
beta 1,4-endoglucanase	5020110	<i>Cherax quadricarinatus</i>	6e-08
Cathepsin I	1483570	<i>L. vannamei</i>	5e-79
Dimeric dihydrodiol dehydrogenase	5766901	<i>Canis familiaris</i>	1e-57
Dipeptidyl-peptidase I precursor	50731191	<i>Gallus gallus</i>	6e-46
Carboxypeptidase B	115881	<i>Astacus astacus</i>	2e-70
Guanine nucleotide exchange factor for Rho/Rac/Cdc42-like GTPases	72005061	<i>Anopheles gambiae</i>	2e-48
Ruvbl2-prov protein	29126859	<i>Xenopus laevis</i>	1e-125
DNA topoisomerase II, α isozyme isoform 3	73966011	<i>Canis familiaris</i>	9e-80
Cystatin B	23344732	<i>Theromyzon tessulatum</i>	1e-12
Low-density lipoprotein receptor (LDL receptor)	126074	<i>Oryctolagus cuniculus</i>	2e-42
Calcitonin gene-related peptide-receptor component	55628662	<i>Pan troglodytes</i>	2e-28
Succinate dehydrogenase (ubiquinone)	67043769	<i>Lysiphlebus testaceipes</i>	1e-8
Destabilase I	1255718	<i>Hirudo medicinalis</i>	1e-18
RNA-binding protein	23613209	<i>Plasmodium falciparum</i>	3e-9
Cysteine dioxygenase, type I	37748497	<i>Dario rerio</i>	6e-22
Ornithine aminotransferase	790956	<i>Drosophila ananassae</i>	2e-63
Serine proteinase inhibitor	33590491	<i>Procambarus clarkii</i>	9e-19
Claw			
70 kDa heat shock protein	55233307	<i>Anopheles gambiae</i>	3e-21
Histone deacetylase 11	50754303	<i>Gallus gallus</i>	3e-26
Ficolin	47523126	<i>Sus scrofa</i>	9e-20
Ornithine decarboxylase	1200128	<i>Panagrellus redivivus</i>	3e-15
Mitochondrial short-chain enoyl-coenzyme A hydratase 1 precursor	12707570	<i>Homo sapiens</i>	1e-23
Chaperonin subunit 8 theta	44969706	<i>Gallus gallus</i>	2e-74
Eukaryotic translation elongation factor 1 beta 2	47940399	<i>Dario rerio</i>	9e-50
Cystathionine gamma-lyase	1705787	<i>Caenorhabditis elegans</i>	2e-22
Basic transcription factor 3-like 4	41152344	<i>Danio rerio</i>	1e-33
Transcription elongation factor B (SIII), polypeptide 1	63100863	<i>Xenopus tropicalis</i>	4e-27
Whole Crab			
Putative helicase	37535038	<i>Oryza sativa</i>	3e-13
Tenascin-R	1617316	<i>Homo sapiens</i>	5e-45
Site-specific recombinase	30248840	<i>Nitrosomonas europaea</i>	2e-35

Table A1 Continued

Gene name	gi#	Species	e-value
Tissue factor pathway inhibitor precursor (TFPI) (Lipoprotein-associated coagulation inhibitor) (LACI)	401174	<i>Rattus norvegicus</i>	7e-38
Fasciclin IV precursor	160845	<i>Schistocerca americana</i>	2e-72
Prophenoloxidase activating factor	66513601	<i>Apis mellifera</i>	2e-34
U2 small nuclear ribonucleoprotein 35 kDa subunit	2833265	<i>Homo sapiens</i>	0.038
Activating signal cointegrator 1 complex subunit 1	50749326	<i>Gallus gallus</i>	2e-61
Zinc-binding dehydrogenase	56118580	<i>Xenopus tropicalis</i>	2e-28
Zinc finger CCCH-type, antiviral 1	61098418	<i>Gallus gallus</i>	1e-14
Multisubstrate deoxyribonucleoside kinase	28629060	<i>Anopheles gambiae</i>	3e-49
Guanosine monophosphate reductase 2	19527300	<i>Mus musculus</i>	5e-27
Glycine cleavage system H protein	67465060	<i>Mus musculus</i>	5e-40
Na ⁺ /K ⁺ ATPase alpha subunit	23380400	<i>H. americanus</i>	1e-149