EXPRESSED SEQUENCE TAG ANALYSIS OF GENES EXPRESSED DURING DEVELOPMENT OF THE TROPICAL ABALONE *HALIOTIS ASININA*

Author(s): DANIEL J. JACKSON and BERNARD M. DEGNAN

Source: Journal of Shellfish Research, 25(1):225-231.
Published By: National Shellfisheries Association
DOI: http://dx.doi.org/10.2983/0730-8000(2006)25[225:ESTAOG]2.0.CO;2
URL: http://www.bioone.org/doi/full/10.2983/0730-8000%282006%2925%5B225%3AESTAOG%5D2.0.CO%3B2

Journal of Shellfish Research, Vol. 25, No. 1, 225–231, 2006.

EXPRESSED SEQUENCE TAG ANALYSIS OF GENES EXPRESSED DURING DEVELOPMENT OF THE TROPICAL ABALONE *HALIOTIS ASININA*

DANIEL J. JACKSON AND BERNARD M. DEGNAN*
School of Integrative Biology, University of Queensland, Brisbane QLD 4072 Australia

ABSTRACT The tropical abalone, Haliotis asinina, is an ideal species to investigate the molecular mechanisms that control development, growth, reproduction and shell formation in all cultured haliotids. Here we describe the analysis of 232 expressed sequence tags (EST) obtained from a developmental H. asinina cDNA library intended for future microarray studies. From this data set we identified 183 unique gene clusters. Of these, 90 clusters showed significant homology with sequences lodged in GenBank, ranging in function from general housekeeping to signal transduction, gene regulation and cell-cell communication. Seventy-one clusters possessed completely novel ORFs greater than 50 codons in length, highlighting the paucity of sequence data from molluscs and other lophotrochozoans. This study of developmental gene expression in H. asinina provides the foundation for further detailed analyses of abalone growth, development and reproduction.

KEY WORDS: abalone, EST, Lophotrochozoan, development, *Haliotis asinine*

INTRODUCTION

Abalone (*Haliotis* spp.) are a particularly attractive aquaculture species because of their high demand, high market value and limited supply (Gordon & Cook 2001, Oakes & Ponte 1996). Whereas temperate abalone form the majority of abalone exports worldwide, there is a growing demand for tropical species (Jarayabhand & Paphavasit 1996). The tropical Indo-Pacific, including Australia, is well positioned to capture and further develop this market through the aquaculture of the rapidly growing tropical species *Haliotis asinina* Linnaeus. This species reaches marketable size far quicker than the temperate species (Capinpin & Corre 1996, Capinpin et al. 1999) and is ideal for the new "cocktail"-sized abalone market (Nateewathana & Hylleberg 1986, Singhagraiwan & Doi 1993).

We have been using the tropical abalone to address questions pertaining to development (e.g., Giusti et al. 2000, Hinman & Degnan 2002, Hinman et al. 2003, O'Brien & Degnan 2002a, 2002b, O'Brien & Degnan 2003, Jackson et al. 2005) and as a model to investigate the molecular basis of commercially important developmental and physiological traits (Jackson et al. 2005, Selvamani et al. 2001, Selvamani et al. 2000). With a frequent and predictable spawning cycle (Counihan et al. 2001, Jebreen et al. 2000) and established techniques for culturing to sexual reproduction (Jackson et al. 2001), gene expression can be easily studied at any stage of the life cycle. As a member of the Lophotrochozoa, the least investigated bilaterian superphylum (Adoutte et al. 2000), analyses of *H. asinina* development also contributes to a general understanding of metazoan evolution and development.

As a prelude to the sequencing of the human genome, over 170,000 ESTs were sequenced in an attempt to identify new genes and their expression patterns (Adams et al. 1995). This approach has been applied to a wide range of species, permitting the identification of homologous genes from organisms not previously studied (see http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html for a comprehensive list). Included in this public database are a number of marine invertebrates: the ascidians, *Ciona intestinalis* (684,319 ESTs), *Halocynthia roretzi* (4192 ESTs) and *Polyandrocarpa misakiensis* (465 ESTs), the lancelet *Branchiostoma floridae* (277,538), the sea urchin *Strongylocentrotus purpuratus* (130,988 ESTs), the oyster *Crassostrea virginica* (9018 ESTs) and the prawn *Penaeus monodon* (1611 ESTs) (dbEST release July 22, 2005). However there is a significant bias in this database towards vertebrates, model organisms and terrestrial species of commercial value. Here we describe a small-scale EST analysis of cDNAs obtained from a cDNA library constructed from a range of developmental stages of the tropical abalone *H. asinina*.

MATERIALS AND METHODS

RNA Extraction and Library Construction

Total RNA was extracted according to the method of Chomczynski and Sacchi (1987) from the following developmental stages: egg, gastrula, early trochophore (13 h post fertilization; hpf), mid torsion (18 hpf), early veliger (24 hpf), mid veliger (40 hpf), early competent veliger (72 hpf), late competent veliger (134 hpf), 1 day post metamorphosis and 5 days post metamorphosis. All RNAs were quantified spectrophotometrically and quality-assessed by inspection of samples that had been electrophoresed through a 1% formaldehyde gel (Sambrook & Russell 2001). To maximize gene representation in the cDNA library 200 ng of total RNA from each stage was combined for library construction using the directional Clontech SMART cDNA library synthesis kit. Following the manufacturers instructions, double stranded cDNA was PCR amplified for 23 cycles prior to cloning. Five-thousand plaques were individually hand picked following methods outlined by Sambrook and Russell (2001) for microarray printing. Of these 288 plaques were PCR amplified using the λTriple X forward (CTCGGGAAGCGCGCCATTGTGTTGGT) and reverse (TAATACGACTCACTATAGGGCGAATTGGCC) primers. Successful PCR reactions were subsequently purified with the Millipore multiscreen PCR clean up kit. Two-hundred and thirty-two clones were single-pass sequenced using the λTriple X forward primer and ABI Big Dye terminator mix (version 2) following standard procedures (2001).

Sequence Analysis

Sequencing chromatograms were assessed manually using 4Peaks (v1.5) and were truncated to eliminate ambiguous terminal base calls and vector sequence. BLASTx and BLASTn homology searches were conducted against the NCBI nr databases using the

*Corresponding author. E-mail: b.degnan@uq.edu.au

default conditions (low complexity filter on, an expect value of 10 and the BLOSUM62 matrix with existence 11 and extension 1 gap costs). Sequences that returned BLAST matches with an E value of $1e^{-06}$ or less were considered significant, and allowed a functional classification to be loosely made following the scheme of Lee et al. (1999). To assess sequence redundancy or "uniqueness," sequences were clustered using ClustalW (http://www.ebi.ac.uk/clustalw/index.html?), (Chenna et al. 2003). The integrity of each cluster was assessed manually by visually inspecting cluster alignments for overlap quality. Where cluster members were found to share more than 95% identity the cluster was usually maintained.

ESTs were also assessed for the presence of a putative poly (A) tail and an associated polyadenylation signal based on the classification of Beaudoing et al. (2000). Those sequences that did not return any significant BLAST result were assessed for the presence of an open reading frame (ORF) using the European Bioinformatics Institute Transeq server (http://www.ebi.ac.uk/emboss/transeq/). Only ORFs longer than 50 amino acid residues were accepted (Lee et al. 1999, Lehninger et al. 1993).

RESULTS

The *Haliotis asinina* developmental cDNA library contained 5×10^7 pfu/mL, of which approximately 93% were recombinant clones. These had an average insert size of 1044 ± 411 bp. Clustering of the 232 sequences revealed that 157 sequences occurred only once, 16 sequences occurred twice, 3 sequences occurred three times, 3 sequences occurred four times, 3 sequences occurred 5 times and 1 sequence occurred 7 times, producing 183 unique clusters (Table 1).

Homology searches using the BLASTx and BLASTn algorithms were conducted for all sequences. 90 clusters returned a significant BLAST result and are grouped according to function in Table 2. Sixty-eight clusters fell into the category of genes used by many cell types, 8 clusters showed homology with genes involved in cell-cell communication, 3 clusters showed homology with transcription factors, and they are likely to play important developmental or biosynthetic roles in *H. asinina*. Eleven clusters showed significant homology with an undescribed ORF and could not be classified (Fig. 1; Table 1). These EST sequences have been submitted to GenBank under accession numbers DY402832–DY403158.

Polyadenylation Signal and ORF Analysis

Of the 183 unique clusters, 81 were completely sequenced in one pass and possessed a poly (A) -rich region at the 3′ end of the sequence. Of these 81, 54 possessed a putative polyadenylation signal a mean distance of 17 nucleotides upstream of the poly (A) tail. The frequency of polyadenylation signals alternative to the predominant AATAAA signal was remarkably similar to that of human signals reported by Beaudoing et al. (2000) (Table 3). We also found that out of all ESTs with a polyadenylation signal, 5 (2.1%) had dual polyadenylation signals an appropriate distance

TABLE 1.

Summary of ESTs grouped by category and BLAST results.

Code	Class	No. of Clusters	No. of Clones
A. Functions that many kinds of cells use			
AI	Transport and binding proteins for ions and other small molecules	1	1
AII	RNA processing, polymerising, splicing and binding proteins	4	4
AIII	Cell replication, histones, cyclins and allied kinases, DNA polymerases	4	5
AIV	Cytoskeleton and membrane proteins, cellular organisers	7	8
AV	Protein synthesis cofactors, ribosomal proteins, rRNAs	29	47
AVI	Intermediary metabolism and catabolism enzymes	7	7
AVII	Stress response, detoxification, and cell defense proteins	8	18
AVIII	Protein degradation and processing, proteases	3	3
AIX	Transportation and binding proteins for proteins and other macromolecules	3	4
AX	Proteins involved in motility including muscle components	2	1
	Total	68	98
B. Cell-cell communication			
BI	Signalling receptors and ligands	4	4
BII	Intracellular signaling molecules, including kinases and signal intermediates	3	3
BIII	Extracellular matrix proteins and cell adhesion	1	1
	Total	8	8
C. Transcription factors and other regulatory proteins			
CI	Sequence-specific DNA-binding proteins	3	3
CII	Non-DNA binding proteins that have positive or negative roles	0	0
CII	Chromatin proteins other than AIII with regulatory function	0	0
	Total	3	3
D. Not enough information to classify			
DI	Not enough information to classify	11	11
DII	No significant similarities to known proteins	93	112
	Total	104	123
	Overall total	183	232
	Single copy sequences	157	
	Total no. of ESTs represented more than once	75	
	Total no. of redundant sequences	49	
	Total number of unique sequences (total number of clusters)	183	
	Clusters returning significant BLAST match	90	
	Clusters without any GenBank homologue	93	

TABLE 2.

Significant BLAST results grouped according to function.

Clone	e Value	Accession	Gene Name	Organism	Scientific Name
AI. Transport and binding proteins for ions and other small molecules					
2G2	5e-14	CAG04534	GM2 ganglioside activator*	Pufferfish	*T. nigroviridis*
AII. RNA processing, polymerising, splicing and binding proteins and enzymes					
1H1	1e-45	AAH20773	Exosome component 8	Human	*H. sapiens*
2C9	3e-18	XP_391989	Similar to thickveins	Honey bee	*A. mellifera*
15C8	6e-27	CAF93482	G patch domain*	Pufferfish	*T. nigroviridis*
15D4	2e-30	IFS2_D	RNA polymerase II elongation factor*	Human	*H. sapiens*
AIII. Cell replication, histones, DNA polymerases, topoisomerases, DNA modification					
2B6	1e-48	P08991	Histone H2A variant	Urchin	*S. purpuratus*
15B4	6e-44	AAC47489	Enhancer of rudimentary	Mosquito	*A. aegypti*
15G2	2e-21	CAA28177	Histone H1-beta*	Painted urchin	*L. pictus*
2C7	4e-13	XM_315130	Chief histone H3* **(2)**	Mosquito	*A. gambiae*
AIV. Cytoskeleton and membrane proteins, cellular organisers					
1A11	4e-38	NP_990573	Low density lipoprotein	Chicken	*G. gallus*
1B5	2e-49	CAF21863	Gelsolin	Sponge	*S. ficus*
1D7	1e-108	U30467	Alpha-tubulin	Spoon worm	*U. caupo*
1H7	3e-45	CAD91425	Actin related protein	Oyster	*C. gigas*
2G1	2e-15	CAG07401	Low density lipoprotein*	Pufferfish	*T. nigroviridis*
15D9	2e-82	AY026071	Beta-tubulin	Choanoflagellate	*M. brevicollis*
15D10	2e-43	AF510206	Beta-tubulin	Ciliate	*O. longa*
AV. Protein synthesis cofactors, tRNA synthetase, ribisomal proteins, rRNAs					
15H3	e-147	AY163259	16S ribosomal RNA **(7)**	Tropical abalone	*H. asinina*
2A7	4e-42	AAN05608	Ribosomal protein L26 **(2)**	Scallop	*A. irradians*
15E5	3e-13	U51989	16S ribosomal RNA gene **(2)**	Abalone	*H. diversicolr*
2H5	e-105	AY163259	16S ribosomal RNA **(4)**	Tropical abalone	*H. asinina*
2A5	1e-60	AY145418	28S ribosomal RNA **(2)**	Abalone	*H. discus*
15F5	1e-52	AY588938	Mitochondrion, complete genome **(2)**	Black lip abalone	*H. rubra*
15C9	8e-97	AY163259	16S ribosomal RNA **(2)**	Tropical abalone	*H. asinina*
1B10	3e-48	AAP33157	Beta-NAC-like protein	Termite	*R. flavipes*
1E3	4e-25	AAC15656	60S ribosomal protein P2	Chiton	*C. stelleri*
1E10	1e-22	Z18289	16S rRNA gene	Red alga	*P. palmata*
1G4	5e-34	AF120512	18S ribosomal RNA	Slit shell	*S. confusa*
1G10	2e-98	AY588938	Mitochondrion, complete genome	Black lip abalone	*H. rubra*
2A12	5e-30	X80345	28S ribosomal RNA	Aquatic fungus	*H. catenoides*
2B3	2e-11	EAA59486	Eukaryotic initiation factor*	Fungi	*A. nidulans*
2E12	4e-98	AY145418	28S ribosomal RNA	Abalone	*H. discus*
2F7	e-114	AY145418	28S ribosomal RNA	Abalone	*H. discus*
2H12	2e-42	M98364	Ribosomal RNA	Ciliate	*Coleps sp.*
15A3	4e-57	AAA70102	40S ribosomal protein S24	Slime mould	*D. discoideum*
15A4	9e-12	CAG13823	Mitochondrial ribosomal protein L27*	Pufferfish	*T. nigroviridis*
15A7	9e-35	EAA01025	60S ribosomal protein L6*	Mosquito	*A. gambiae*
15B1	8e-25	AY691376	23S ribosomal RNA	Bacteria	*Burkholdera sp.*
15B2	9e-40	AAM94271	Ribosomal protein S2	Scallop	*C. farreri*
15B7	1e-74	AY588938	Mitochondrion, complete genome	Black lip abalone	*H. rubra*
15B12	4e-24	BX950290	Ribosomal protein L38*	Chicken	*G. gallus*
15E4	4e-66	AAN05595	Ribosomal protein S8	Scallop	*A. irradians*
15E8	6e-29	AAP21827	Ribosomal protein S29	Amphioxus	*B. tsingtaunese*
15F10	2e-32	AAH86809	Translation initiation factor*	Zebrafish	*D. rerio*
15G9	6e-40	CAG02712	Elongation Factor Tu*	Pufferfish	*T. nigroviridis*
15G10	5e-70	AB003720	Elongation factor 1	Turbo	*B. cornutus*
AVI. Intermediary metabolism and catabolism enzymes					
2D3	6e-17	AY431429	Carboxypeptidase A	Mosquito	*A. aegypti*
1C1	2e-26	NP_918077	Putative carboxymethylenebutenolidase	Rice	*O. sativa*
1F4	1e-38	XP_316948	Nudix*	Mosquito	*A. gambiae*
1F10	1e-76	CAE70680	Isocitrate dehydrogenase*	Nematode	*C. elegans*
1H6	3e-27	XP_308198	Pancreatic lipase*	Mosquito	*A. gambiae*
15C3	2e-20	CAA93088	Giutathione S-transferase*	Nematode	*C. elegans*
15G3	9e-11	EAL31332	Maleylacetoacetate isomerase*	Fly	*D. pseudoobscura*

continued on next page

TABLE 2.

continued

Clone	e Value	Accession	Gene Name	Organism	Scientific Name
AVII. Stress response, detoxification, and cell defence proteins					
15E2	2e-53	YP_026076	Cytochrome b **(3)**	Black lip abalone	*H. rubra*
2E5	1e-37	AAR11781	Heat shock protein 90 **(2)**	Scallop	*C. farreri*
2E10	1e-15	AAK56498	Putative metallothionein **(2)**	Periwinkle	*L. littorea*
15F3	4e-09	AAM20842	SARP-19 precursor **(5)**	Periwinkle	*L. littorea*
1F9	3e-27	CAF99069	Ubiquinol-cytochrome C*	Fish	*T. nigroviridis*
1G9	1e-14	ZP_00316378	Fucose binding lectin*	Bacteria	*M. degradans*
2H11	5e-11	EAA03739	UQCRH protein*	Mosquito	*A. gambiae*
15B9	1e-25	AAP31550	HSB1-like protein	Fruit fly	*D. melanogaster*
AVIII. Protein degradation and processing, proteases					
2C2	3e-26	AAH87343	Proteasome 26S subunit*	African clawed frog	*X. laevis*
2C4	3e-52	AAH42820	PSMA8 protein	Human	*H. sapiens*
2E11	2e-29	XP_394993	Proteasome subunit beta	Honey bee	*A. mellifera*
AIX. Transportation and binding proteins for proteins and other macromolecules					
1B12	1e-67	AAT44866	Translocon-associated protein **(2)**	Amphioxus	*B. tsingtaunese*
1B9	6e-14	XP_518767	Heme binding protein*	Chimpanzee	*P. troglodytes*
15F6	3e-24	AAH54875	TIMM9 protein	Human	*H. sapiens*
AX. Proteins involved in motility including muscle components					
15F12	8e-50	XP_537691	Dynein light chain-2	Dog	*C. familiaris*
2F9	1e-10	EAL32148	Myosin II*	Fly	*D. miranda*
BI. Signalling receptors including cytokine and hormone receptors, and signalling ligands					
1A12	2e-13	F44840	FMRF amide	Pond snail	*L. stagnalis*
1C5	4e-7	AF076823	Lysin precursor	Red abalone	*H. rufescens*
1E11	1e-14	NP_001008891	Signal sequence receptor	Rat	*R. norvegicus*
2F12	4e-37	AAT00460	Endozepine	Carp	*C. carpio*
BII. Intracellular signal transduction pathway molecules, including kinases and signal intermediates					
15D12	4e-12	EAL51743	Protein kinase	Entamoeba	*E. histolytica*
15E1	3e-14	AAD34418	Calmodulin mutant	Synthetic construct	*x*
15F11	1e-24	AAO62074	Rho-GTPase-activating protein	Mouse	*M. musculus*
BIII. Extracellular matrix proteins and cell adhesion					
2H9	8e-11	Q01528	Aggregation factor	Horseshoe crab	*L. polyphemus*
CI. Sequence-specific DNA-binding proteins					
2A9	7e-10	XP_470396	Putative pirin-like protein	Rice	*O. sativa*
2C10	2e-19	AAN31640	High mobility group protein 1	Snail	*B. glabrata*
2F1	5e-38	AAP97158	EVORF	Human	*H. sapiens*

* Indicates gene names that do not belong to the most significant BLAST match for that EST. The presented gene name is taken from the next most significant BLAST match that allowed classification.

Numbers in bold parenthesis represent the total number of clones in that cluster.

upstream (both within 35 bp) of the poly A tail. This compares with 16.6% of human genes with dual polyadenylation signals reported by Beaudoing et al. (2000).

Of the 93 clusters that did not produce a significant BLAST result, 71 possessed ORFs equal to or longer than 50 codons (150 nucleotides), suggesting that they represent transcribed genes with no reported homologue.

DISCUSSION

A decade ago Adams et al. (1995) established a set of stringent criteria for selection of cDNA libraries of adequate quality to be used in a large scale human EST analysis. These criteria included the following parameters: (1) less than 20% of the clones in library should have no insert or ribosomal and mitochondrial sequences; (2) the average insert lengths should be 1 kb or greater; (3) at least 50% of the cDNAs should encode novel genes and (4) no gene or group of genes should dominate the distribution. These authors advocate the sequencing of 100–200 clones as an excellent means of assessing library quality based on these criteria. Here we have constructed a cDNA library from 10 distinct stages of development from the tropical abalone *Haliotis asinina*. Upon sequencing 232 randomly selected clones, we found that this library satisfies all of the criteria outlined by Adams et al.(1995). Searching the NCBI databases using the BLAST algorithms, 90 of the 183 clusters (49.1%) are significantly similar to previously described sequences.

As Lee et al. (1999) points out, one of the great advantages of an EST analysis of a novel cDNA library constructed from relatively poorly studied tissues is the generation of probes for "interesting" genes. This study has revealed a number of ESTs that deserve further investigation in terms of their temporal and spatial expression profiles, and the developmental and physiological roles

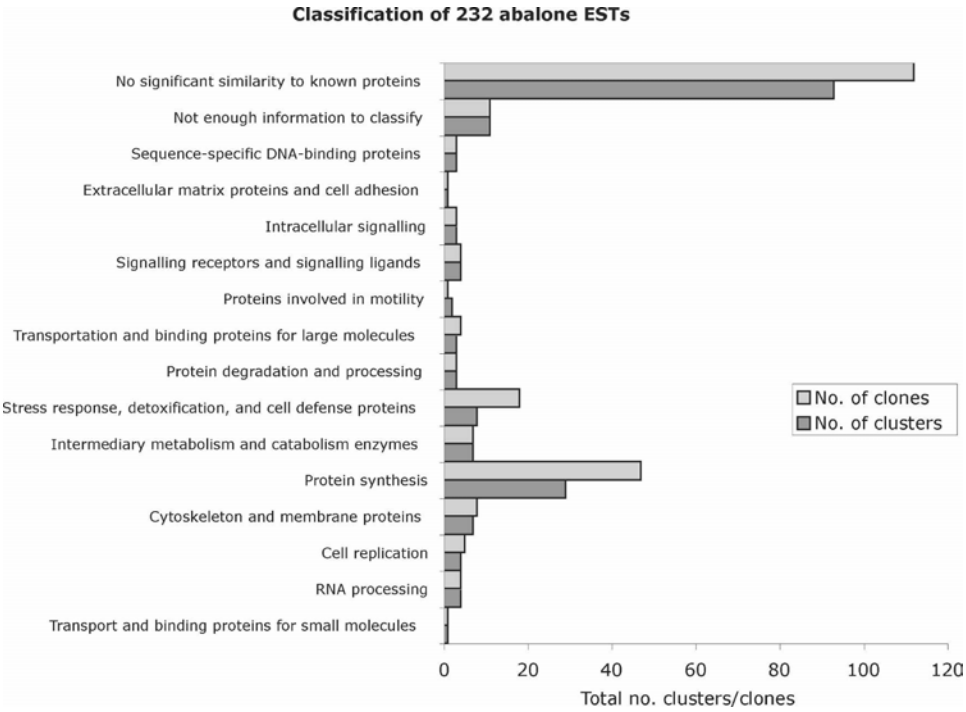

Figure 1. Prevalence of abalone ESTs grouped according to function.

that they play in the tropical abalone. For example, a significant proportion of the ESTs encode proteins that appear to be involved with stress response, detoxification, cell defense and innate immunity (Fig. 1). Analyses of differential gene expression in the ascidians *Boltenia villosa* and *Herdmania momus* during larval development and metamorphosis have revealed that a significant proportion of the expressed genes are involved with the innate immune system, with most of these being up-regulated during metamorphosis (Davidson & Swalla 2002, Woods et al. 2004). Components of the innate immune system may therefore be responsible for recognizing and responding to bacterial settlement cues associated with a suitable settlement environment. Bacterial biota associated with coralline algae may be responsible for induction of abalone settlement and metamorphosis (Bryan & Qian 1998, Roberts 2001).

TABLE 3.

Prevalence of polyadenylation signals among abalone and human mRNA transcripts.

Polyadenylation Signal	Prevalence (%) and Location[1] in Abalone ESTs	Prevalence (%) and Location[1] in Human ESTs
AATAAA	(55.1) 15.3 ± 3.9	(58.2) 16 ± 4.7
ATTAAA	(12.3) 19.3 ± 7.9	(14.9) 17 ± 5.3
AGTAAA	(6.1) 16.7 ± 1.5	(2.7) 16 ± 5.9
TATAAA	(0.0)	(3.2) 18 ± 7.8
CATAAA	(2.0) 13.0 ± NA	(1.3) 17 ± 5.9
GATAAA	(2.0) 18 ± NA	(1.3) 18 ± 6.9
AATATA	(4.1) 18 ± 1.4	(1.7) 18 ± 6.9
AATACA	(2.0) 24 ± NA	(1.2) 18 ± 8.7
AATAGA	(2.0) 15 ± NA	(0.7) 18 ± 6.3
AAAAAG	(4.1) 22.5 ± 13.4	(0.8) 18 ± 8.9
ACTAAA	(2.0) 26 ± NA	(0.6) 17 ± 8.1
GACAAA[2]	(2.0) 21 ± NA	—
ATGAAA[2]	(4.1) 18 ± 2.8	—
AGAAAA[2]	(2.0) 16 ± NA	—
Total	100%	87%

[1] The location reported here is the position of the sixth nucleotide of the hexamer relative to the beginning of the poly A tail.

[2] These 3 hexamers were not reported by Beaudoing et al. (2000) but are included here as they only differ by one nucleotide from recognized polyadenylation signals, and were located within 30 nucleotides of the poly A rich 3′ region, suggesting that they may be alternative abalone polyadenylation signals.

Genes involved in cellular metabolism are also well represented in this EST set (7 clusters). Because larvae of *H. asinina* are lecithotrophic (nonfeeding), maternal energy reserves are utilized during larval development and metamorphosis. Genes encoding certain digestive enzymes however only appear to be activated during late larval development (Degnan et al. 1995, Spaulding & Morse 1991), whereas enzymes required for algal digestion are not detectable until postlarvae are significantly older (Takami et al. 1998). The proteases and metabolic enzymes identified here provide a platform from which to further our understanding of the processes that regulate and limit larval and juvenile growth.

An FMRF-encoding gene was also detected in this EST survey. In the gastropod mollusc *Lymnaea stagnalis*, expression of this gene is affected by infection of a schistosome parasite. This seems to be a strategy of the parasite to induce physiological (growth and reproduction) and behavioral changes in the host (Hoek et al. 1997). Among natural populations *H. asinina* can become infected by a digenean parasite that has been suggested to influence the antagonism between growth and reproduction cycles (Lucas et al. 2005), with infected animals effectively being castrated. The FMRF-neuropeptide identified here provides a means with which to study this phenomenon (Rice & Degnan, unpublished data).

Of the 90 unique clusters that returned significant BLAST

matches, 11 (12.2%) could not be assigned a function. This value is less than that reported by Adams et al. (1995) for a human EST study (24.8% of 266,714 ESTs), but much larger than 5.2% reported by Lee et al. (1999) for a sea urchin EST analysis, and 4.1% reported by Davidson and Swalla (2002) for an ascidian EST analysis. These figures reflect both the exponential increase in the number of annotated sequences housed in the NCBI database since 1995 (the 1995 Genbank database was 3.7% the size of the 2001 database), and the fact that *H. asinina* belongs to the poorly studied Lophotrochozoan clade in terms of sequence representation and annotation.

After an analysis of those ESTs that showed no homology with any sequences in the GenBank database, Lee et al. (1999) estimated that 65% to 80% of these sequences were genuine protein coding sequences that were simply too divergent to match anything in the largely mammalian GenBank database. This estimate was based on an analysis of observed versus expected ORF length within each EST. Briefly, randomly generated sequences equal in length to those ESTs analyzed were assessed for the presence of the longest ORF. These derived ORFs rarely exceeded 150 base pairs. Here we analyzed novel clusters (clusters that showed no homology with GenBank sequences) for the presence of ORFs greater than 150 nucleotides in length, in an effort to identify genuine coding transcripts that have no homologues within GenBank. Of the 93 novel clusters, 71 (76.3%) possessed an ORF in excess of 50 codons. This is similar to Lee et al.'s (1999) estimate of *Strongylocentrotus purpuratus* specific mRNAs of 65% to 80%. Current developmental studies on *H. asinina* further support the notion of a high proportion of *H. asinina* specific genes (Jackson et al. 2005). Interestingly, studies on *C. elegans* and *Drosophila,* for which entire genomes are available, have revealed lower estimates of organism-specific coding sequences: 50% and 30% respectively of all recognized coding sequences (Harrison et al. 2002). If representative, this high proportion of novel coding mRNA transcripts for *H. asinina* further highlights the need for more studies of this kind to be conducted on the poorly studied Lophotrochozoan clade.

CONCLUSION

This small EST project has demonstrated that the *H. asinina* library we have constructed consists of a diverse set of cDNAs that are of a large enough size to be useful in future microarray and gene characterization studies. It has also shown that during development, embryos and larvae of the tropical abalone express genes involved in a range of functions, ranging from cell defense and innate immunity to signal transduction and metabolism. The percentage of EST sequences that do not have homologues in the NCBI database highlights the need for further EST projects to be conducted on more diverse organisms, especially taxa belonging to the Lophotrochozoa.

ACKNOWLEDGMENTS

This research was funded by an Australian Research Council grant to BMD.

LITERATURE CITED

Adams, M. D., A. R. Kerlavage, R. D. Fleischmann, R. A. Fuldner, C. J. Bult, N. H. Lee, E. F. Kirkness, K. G. Weinstock, J. D. Gocayne & O. White. 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 377(Supplement 6547):3–174.

Adoutte, A., G. Balavoine, N. Lartillot, O. Lespinet, B. Prud'homme & R. de Rosa. 2000. The new animal phylogeny: reliability and implications. *Proc. Natl. Acad. Sci. USA* 7:4453–4456.

Beaudoing, E., S. Freier, J. R. Wyatt, J. M. Claverie & D. Gauheret. 2000. Patterns of variant polyadenylation signal usage in human genes. *Gen. Res.* 10:1001–1010.

Bryan, P. J. & P. Y. Qian. 1998. Induction of larval attachment and metamorphosis in the abalone *Haliotis diversicolor* (Reeve). *J. Exp. Mar. Biol. Ecol.* 223:39–51.

Capinpin, E. C. & K. G. Corre. 1996. Growth rate of the Philippine abalone, *Haliotis asinina* fed an artificial diet and macroalgae. *Aquaculture* 144:81–89.

Capinpin, E. C., J. D. Toledo, V. C. Encena & M. Doi. 1999. Density dependent growth of the tropical abalone Haliotis asinina in cage culture. *Aquaculture* 171:227–235.

Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins & J. D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31:3497–3500.

Chomczynski, P. & N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156–159.

Counihan, R., D. C. McNamara, D. C. Souter, E. J. Jebreen, N. P. Preston, C. R. Johnson & B. M. Degnan. 2001. Pattern, synchrony and predictability of spawning of the tropical abalone *Haliotis asinina* from Heron Reef, Australia. *Mar. Ecol. Prog. Ser.* 213:193–202.

Davidson, B. & B. J. Swalla. 2002. A molecular analysis of ascidian metamorphosis reveals activation of an innate immune response. *Development* 129:4739–4751.

Degnan, B. M., J. C. Groppe & D. E. Morse. 1995. Chymotrypsin mRNA expression in digestive gland amoebocytes: cell specification occurs prior to metamorphosis and gut morphogenesis in the gastropod, *Haliotis rufescens. Roux. Arch. Dev. Biol.* 205:97–101.

Giusti, A. F., V. F. Hinman, S. M. Degnan, B. M. Degnan & D. E. Morse. 2000. Expression of a Scr/Hox5 gene in the larval central nervous system of the gastropod Haliotis, a non-segmented spiralian lophotrochozoan. *Evol. Dev.* 2:294–302.

Gordon, H. R. & P. Cook. 2001. World abalone supply, markets and pricing: historical, current and future. *J. Shellfish Res.* 20:567–570.

Harrison, P. M., A. Kumar, N. Lang, M. Snyder & M. Gerstein. 2002. A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.* 30:1083–1090.

Hinman, V. F. & B. M. Degnan. 2002. *Mox* homeobox expression in muscle lineage of the gastropod *Haliotis asinina*: evidence for a conserved role in bilaterian myogenesis. *Dev. Genes Evol.* 212:141–144.

Hinman, V. F., E. K. O'Brien, G. S. Richards & B. M. Degnan. 2003. Expression of anterior Hox genes in larvae of the gastropod. Haliotis asinina. *Evol. Dev.* 5:508–521.

Hoek, R. M., R. E. van Kesteren, A. B. Smit, M. de Jong-Brink & W. P. M. Geraerts. 1997. Altered gene expression in the host brain caused by a trematode parasite: Neuropeptide genes are preferentially affected during parasitosis. *Proc. Natl. Acad. Sci. USA* 94:14072–14076.

Jackson, D., K. Williams & B. M. Degnan. 2001. Suitability of Australian formulated diets for the aquaculture of the tropical abalone *Haliotis asinina* Linnaeus. *J. Shellfish Res.* 20:627–636.

Jackson, D. J., N. Ellemor & B. M. Degnan. 2005. Correlating gene expression with larval competence, and the effect of age and parentage on metamorphosis in the tropical abalone *Haliotis asinina. Mar. Biol.* 147:681–697.

Jarayabhand, P. & N. Paphavasit. 1996. A review of the culture of tropical abalone with special reference to Thailand. *Aquaculture* 140:159–168.

Jebreen, E. J., R. T. Counihan, D. R. Fielder & B. M. Degnan. 2000.

Synchronous oogenesis during the semilunar spawning cycle of the tropical abalone *Haliotis asinina. J. Shellfish Res.* 19:845–851.

Lee, Y.-H., G. M. Huang, R. A. Cameron, G. Graham, E. H. Davidson, L. Hood & R. J. Britten. 1999. EST analysis of gene expression in early cleavage-stage urchin embryos. *Development* 126:3857–3867.

Lehninger, A. L., D. L. Nelson & M. M. Cox. 1993. Principles of biochemistry. New York: Worth Publishers.

Lucas, T., E. K. O'Brien, T. Cribb & B. M. Degnan. 2005. Digenean trematodes infecting the tropical abalone *Haliotis asinina* have species-specific cercarial emergence patterns that follow daily or semilunar spawning cycles. *Mar. Biol.* 148:285–292.

Nateewathana, A. & J. Hylleberg. 1986. Thai species of abalone (*Haliotis* spp.). Is aquaculture feasible? *Thai Fisheries Gazette* 39:10–17.

O'Brien, E. K. & B. M. Degnan. 2002a. Developmental expression of a class IV *POU* gene in the gastropod *Haliotis asinina* supports a conserved role in sensory cell development in bilaterians. *Dev. Genes Evo.* 212:394–398.

O'Brien, E. K. & B. M. Degnan. 2002b. Pleiotropic developmental expression of *HasPOUIII*, a class III *POU* gene, in the gastropod *Haliotis asinina. Mech. Dev.* 114:129–202.

O'Brien, E. K. & B. M. Degnan. 2003. Expression of *Pax258* in the gastropod statocyst: insights into the antiquity of metazoan geosensory organs. *Evol. Dev.* 5:572–578.

Oakes, F. R. & R. D. Ponte. 1996. The abalone market: opportunities for cultured abalone. *Aquaculture* 140:187–195.

Roberts, R. 2001. A review of settlement cues for larval abalone (Haliotis spp.). *J. Shellfish Res.* 20:571–586.

Sambrook, J. & D. W. Russell. 2001. Molecular cloning: a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Selvamani, M. J. P., S. M. Degnan & B. N. Degnan. 2001. Microsatellite genotyping of individual abalone larvae: parentage assignment in aquaculture. *Mar. Biotechnol.* 3:478–485.

Selvamani, M. J. P., S. M. Degnan, D. Paetkau & B. M. Degnan. 2000. Highly polymorphic microsatellite loci in the Heron Reef population of the tropical abalone Haliotis asinina. *Mol. Ecol.* 9:1184–1186.

Singhagraiwan, T. & M. Doi. M. 1993. Seed production and culture of a tropical abalone, *Haliotis asinina* Linne: the research project of fishery resource development in the kingdom of Thailand.

Spaulding, D. C. & D. E. Morse. 1991. Purification and characterization of sulfatases from *Haliotis rufescens*: evidence for changes in synthesis and heterogeneity during development. *J. Comp. Physiol.* 161:498–515.

Takami, H., T. Kawamura & Y. Yamashita. 1998. Development of polysaccharide degredation activity in postlarval abalone *Haliotis discus hannai. J. Shellfish Res.* 17:723–727.

Woods, R. G., K. Roper, M. Gauthier, L. Bebell, K. Sung, M. F. Lavin & B. M. Degnan. 2004. Gene expression during early ascidian metamorphosis requires signaling by Hemps, an EGF-like protein. *Development* 131:2921–2933.