Distinct profiles of expressed sequence tags during intestinal regeneration in the sea cucumber *Holothuria glaberrima*

Carmencita Rojas-Cartagena,1 Pablo Ortíz-Pineda,1 Francisco Ramírez-Gómez,1 Edna C. Suárez-Castillo,1 Vanessa Matos-Cruz,1 Carlos Rodríguez,2 Humberto Ortíz-Zuazaga,2 and José E. García-Arrarás

1Department of Biology, Rio Piedras Campus, and 2High Performance Computing Facility, University of Puerto Rico, San Juan, Puerto Rico

Submitted 20 October 2006; accepted in final form 19 June 2007

Distinct profiles of expressed sequence tags during intestinal regeneration in the sea cucumber *Holothuria glaberrima*. Physiol Genomics 31: 203–215, 2007. First published July 3, 2007; doi:10.1152/physiolgenomics.00228.2006.—Repair and regeneration are key processes for tissue maintenance, and their disruption may lead to disease states. Little is known about the molecular mechanisms that underlie the repair and regeneration of the digestive tract. The sea cucumber *Holothuria glaberrima* represents an excellent model to dissect and characterize the molecular events during intestinal regeneration. To study the gene expression profile, cDNA libraries were constructed from normal, 3-day, and 7-day regenerating intestines of *H. glaberrima*. Clones were randomly sequenced and queried against the nonredundant protein database at the National Center for Biotechnology Information. RT-PCR analyses were made of several genes to determine their expression profile during intestinal regeneration. A total of 5,173 sequences from three cDNA libraries were obtained. About 46.2, 35.6, and 26.2% of the sequences for the normal, 3-days, and 7-days cDNA libraries, respectively, shared significant similarity with known sequences in the protein database of GenBank but only present 10% of similarity among them. Analysis of the libraries in terms of functional processes, protein domains, and most common sequences suggests that a differential expression profile is taking place during the regeneration process. Further examination of the expressed sequence tag dataset revealed that 12 putative genes are differentially expressed at significant level ($R > 6$). Experimental validation by RT-PCR analysis reveals that at least three genes (unknown C-4677-1, melanotransferrin, and centaurin) present a differential expression during regeneration. These findings strongly suggest that the gene expression profile varies among regeneration stages and provide evidence for the existence of differential gene expression.

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org). Address for reprint requests and other correspondence: J. García-Arrarás, Dept. of Biology, Univ. of Puerto Rico, JGD Bldg. Rm. 220, San Juan, PR 00931 (e-mail: jegarcia@hcpf.upr.edu).

1094-8341/07 $8.00 Copyright © 2007 the American Physiological Society
11, 39), little is known about the genes involved. Previous attempts to identify the genes associated with intestinal regeneration in H. glaberrima have been made in our laboratory using a gene by gene strategy, which is time consuming and provided limited information (37, 49, 51). Expressed sequence tags (ESTs) provide an alternative method for identifying regeneration-associated genes. EST analysis allows rapid gene discovery, clarification of gene function, and the identification of stage specific gene expression profiles (1, 29). EST analysis was used to test our hypothesis that the genetic expression profile changes during the different stages of regeneration. We generated an EST dataset from three cDNA libraries: two libraries from stages 1 and 2 of regenerating intestine and one from normal noneviscerated intestine of H. glaberrima. The ESTs obtained provide a preliminary view of the gene expression pattern in regenerating vs. normal animals and a valuable insight into the genes that may be involved in the regeneration process. The gene expression profile helps identify genes differentially expressed during the initial regeneration stages and in particular serves to discover genes homologous to those associated with disease or healing in humans.

MATERIALS AND METHODS

Animals. Adult sea cucumbers, H. glaberrima, were obtained from the coastal waters of Puerto Rico and maintained in seawater aquaria at 22–24°C. Evisceration was induced by injections (3–5 ml) of 0.35 M KCl into the coelomic cavity. Prior to dissection, animals were anesthetized by placement in ice cold water for 1 h and then killed by cutting through the calcareous ring. The intestines of noneviscerated and regenerating animals at 3dpe and 7dpe were removed under RNase-free conditions and stored in RNAlater (Ambion, Austin, TX) until use.

cDNA libraries. Unidirectional cDNA libraries were constructed in the bacteriophage λ-ZAPII according to the procedures indicated in the cDNA synthesis and in vitro phage packaging of Stratagene’s Uni-Zap kit (Stratagene, La Jolla, CA). The 7dpe (37) and the noneviscerated (normal) (49) cDNA libraries were previously described. The 3dpe cDNA library was constructed by Ampiclon Express (Pullman, WA) using 2 mg of total RNA extracted with the RNAeasy midi prep kit from Qiagen (Valencia, CA). Mass excisions of the three cDNA libraries were performed using the Escherichia coli strain XL-1 Blue and phage supernatant according to the cDNA library manual (Stratagene). All three libraries were amplified but not normalized.

Transformed bacteria were plated in medium containing ampicillin to select for growth of those with an insert. Individual colonies were randomly picked and inoculated into 96-well microtiter plates. Colony PCR were performed using M13 primers (Sigma Genosys, St. Louis, MO). The PCR products were analyzed by electrophoresis using 1.0% agarose gels and cleaned by ExoSap (USB, Cleveland, OH) prior to sequencing. The percentage of insert amplification was determined by dividing the number of amplified clones by the number of total clones submitted to PCR.

Nucleotide sequencing. Single-pass sequencing was performed on each PCR template using the T3 primer (5′-AATTAACCCCTACTAAAGGG-3′). PCR products were sequenced in 96-well microplates using DYEnamic ET dye terminator kit diluted four times with reaction buffer (1 M Tris-HCl, pH 9.0, 1 M MgCl) (Amersham Pharmacia Biotech, Piscataway, NJ) Sequencing was performed on a MegaBACE 500 capillary sequencer (Amersham Pharmacia Biotech). Samples were injected at 3 Kv for 50 s and run at 9 Kv for 120 min.

Sequence analysis, contig assembly, and BLAST similarity. The chromatogram produced for each clone was analyzed using Chromas 2.13 (Techneleysium Pty, Helensvale, Australia). Quality control of sequenced ESTs was performed with Phred (21) with a cutoff of 20 for trimming low-quality regions, and Cross-Match (16) for vector trimming.

ESTAP (EST Analysis Pipeline) (35), a specialized database for ESTs, was used for assembly and analysis of the sequences. The ESTs obtained from cDNA sequencing were grouped into three projects: two for regenerating tissues (3 days and 7 days) and one for normal intestinal tissue. Initially, two subprojects for the 7dpe cDNA library (7dpe A and 7dpe B performed at different times and conditions) were maintained to corroborate the reproducibility of the results. The 7dpe A project mass excision and sequencing had been performed 2 yr before the 7dpe B project. After reproducibility was determined, the projects were fused together for further analysis. The contig assembly for the three libraries was performed using CAP3 (25) set to the default parameters. The contigs for each library were analyzed using a Venn diagram to determine the percentage of similarity among them (21, 54, 61).

Each cDNA sequence was queried against the nonredundant protein database at the National Center for Biotechnology Information (NCBI, Bethesda, MD) using the BLASTX algorithm (4). In both cases, the default BLAST parameters were used. BioParser, an EST database manager (12) was used to classify Blast results using an E-value of 1e-3 or less. After analysis, a putative product was assigned to the results from the similarity search (BLASTX). ESTs with an E-value < 1e-7 were classified according to their probable function. The ESTs significantly similar to hypothetical proteins were placed in the “unclassified” category. Sequences were also queried against those at the NCBI for the Holothuroidea and against a custom database of all known Strongyllocentrotus purpuratus (NCBI, taxonomy ID: 7668), sequences extracted from GenBank (8) with the NCBI taxonomy Browser (62). The sea urchin S. purpuratus was selected because of its phylogenetically proximity and the extent of the sequencing efforts in this organism (45).

Domain and electronic annotation for contigs. Domains were searched with RPS-BLAST against the conserved domain database (35) from NCBI. BLAST, Gene Ontology (GO) annotation, and domain search output files were parsed for homologous sequences, with an E-value of 1e-7 as a cutoff for BLASTX searches against the sequence databases. The default cut-off of 0.01 was considered to yield significant similarity to conserved domains from the Conserved Domain Database (CDD). Biological process, molecular function, and cellular component were assigned to contigs based on the GO annotation of the closest annotated homolog (6). For this purpose, the GenBank annotation files from the GO database were downloaded and parsed for the gene identifier numbers of previously identified homologs.

In silico differential display of H. glaberrima contigs from regenerating and normal intestine. The statistic R test (55) was used to compare the abundance of genes in the three cDNA libraries (3dpe, 7dpe, and normal) and between normal and regenerating datasets under the null hypothesis that all libraries are equally distributed. A high value indicates high divergence of the gene between libraries.

The statistic R was obtained using the logarithmic likelihood ratio between the observed amount of fragments for a specific gene and its expected values. This statistic tends asymptotically to a X² distribution (55). To validate the statistic R and select the significance threshold, the two methods suggested by Stekel and colleagues (55) were applied. First, a randomization procedure to determine the true differential expression and, second, application of the theory of large deviations.

We randomly generated 1,000 datasets and calculated the R statistic under the null hypothesis. The number of genes for each R value was averaged over the 1,000 runs and the three libraries. The significance was established under 99% of believability. The biological significance is given by the putative genes for which the statistic R value was greater than the threshold and consequently in the 1% of the sample
distribution. In our case, the statistic significant values of $R$ were $>0.6$.

Gene expression analysis. The differential gene expression of several genes (unknown c-4766-1, unknown NADH, melanotransferrin, and centaurin) was semiquantitatively determined by gene-specific relative RT-PCR using cytochrome $b$ (168 bp) for normalization. Total RNA was extracted from the intestine tissues after homogenization using Tri-Reagent (MRC, Cincinnati, OH), following the manufacturer’s specifications. Quality of the extracted RNA was evaluated by agarose electrophoresis, and quantity was determined by spectrophotometry at 260-nm optical density in a NanoDrop (ND-1000). For gene-specific relative RT-PCR, first-strand synthesis was performed with the RETROscript First Strand Synthesis Kit following the manufacturer’s specifications (Ambion). In brief, 1 µg of total RNA was incubated with random decamers for 3 min at 80°C. After incubation, dNTPs, buffer, RNase inhibitor (Ambion), and reverse transcriptase were added, and the reaction was incubated at 44°C for 1 h. This was followed by a 10-min incubation at 92°C. The single-stranded cDNA obtained was amplified by PCR as described below. The primers were designed for optimal performance using the Whitehead Institute’s Primer3 and Net primer. The primers were design to amplify a region of 207 bp of the unknown C-4766-1, 237 bp for unknown 4860-1, 241 bp for NADH, 335 bp for melanotransferrin, and 257 for centaurin. Primers for melanotransferrin, centaurin, and cytochrome $b$ were synthesized at Sigma Genosys (St. Louis, MO). Primers for NADH and the unknown C-4766-1 at AlfaDNA (Montreal, Canada). Cycling conditions for the amplified products were as follow: unknown C-4766-1, unknown 4860-1, and NADH 94°C 45 s, 55°C 1 min, 72°C; melanotransferrin 94°C 45 s, 58°C 1 min, 72°C; centaurin 94°C 45 s, 56°C 1 min, 72°C; all performed for 35 cycles.

Data analysis for RT-PCR was carried out using the SPSS Base 8.0.1 statistical package (SPSS). Values are presented as the means ± SE. Differences between the means of each group were analyzed using the Student $t$-test.

RESULTS

cDNA libraries and EST sequence quality analysis. Three independent cDNA libraries from normal and regenerating intestine (stage 1, 3dpe and stage 2, 7dpe) were sequenced to generate an EST dataset. A summary comparison of the three cDNA libraries is provided on Table 1. We found that the titers and the percent of phagemids with inserts were similar for the normal, 3dpe, and 7dpe intestine cDNA libraries (106–107 and the percent of phagemids with inserts were similar for the cDNA libraries is provided on Table 1. We found that the titers generate an EST dataset. A summary comparison of the three libraries. pfu, Plaque-forming unit; dpe, day posteviceration.

Table 1. Comparison of the 3dpe, 7dpe, and nonevicerated cDNA libraries

<table>
<thead>
<tr>
<th>cDNA Library</th>
<th>pfu</th>
<th>% of Insert Amplification</th>
<th>Average Insert Size</th>
<th># of Sequenced Clones</th>
<th># of Clean ESTs</th>
<th>Read Sequence Length Average</th>
<th>Average Sequence Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>3dpe</td>
<td>$10^6$</td>
<td>90</td>
<td>800</td>
<td>2,161</td>
<td>1,252</td>
<td>400</td>
<td>37</td>
</tr>
<tr>
<td>7dpe</td>
<td>$10^7$</td>
<td>97</td>
<td>1,249</td>
<td>2,884</td>
<td>2,217</td>
<td>600</td>
<td>45</td>
</tr>
<tr>
<td>Normal</td>
<td>$10^8$</td>
<td>98</td>
<td>1,278</td>
<td>1,906</td>
<td>1,704</td>
<td>700</td>
<td>50</td>
</tr>
</tbody>
</table>

Percentage of amplification and insert size as well as the number of expressed sequence tags (ESTs) and their sequence quality are shown for the 3 cDNA libraries. pfu, Plaque-forming unit; dpe, day posteviceration.

cDNA efficiency assessment. The efficiency and reproducibility of our protocols were corroborated by comparing two sets of ESTs for the 7dpe library performed at different times. The 7dpe (A) library had 773 ESTs and the 7dpe (B) had 1,444 ESTs, for a total of 2,217 sequences. The efficiencies of the two datasets were similar (72% for 7dpe A and 80% for 7dpe B). Moreover, genes that were previously reported in the 7dpe (A) such as actin, ferritin, thyromosin, and QM protein were also identified in the 7dpe (B) dataset as well. In addition, genes such as collagen, serum amyloid, and ependymin, which were previously found by Northern blot and/or RT-PCR analyses to be expressed at 7dpe (49, 51, 56), were identified in both EST datasets.

EST BLASTX analysis. ESTs were subjected to BLASTX searches against the nonredundant database (NCBI) for the identification of putative homologies to known proteins as of August 2006. Of the total dataset, 3,242 sequences (72.2%) had no similarity with known genes. About 90% of these sequences presented an open reading frame (Table 2). A total of 1,112 sequences (21%) shared similarity with previously described proteins. The E-value distribution of these 1,112 sequences in the nonredundant database is shown in Fig. 2A. The 3% of the ESTs that matched a protein with an E-value $<1e-100$ could be considered true orthologs (23). About 65% of the ESTs had an E-value between 1e-20 and 1e-99, and 32% E-value 1e-03 to 1e-19, presenting significant and weak similarity, respectively. We also compared our ESTs dataset against the ESTs database of the sea urchin S. purpuratus (NCBI), finding that 1,905 (38%) of the sequences shared similarity with a putative protein in the sea urchin database. The E-value distribution of these 1,905 sequences in the sea urchin is shown on Fig. 2B.
About 2% of the ESTs have an E-value < 1e-100, which might be considered as true orthologs, 77% had E-values between 1e-20 and 1e-99, presenting significant similarity, and 21% had weak similarity (E-value 1e-03 to 1e-19). Our dataset was also searched against the 657 ESTs and 14 genes reported in the NCBI for different species of holothurians and found that 97.4% of these ESTs are present in our dataset. These sequences correspond to 639 ESTs in our dataset; thus 87% of our sequences are novel, enriching the existing database for holothurians with 4,534 new sequences. From these new 4,534 sequences, 2,348 are nonredundant.

We analyzed the extent of overlap among the three cDNA libraries generated from regenerating and normal intestine to determine the percentage of shared transcripts between them. The libraries were ~10% similar, which suggested that stage-specific gene expression is represented in each cDNA library. To further characterize genetic profile of the libraries, the 10 contigs with the highest EST frequency for each library were determined (Table 3). As expected, some contigs such as cytochrome c oxidase, ATP synthase, and NADH dehydrogenase were shared among the libraries. Also, even though ribosomal ESTs were present among the libraries, some ribo-

---

Table 2. Match to GenBank sequences of ESTs from 3dpe, 7dpe, and noneviscerated intestines cDNA libraries

<table>
<thead>
<tr>
<th>cDNA Library</th>
<th>% of Identified Proteins, E-value &lt; 10^-7</th>
<th>% of Hypothetical Proteins, E-value &lt; 10^-4</th>
<th>% of Unidentified Sequences</th>
<th>% Sequences With ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>3dpe</td>
<td>35.6 (9.37 ribosomal)</td>
<td>12.4</td>
<td>52.0</td>
<td>93.5</td>
</tr>
<tr>
<td>7dpe</td>
<td>26.2 (10.70 ribosomal)</td>
<td>8.6</td>
<td>65.2</td>
<td>90.7</td>
</tr>
<tr>
<td>Normal</td>
<td>46.2 (4.96 ribosomal)</td>
<td>5.5</td>
<td>48.3</td>
<td>89.0</td>
</tr>
</tbody>
</table>

Percentage of the identified proteins, hypothetical proteins, and unidentified sequences present in the 3 cDNA libraries. “% of Hypothetical Proteins” represents the group of ESTs in each cDNA library that presents a match with a hypothetical protein in the database. “% of Unidentified Sequences” represents the group of ESTs in each cDNA library without any similarity match in GenBank (unknowns). “% Sequences With ORF” represents the group of unknown ESTs with an open reading frame (ORF).

---

Table 3. Summary of most abundant contigs among normal, 3dpe, and 7dpe cDNA libraries of Holothuria glaberrima

<table>
<thead>
<tr>
<th>No. of ESTs, %</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>cytochrome c oxidase subunit 1</td>
</tr>
<tr>
<td>138 (8.1)</td>
<td>unknown (4277-3)</td>
</tr>
<tr>
<td>96 (5.6)</td>
<td>ferritin</td>
</tr>
<tr>
<td>58 (3.4)</td>
<td>cytochrome c oxidase subunit 3</td>
</tr>
<tr>
<td>33 (1.9)</td>
<td>cytochrome c oxidase subunit 2</td>
</tr>
<tr>
<td>21 (1.2)</td>
<td>cytochrome b</td>
</tr>
<tr>
<td>14 (0.8)</td>
<td>thymosin beta-4 (T beta 4)</td>
</tr>
<tr>
<td>13 (0.8)</td>
<td>unknown (125-4)</td>
</tr>
<tr>
<td>13 (0.8)</td>
<td>ATP synthase a chain</td>
</tr>
<tr>
<td>11 (0.6)</td>
<td>NADH dehydrogenase subunit 4</td>
</tr>
<tr>
<td>3dpe</td>
<td>unknown (4277-3)</td>
</tr>
<tr>
<td>104 (8.3)</td>
<td>unknown (125-4)</td>
</tr>
<tr>
<td>18 (1.4)</td>
<td>unknown (4991-1)</td>
</tr>
<tr>
<td>14 (1.1)</td>
<td>cytochrome b</td>
</tr>
<tr>
<td>13 (1.0)</td>
<td>unknown (4791-1)</td>
</tr>
<tr>
<td>13 (1.0)</td>
<td>cytochrome c oxidase subunit III</td>
</tr>
<tr>
<td>12 (1.0)</td>
<td>ferritin GF1</td>
</tr>
<tr>
<td>12 (1.0)</td>
<td>ATP synthase F0 subunit 6</td>
</tr>
<tr>
<td>12 (1.0)</td>
<td>unknown (4766-1)</td>
</tr>
<tr>
<td>11 (0.9)</td>
<td>ribosomal protein L7-like</td>
</tr>
<tr>
<td>7dpe</td>
<td>unknown (4277-3)</td>
</tr>
<tr>
<td>122 (5.5)</td>
<td>cytochrome c oxidase subunit i</td>
</tr>
<tr>
<td>44 (2.0)</td>
<td>cytochrome b</td>
</tr>
<tr>
<td>31 (1.4)</td>
<td>unknown (5020-1)</td>
</tr>
<tr>
<td>19 (0.9)</td>
<td>ATP synthase F0 subunit 6</td>
</tr>
<tr>
<td>17 (0.8)</td>
<td>cytochrome c oxidase subunit II</td>
</tr>
<tr>
<td>16 (0.7)</td>
<td>cytochrome c oxidase subunit III</td>
</tr>
<tr>
<td>14 (0.6)</td>
<td>unknown (125-4)</td>
</tr>
<tr>
<td>13 (0.6)</td>
<td>actin</td>
</tr>
<tr>
<td>11 (0.5)</td>
<td>NADH dehydrogenase subunit 4</td>
</tr>
</tbody>
</table>

The 10 most frequent ESTs are shown for each library. Percentages are based on the total ESTs for each cDNA library. Percentage of the identified proteins, hypothetical proteins, and unidentified sequences present in the 3 cDNA libraries.
somal sequences were only identified in one of the libraries (L6 for normal, L36 for 3dpe, and L4 for 7dpe), suggesting a differential expression profile among them.

The ESTs were assembled into contigs to identify sequences representing the same transcript. Contig assembly was performed for each library, and 154, 134, and 262 contigs were found in the normal, 3dpe, and 7dpe cDNA libraries, respectively (Table 2). To determine the diversity of the libraries, redundancy was estimated based on the fraction of sequences that assembled into contigs. The estimated redundancy of the normal, 3dpe, and 7dpe cDNA libraries was 55, 58, and 53%, respectively. Some contigs were highly represented in one library but not the others. For example, sequences that have a match with actin were one of the largest contigs in the 7dpe cDNA library, while sequences with similarity to B-thymosin formed one of the largest contigs in the normal (nonviscerated) cDNA library. Interestingly, contigs without similarity to sequences in the databank were also among the largest in all three libraries. Some of these (4911-1, 4791-1, and 4766-1) were found to be mostly represented in the 3dpe library.

**Common domains in the H. glaberrima dataset.** Potential domains for the H. glaberrima contigs were identified performing an RPS-BLAST search against the conserved database (CDD, NCBI). About 42.9, 54.5, and 42.8% of the contigs of 3dpe, 7dpe, and normal libraries had a known protein domain. Domains such as Cox and ribosomal were identified in all three libraries (Table 4). Cellular functions were assigned to the identified domains and analyzed among the contigs. The most common domains over all three libraries were in the “oxidoreductase activity and protein biogenesis” category, which may have resulted from redundant annotations of cytochromes and ribosomal sequences. The “generation of precursor metabolites and energy” was the second most abundant category followed by “mitochondrial electron transport” category. Among the protein domains identified in the 3dpe library are nucleosplasmin, SAA, GroEL, A2M, and EGF-Lam. Also, proteins domain such as Band 7-prohibitin, vWFA, SapB, PP2c, and NDK were found only in the 7dpe library. Interestingly, one of the unknown contigs (#4766-1) highly represented in the 3dpe library presented a EF-hand domain, however, further characterization of this contig is needed.

**Functional category analysis.** The GO database was used to identify the biological process of the putative homologous proteins found among the ESTs libraries. We examined the assigned biological process to determine differences in the distribution of these processes among the libraries. For this purpose, the biological process was split into different categories (Fig. 3). The most prominent functional category in the normal cDNA library was metabolism. In the 3dpe and 7dpe libraries, metabolism and protein metabolism/modification appeared to have a similar representation. The percentage of ESTs involved in cell proliferation/apoptosis, cell adhesion, and intracellular signaling appeared to be higher in the 7dpe (7.9, 8.6, and 3.7%, respectively) than in the 3dpe (4.0, 0.5, and 1.9%, respectively) and normal (0.2, 1.4, and 0.7%, respectively) libraries (Fig. 3). In the 3dpe library the percentages of ESTs involved in inflammatory response were higher (3.5%) when compared with the 7dpe (2.0%) and normal (2.2%) libraries.

In addition to this overview of the expression profiles, we also examined our datasets for individual sequences potentially useful in studying regeneration, healing, and developmental processes (Table 4). We searched for sequences involved in regulating cell cycle, proliferation, and apoptosis or those that could play a role during extracellular matrix remodeling and signal transduction pathways. Several sequences matched putative homolog genes involved in the processes mentioned above such as matrix metalloproteinases (MMPs), melanotransferrin, and centaurin (Table 5).

**In silico differential display of the regenerating and normal H. glaberrima contigs.** To determine differences in sequence frequency at each stage, we compared the representation of the ESTs in the three libraries by performing an in silico differential display analysis. From 5,173 transcripts of the three cDNA libraries, 12 of them showed a significant differential expression under the null hypothesis that all genes are expressed equally among the libraries (Table 6). The statistic R resulted in a range of values from 0 to 25.75. These values were dependent on the amount of transcripts obtained in all libraries (Fig. 4). In our case, R < 6.0 meant only one false positive for every 17 significant genes. In contrast, if the statistic R > 6.0, the amount of significant genes was 12 with no false positives. Thus, our cut-off for statistical significance was R > 6.0.

**Gene expression analysis.** Gene-specific relative RT-PCR was performed to validate the expression of some of the ESTs that were found in our analyses (Fig. 5). The sequence of the primers used is presented in Table 7. Two of these were chosen from the R analyses and represented sequences that were expected to be overexpressed in animals at 3 days post-eversion. Two other sequences corresponded to housekeeping genes, NADH and cytochrome b that according to the R analyses were not expected to change their expression during regeneration. These genes are usually used as a reference in determining expression changes. Finally, two other genes that appeared only in the regenerating libraries were analyzed. These genes had high similarities to melanotransferrin and centaurin. Melanotransferrin formed a contig of 8 ESTs that were found in the 3dpe and 7dpe libraries, but not in the normal intestinal library. Unpublished experiments in our laboratory had shown the gene to be overexpressed in coelomocytes. Centaurin, was a unique sequence found in the 7dpe library, thus we were interested in studying the possibility that genes found only within the regenerating libraries were overexpressed during the regeneration process.

Our results show that only one of the analyzed genes was overexpressed at the 3dpe stage compared with 7dpe and normal intestines. This was the unknown C-4766-1, providing some validation to our R analyses that predicted this gene to be overexpressed at 3dpe (Fig. 5A). However, a second unknown showed no significant changes in expression, which might be due to a limitation in the sensitivity of the PCR analyses, to large variability in the expression within individuals, or to the possibility that not all genes identified by R are true positives.

Two other genes showed changes in their expression. Centaurin was overexpressed in the 7dpe animals compared with 3dpe and normal intestines, and melanotransferrin showed a trend for an increase in both 3 and 7dpe compared with normal intestines (Fig. 5A). This initial observation was followed with a more complete stage analysis to determine the temporal expression profile of these genes. We found that the expression of both genes was significantly overexpressed during intestinal regeneration but their expression profile differed. Centaurin
Table 4. **Total EST clones displaying similarity (BLASTX, E-value < 1E−07) to proteins in the nonredundant protein NCBI database, grouped into functional categories according to expressed gene anatomy database, partial representative list**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Protein Match</th>
<th>Species Match</th>
<th>E-value</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP8F3</td>
<td>histone protein Hist2 h3c1</td>
<td>Gallus gallus</td>
<td>9e-36</td>
<td>XP_425464</td>
</tr>
<tr>
<td>NP9C12</td>
<td>H3.3 histone</td>
<td>Paracentrotus lividus</td>
<td>1e-35</td>
<td>CAAS3692</td>
</tr>
<tr>
<td>7dpB2D4</td>
<td>development-specific protein-217</td>
<td>Tripneustes gratilla</td>
<td>4e-31</td>
<td>A30033</td>
</tr>
<tr>
<td>7dpB11F7</td>
<td>MAP kinase</td>
<td>Mus musculus</td>
<td>4e-22</td>
<td>NP_067437</td>
</tr>
<tr>
<td>7dpA15B6</td>
<td>G protein</td>
<td>Xenopus laevis</td>
<td>5e-92</td>
<td>AHA1541</td>
</tr>
<tr>
<td>NP7H1</td>
<td>calreticulin</td>
<td>Strongylocentrotus purpuratus</td>
<td>9e-54</td>
<td>NP_999643</td>
</tr>
</tbody>
</table>

**Chromosome structure/Signal transduction**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Protein Match</th>
<th>Species Match</th>
<th>E-value</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP18F1</td>
<td>cathepsin C</td>
<td>Gallus gallus</td>
<td>7e-36</td>
<td>XP_417207</td>
</tr>
<tr>
<td>3dp13F3</td>
<td>alpha-2-macroglobulin</td>
<td>Apis mellifera</td>
<td>7e-18</td>
<td>XP_392454</td>
</tr>
<tr>
<td>3dp8A6</td>
<td>serum amyloid A protein</td>
<td>Holothuria glaberrima</td>
<td>2e-41</td>
<td>A2G4633</td>
</tr>
<tr>
<td>7dpB9A1</td>
<td>fibrinogen-related protein</td>
<td>Parastichopus parvimensis</td>
<td>6e-28</td>
<td>A53084</td>
</tr>
<tr>
<td>7dpB13A10</td>
<td>lactotransferrin</td>
<td>Strongylocentrotus purpuratus</td>
<td>2e-47</td>
<td>XP_786179</td>
</tr>
<tr>
<td>NP15H1</td>
<td>lactatherin</td>
<td>Strongylocentrotus purpuratus</td>
<td>3e-21</td>
<td>XP_793834</td>
</tr>
<tr>
<td>7dpA14A6</td>
<td>Fc fragment IgG</td>
<td>Strongylocentrotus purpuratus</td>
<td>2e-16</td>
<td>XP_780608</td>
</tr>
<tr>
<td>7dpE3</td>
<td>development-specific protein-217</td>
<td>Drosophila yakuba</td>
<td>7e-13</td>
<td>AAJ9708</td>
</tr>
<tr>
<td>NP13F3</td>
<td>similar to interferon</td>
<td>Strongylocentrotus purpuratus</td>
<td>4e-10</td>
<td>XP_783602</td>
</tr>
<tr>
<td>NP16C11</td>
<td>similar to fibronectin</td>
<td>Strongylocentrotus purpuratus</td>
<td>6e-15</td>
<td>XP_797088</td>
</tr>
<tr>
<td>NP16E2</td>
<td>GalNac-specific lectin</td>
<td>Astera pectinifera</td>
<td>7e-12</td>
<td>BAA7859</td>
</tr>
<tr>
<td>3dp13D10</td>
<td>superoxide dismutase</td>
<td>Gryllotalpa orientalis</td>
<td>6e-57</td>
<td>AAV73809</td>
</tr>
</tbody>
</table>

**Stress response/Cell defense**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Protein Match</th>
<th>Species Match</th>
<th>E-value</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP2E7</td>
<td>proteasome</td>
<td>Strongylocentrotus purpuratus</td>
<td>6e-61</td>
<td>XP_780413</td>
</tr>
<tr>
<td>3dp8G2</td>
<td>poly(A)-binding protein 1</td>
<td>Gallus gallus</td>
<td>9e-28</td>
<td>XP_428547</td>
</tr>
<tr>
<td>3dp21A2</td>
<td>heterogeneous nuclear ribonuleoprotein K</td>
<td>Homo sapiens</td>
<td>5e-23</td>
<td>CAI16019</td>
</tr>
<tr>
<td>3dp20C1</td>
<td>translation elongation factor 1-alpha</td>
<td>Malassezia pachydermatis</td>
<td>4e-26</td>
<td>AAY46260</td>
</tr>
<tr>
<td>3dp24H9</td>
<td>protein transport (SEC61b)</td>
<td>Strongylocentrotus purpuratus</td>
<td>6e-17</td>
<td>XP_786999</td>
</tr>
<tr>
<td>7dpA3C7</td>
<td>transcription factor BIGMAX</td>
<td>Drosophila melanogaster</td>
<td>5e-19</td>
<td>AAG40148</td>
</tr>
<tr>
<td>NP2D2</td>
<td>vacuolar sorting protein</td>
<td>Box taurs</td>
<td>2e-52</td>
<td>XP_615172</td>
</tr>
<tr>
<td>7dpB6A3</td>
<td>QM</td>
<td>Pinicuda fucata</td>
<td>2e-94</td>
<td>AAN85578</td>
</tr>
<tr>
<td>7dpA15 h3</td>
<td>SSR gamma</td>
<td>Homo sapiens</td>
<td>7e-39</td>
<td>AA72075</td>
</tr>
<tr>
<td>7dpB9D10</td>
<td>solute carrier protein</td>
<td>Paralichthys olivaceus</td>
<td>3e-45</td>
<td>BA4E8204</td>
</tr>
<tr>
<td>7dpB8G2</td>
<td>matrix metalloproteinase</td>
<td>Notophthalmus viridescens</td>
<td>3e-27</td>
<td>AAX14804</td>
</tr>
<tr>
<td>7dpB5G7</td>
<td>ferritin</td>
<td>Crassostrea gigas</td>
<td>2e-55</td>
<td>AAPP8739</td>
</tr>
<tr>
<td>7dpB6E3</td>
<td>alpha 1 (V) collagen</td>
<td>Gallus gallus</td>
<td>7e-24</td>
<td>AAF28099</td>
</tr>
<tr>
<td>3dp23B3</td>
<td>actin</td>
<td>Hydrodida elegans</td>
<td>3e-28</td>
<td>BAE91915</td>
</tr>
<tr>
<td>NP2F2</td>
<td>myosin light chain</td>
<td>Gallus gallus</td>
<td>2e-11</td>
<td>P02065</td>
</tr>
<tr>
<td>NP2C3</td>
<td>tubulin beta2</td>
<td>Danio rerio</td>
<td>2e-85</td>
<td>AAQ78589</td>
</tr>
<tr>
<td>NP5C5</td>
<td>thymosin beta-4 precursor</td>
<td>Rattus norvegicus</td>
<td>7e-11</td>
<td>J52084</td>
</tr>
<tr>
<td>7dpA4F9</td>
<td>tektin</td>
<td>Strongylocentrotus purpuratus</td>
<td>2e-59</td>
<td>NP_999787</td>
</tr>
<tr>
<td>7dpA2E2</td>
<td>dynein</td>
<td>Canis familiaris</td>
<td>9e-44</td>
<td>XP_537691</td>
</tr>
</tbody>
</table>

**Protein synthesis, processing, storage, transport, and degradation**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Protein Match</th>
<th>Species Match</th>
<th>E-value</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7dpA3B9</td>
<td>translationally control tumor protein</td>
<td>Branchiostoma belcheri</td>
<td>7e-23</td>
<td>AAK84394</td>
</tr>
<tr>
<td>7dpB10F8</td>
<td>survivin</td>
<td>Xenopus laevis</td>
<td>2e-22</td>
<td>A1K0480</td>
</tr>
<tr>
<td>7dpA8F11</td>
<td>programmed cell death 4</td>
<td>Xenopus tropicalis</td>
<td>5e-11</td>
<td>AHA76698</td>
</tr>
<tr>
<td>7dpA15D10</td>
<td>novel death regulatory protein (GRIM-19)</td>
<td>Rattus norvegicus</td>
<td>7e-35</td>
<td>XP_214305</td>
</tr>
</tbody>
</table>

**Cell proliferation/Apoptosis**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Protein Match</th>
<th>Species Match</th>
<th>E-value</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7dpA15B8</td>
<td>slit</td>
<td>Gallus gallus</td>
<td>2e-16</td>
<td>XP_416992</td>
</tr>
<tr>
<td>7dpB6E10</td>
<td>ependymin</td>
<td>Holothuria glaberrima</td>
<td>4e-55</td>
<td>AAR89380</td>
</tr>
<tr>
<td>7dpB10H6</td>
<td>laminin</td>
<td>Tripneustes gratilla</td>
<td>4e-65</td>
<td>U02371</td>
</tr>
<tr>
<td>7dpA1D9</td>
<td>synaptotagmin</td>
<td>Danio rerio</td>
<td>6e-14</td>
<td>NP_999861</td>
</tr>
</tbody>
</table>

**Nervous system**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Protein Match</th>
<th>Species Match</th>
<th>E-value</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7dpA8F5</td>
<td>vitellogenin</td>
<td>Pseudocentrotus depressus</td>
<td>4e-44</td>
<td>AAK57983</td>
</tr>
<tr>
<td>3dp7D7</td>
<td>Par-3</td>
<td>Strongylocentrotus purpuratus</td>
<td>6e-13</td>
<td>XP_785679</td>
</tr>
<tr>
<td>7dA9D3</td>
<td>Wnt-14</td>
<td>Gallus gallus</td>
<td>1e-37</td>
<td>XP_784191</td>
</tr>
<tr>
<td>7dB1E9</td>
<td>sprout growth factor</td>
<td>Strongylocentrotus purpuratus</td>
<td>2e-39</td>
<td>XP_784191</td>
</tr>
<tr>
<td>3dp17C1</td>
<td>calponin</td>
<td>Branchiostoma belcheri</td>
<td>3e-23</td>
<td>BAC16745</td>
</tr>
</tbody>
</table>

NCBI, National Center for Biotechnology Information.
was overexpressed at 7dpe to 21dpe with a peak in expression at 14dpe ($P < 0.001$). Melanotransferrin was overexpressed during the first week (3–7dpe, $P < 0.5$) of regeneration compared with the later stages (21–28dpe) (Fig. 5B). No changes were observed in either NADH or in cytochrome $b$, where the latter was used for normalization purposes.

**DISCUSSION**

**EST BLAST analysis.** A total of 5,173 EST sequences were analyzed to determine gene expression in normal intestine and in 3dpe and 7dpe regenerating intestine. Since the three sea cucumber libraries were not normalized, we expected a significant number of cDNAs to correspond to abundant transcripts present in intestinal tissue. However, contig analysis indicated that redundancy was relatively low among the libraries (in the range of 39–55%) compared with that reported for normalized cDNA libraries (54%) (44) and for nonnormalized libraries (74%) (32). The redundancy levels found in our libraries are similar to those seen in EST analyses from other organisms such as the fungi *Mycosphaerella graminicola* (56%) and *Gibberella zeae* (26%) (30, 58). These results suggested that the expression profile obtained for each library was representative of the intestinal regeneration stage.

Of the analyzed ESTs, a high percentage showed no significant similarity to entries in the public protein databases. This was expected in view of the small number of genetic studies in holothurians. Most of the ESTs previously reported for holothurians in the NCBI database were detected in our libraries. Thus a significant achievement from our studies is to increase the existing sequence information for this group with 4,522 new sequences. While the database for *H. glaberrima* is far from complete, our results enrich the current dataset from 639 to 5,161 sequences, providing a comprehensive sequence dataset for this group of organisms. These ESTs were grouped in 2,348 contigs and singletons. However, because the sequences were 5′-reads the possibility exists that clones from the same gene might be more or less complete at the 5′-end and fail to assemble into a contig. On the other hand, this number can be underestimated do to the fact that genes with high sequence similarities, such as actins, might be grouped together into one contig.

Additional comparisons of our EST dataset to the one of the sea urchin *S. purpuratus*, whose genome is almost complete at NCBI, showed an almost doubling of the number of sequences with shared similarities (from 1,112 to 1,905). This is to be expected in view of the evolutionary proximity between holothurians and sea urchins (5). However, the shared sequences are only about one-third of the known holothurian ESTs. This number will possibly increase as the clones are fully sequenced and a larger portion of the putative genes is available for interspecies comparisons. Nonetheless, some of our ESTs are part of gene sequences that are either holothurian specific or are associated with the process of intestinal regeneration that occurs in the sea cucumber but not in sea urchins (20, 27).

**EST representation of stage-related events.** This study provided a preliminary overview of the molecular genetics of regenerating and normal intestine for the sea cucumber *H. glaberrima*. Although, we realize that comparisons of cDNA libraries can be prone to error due to the large number of variables associated with their preparation, certain findings...
suggest that the ESTs found in these libraries are representative of genes expressed in the intestinal tissue from which the cDNA libraries were made. First, similar EST composition and specific putative gene homologs were found in two independent rounds of mass-excision and sequencing of the 7dpe cDNA library. Second, the high number of ESTs associated with housekeeping processes such as metabolism and ribosomal proteins was to be expected, due to the energy production and protein synthesis requirements in both normal and regenerating organs (45). These results are in accordance with Aguero and coworkers (3), who found an even larger percentage (17.4%) of ribosomal protein in Trypanosoma cruzi total ESTs, and Chaudhuri and Lieberman (13), who showed a rise in liver ribosome and protein synthesis after partial hepatectomy. The high representation of some of these ESTs in normal compared with the regenerating tissues is also to be expected. Regenerating tissues need to express additional genes associated with this process, thus decreasing the relative expression of the housekeeping genes.

### Table 5. Partial list of the protein domains identified in the H. glaberrima contigs and their cellular function

<table>
<thead>
<tr>
<th>Normal Domain</th>
<th>Normal Cellular Function</th>
<th>Domain</th>
<th>Cellular Function</th>
<th>Domain</th>
<th>Cellular Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FRQ1</strong></td>
<td>signal transduction mechanisms/ cytoskeleton/cell division and chromosome partitioning</td>
<td>SOUL</td>
<td>heme-binding proteins</td>
<td>TR-Fer</td>
<td>iron ion homeostasis</td>
</tr>
<tr>
<td>Euk-ferritin</td>
<td>metal ion binding</td>
<td>SAA</td>
<td>acute-phase response</td>
<td>Band-7-prohibitin</td>
<td>regulation of apoptosis/histone deacetylation/signal transduction</td>
</tr>
<tr>
<td>Epipend</td>
<td>cell-matrix adhesion</td>
<td>EGF-Lam</td>
<td>cell adhesion, growth migration, and differentiation</td>
<td>vWFA</td>
<td>cell adhesion, migration, homing, pattern formation, and signal transduction</td>
</tr>
<tr>
<td>Beta-tubulin</td>
<td>major components of microtubules</td>
<td>A2M</td>
<td>endopeptidase inhibitor activity</td>
<td>ANK</td>
<td>transcription initiation, cell-cycle regulation, cytoskeletal, ion transport and signal transduction</td>
</tr>
<tr>
<td>H3</td>
<td>nucleosome assembly/chromosome organization and biogenesis</td>
<td>GroEL</td>
<td>cellular protein metabolism</td>
<td>PP2Cc</td>
<td>Intracellular signalling</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>ubiquitin conjugating enzyme activity</td>
<td>Sec61-beta</td>
<td>intracellular protein transport</td>
<td>SOUL</td>
<td>heme-binding proteins</td>
</tr>
<tr>
<td>Cyclophilin-ABH</td>
<td>protein folding</td>
<td>Fubi</td>
<td>tumor suppressor protein</td>
<td>COL 1</td>
<td>extracellular matrix structural constituent</td>
</tr>
<tr>
<td>WD40</td>
<td>adaptor/regulatory modules in signal transduction, premRNA processing and cytoskeleton assembly</td>
<td>TR-Fer</td>
<td>iron ion homeostasis</td>
<td>NDK</td>
<td>nucleoside-diphosphate kinase activity, ATP binding</td>
</tr>
<tr>
<td>PCBP-like HK</td>
<td>mRNA stabilization, translational activation, and translational silencing</td>
<td>nucleosmin</td>
<td>nucleic acid binding</td>
<td>WD40</td>
<td>adaptor/regulatory modules in signal transduction, premRNA processing and cytoskeleton assembly</td>
</tr>
<tr>
<td>AhpC</td>
<td>posttranslational modification, protein turnover, chaperones</td>
<td>RPP1A</td>
<td>translation, ribosomal structure and biogenesis</td>
<td>FRQ1</td>
<td>Signal transduction mechanisms/ Cytoskeleton/Cell division and chromosome partitioning</td>
</tr>
<tr>
<td>CT</td>
<td>regulate ligand-receptor interactions for a variety of signaling pathways</td>
<td>Cox2</td>
<td>cytochrome c oxidase activity</td>
<td>TRAP-beta</td>
<td>cotranslational protein targeting to membrane</td>
</tr>
<tr>
<td>Cox2</td>
<td>cytochrome c oxidase activity</td>
<td>Euk-Ferritin</td>
<td>metal ion binding</td>
<td>Sap B</td>
<td>activation of various lysosomal lipid-degrading enzymes cytochrome-c oxidase activity</td>
</tr>
<tr>
<td>Cox4</td>
<td>cytochrome c oxidase activity</td>
<td>RPL30</td>
<td>translation, ribosomal structure and biogenesis</td>
<td>Cox2</td>
<td></td>
</tr>
<tr>
<td>CyoB</td>
<td>energy production and conversion</td>
<td>ATP-synt-A</td>
<td>proton transport</td>
<td>NADH5</td>
<td>ATP synthesis coupled electron transport structural constituent of cytoskeleton</td>
</tr>
<tr>
<td>Rpxl</td>
<td>translation, ribosomal structure and biogenesis</td>
<td>RpsN</td>
<td>translation, ribosomal structure and biogenesis</td>
<td>Actin</td>
<td>translation, ribosomal structure and biogenesis</td>
</tr>
<tr>
<td>Oxired-q2</td>
<td>mitochondrial electron transport, NADH to ubiquinone</td>
<td>RpsK</td>
<td>translation, ribosomal structure and biogenesis</td>
<td>RPP1A</td>
<td>translation, ribosomal structure and biogenesis</td>
</tr>
<tr>
<td>Ribosomal S8</td>
<td>translation, ribosomal structure and biogenesis</td>
<td>Cyt-c-oxidase</td>
<td>mitochondrial electron transport</td>
<td>OSCP</td>
<td>ATP synthesis coupled proton transport translation, ribosomal structure and biogenesis</td>
</tr>
<tr>
<td>Actin</td>
<td>structural constituent of cytoskeleton</td>
<td>actin</td>
<td>structural constituent of cytoskeleton</td>
<td>RAMP4</td>
<td></td>
</tr>
<tr>
<td>Rpl0</td>
<td>translation, ribosomal structure and biogenesis</td>
<td>ribosomal-S8</td>
<td>translation, ribosomal structure and biogenesis</td>
<td>H2A</td>
<td>nucleosome assembly, chromosome organization and biogenesis</td>
</tr>
<tr>
<td>NADH5</td>
<td>ATP synthesis coupled electron transport</td>
<td>NADHdh</td>
<td>electron transport</td>
<td>40S-S3-KH</td>
<td>translation, ribosomal structure and biogenesis</td>
</tr>
</tbody>
</table>

Partial list of the most relevant and abundant proteins domains identified in the 3 cDNA libraries.
sion profile of ribosomal proteins differed among the three cDNA libraries; however, the relative percentage of ribosomal sequences increased with the regenerating stage. This increment in ribosomal proteins during regeneration may be due to the increased need for proteins synthesis during the regeneration process (13).

Third, the analysis by functional categories suggests that changes in gene expression in the regenerating libraries followed expected trends. For example, the percentage of ESTs associated with inflammatory response in the 3dpe library was almost double that of the other libraries. At this stage, wound healing is occurring and the immune response must be activated against pathogens that might have entered during evisceration (19, 52). Similarly, three functional categories increased in the 7dpe dataset: cell proliferation/apoptosis, cell adhesion, and intracellular signaling. The first was in accordance with our published observations that cell division is low at 3dpe but increases at 9dpe (19) and with our unpublished observations that apoptosis occurs mainly at the 7dpe stage. It has been reported that apoptosis may play an important role during regeneration in other organisms (26). The increase in sequences associated with cell adhesion agreed with our observations that extensive cellular migration and extracellular matrix remodeling occurs between 7 and 14dpe (10, 19, 46). Thus, during the 7dpe stage, genes associated with migration, cell adhesion, and extracellular matrix interactions can be expected to be activated.

Finally, previous work in our laboratory has identified specific genes that appeared to be associated with the regeneration process. Some of these genes, such as ependymin (56), collagen (49), and serum amyloid (49, 51), were also identified in our regenerating EST dataset. Moreover, one of the most prominent contigs in the 7dpe cDNA library corresponds to actin. Our laboratory previously identified actin clones in generating intestine by differential display (49), demonstrating upregulation of various actin isoforms and suggesting an important role during regeneration. The identification in this study of genes previously reported by our laboratory supported our contention that our EST dataset represented the ongoing genetic changes taking place at the time the cDNA libraries were made.

To further dissect the sequences represented in our libraries and determine a differential gene profile among them, we identified potential domains in the *H. glaberrima* contigs using RPS-BLAST searches. Because the libraries were not normalized, we found that Cox and ribosomal domains were among the most abundantly identified domains. However, several domains were identified in only one of the libraries but not the others, suggesting that the expression of these domains may be stage specific. Among the most relevant domains present in the 3dpe library were SAA, A2M, and EGF-lam. The SAA domain, which can induce IL-12 and 23 (24), has been reported previously by our laboratory (51). The domains A2M and EGF-lam have been associated with neuronal growth and neurodegenerative diseases, respectively (54, 63). Domains such as 7-prohibitin, PP2c, and NDK were identified only in the 7dpe library. Proteins presenting these domains have been found to be associated with apoptosis and tumorigenesis (17, 57, 64), suggesting that the expression of the corresponding genes might be highly regulated during regeneration.

Table 6. Differential expression of contigs in *H. glaberrima* cDNA libraries

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Clone Name</th>
<th>Statistic R</th>
<th>↑</th>
<th>↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>2986-2</td>
<td>cytochrome c oxidase subunit I [Cucumaria miniata]</td>
<td>25.57</td>
<td>3</td>
<td>N</td>
</tr>
<tr>
<td>4791-1</td>
<td>unknown (4791-1)</td>
<td>15.95</td>
<td>7</td>
<td>N</td>
</tr>
<tr>
<td>2986-3</td>
<td>cytochrome c oxidase subunit I [Cucumaria miniata]</td>
<td>13.85</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>94-1</td>
<td>cytochrome c oxidase subunit III [Cucumaria miniata]</td>
<td>12.77</td>
<td>3</td>
<td>N</td>
</tr>
<tr>
<td>4277-4</td>
<td>unknown (4277-4)</td>
<td>10.29</td>
<td>7</td>
<td>N</td>
</tr>
<tr>
<td>4860-1</td>
<td>unknown (4860-1)</td>
<td>9.93</td>
<td>7</td>
<td>N</td>
</tr>
<tr>
<td>76-2</td>
<td>ferritin GF1 [Crassostrea gigas]</td>
<td>8.73</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>5087-2</td>
<td>toposome [Paracentrotus lividus]</td>
<td>7.77</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4766-1</td>
<td>unknown (4766-1)</td>
<td>6.91</td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>5141-2</td>
<td>unknown (5141-2)</td>
<td>6.67</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>990-7</td>
<td>cytochrome c oxidase subunit II [Cucumaria miniata]</td>
<td>6.46</td>
<td>3</td>
<td>N</td>
</tr>
<tr>
<td>4911-1</td>
<td>unknown (4911-1)</td>
<td>6.36</td>
<td>7</td>
<td>N</td>
</tr>
</tbody>
</table>

Data are presented from highest to lowest statistic R value. The Gene ID corresponds to local identification, the clone name shows the best BLASTX homology with the NCBI GenBank.

Fig. 4. Number of genes in logarithmic scale as a function of R. The solid line represents the regression for 1 ≤ R ≤ 8 in which the slope is −0.85 with SE of 0.09. This is close to the −1 with 5% of error. For R ≥ 8 the number of genes is above this exponential curve and is greater than predicted by the large deviations calculation and thus, with true biological effect.
Fig. 5. Gene expression profile analysis of unknown 4766-1, unknown 4680-1, NADH, centaurin, and melanotransferrin mRNA during intestinal regeneration. Analysis was performed for all samples by conventional RT-PCR and normalized against cytochrome b (Cyt b, 161 bp). A: compared with normal animals, unknown 4766-1 presents a significant overexpression at 3dpe (*P < 0.01, 2-tailed t-test). Centaurin expression presents a significant increase in 7dpe (*P < 0.01) compared with 3dpe and normal animals. Melanotransferrin presents an increase tendency in 3dpe and 7dpe but is not statistically significant. Unknown 4680-1 and NADH expression didn't show a significant change. B: centaurin presents a significant overexpression in stages 7dpe–21dpe, where the peak of expression appears at 14dpe (***P < 0.001, 2-tailed t-test) compared with normal animals. Overexpression of melanotransferrin was detected in 3dpe+7dpe (*P < 0.05, 2-tailed t-test) compared with later stages of regeneration (21+28dpe) and normal animals. Values are presented as means ± SE of 5 animals per group.
regeneration the MMP activity was increased (46). Interestingly, in the sea cucumber 7dpe dataset we identified various ESTs with high similarity to MMPs. These findings support our work and provide the tools for further exploration of the ECM remodeling processes during regeneration. These results provide the first clue to the identification of molecules that might be important in regeneration and strengthen the use of H. glaberrima as a model to study repair and regeneration of the gastrointestinal tract.

In silico differential display of the regenerating and normal H. glaberrima contigs. To identify putative genes associated with the regeneration process, an in silico differential display was performed. This method provided a more rigorous analysis of the genetic profiles of the libraries. We found that many contigs identified as the most common in the cDNA libraries were also identified by the R analysis to be differentially expressed. The R statistic showed 12 sequences to be differentially expressed at significant levels (R > 6). Some of these were overexpressed in the normal intestine, including cytochrome c oxidase subunits and ferritin. Since these proteins are known to be highly expressed in the mucosal layer of the digestive tract (14), their differential expression is probably due to regular occurrence in the luminal epithelium of normal animals, a tissue layer that would be absent in the regenerating animals.

Some of the most interesting contigs are those found to be differentially expressed by R analysis and that showed no similarity to sequences in the NCBI database. One of these, contig #5141-2 was confirmed by ESTs only found in the normal intestine library. This putative gene may be expressed by a cell type present in normal nonviscerated intestine but absent in regenerating tissues. Alternatively, it may belong to a putative gene that is underexpressed during the regeneration process. In view of our interest in finding novel genes associated with regeneration, the best candidates are those unknown contigs that were found mostly in the 3dpe or in the 7dpe libraries. Contig #5020-1 was among the most common found in the 7dpe library with a large R value (> 5.4), making it a possible putative gene involved in regeneration at 7dpe stage. Two unknowns (#4791-1 and #4860-1) were found mainly in the 3dpe library. However, the predicted overexpression of 4860-1 could not be verified by PCR analyses. The lack of significance in the expression of the unknown 4860-1 during regeneration may be due to the high variability among the animals or to limits in the sensitivity of the conventional RT-PCR. Finally, two other unknowns (#4277-4 and #4766-1) were also present in the 7dpe and normal libraries but were overrepresented in the 3dpe library. One of them (#4766-1) is of particularly interesting, first because it shows an EF-hand calcium binding domain, which have been recognized as a key player in all aspects of cell function, from mitosis to apoptotic death (22). Second, because of the increase in expression predicted by R for this clone was verified by the PCR analyses showing the clone to be significantly higher in the 3dpe than in the 7dpe and normal intestines.

Our PCR analysis also showed further evidence of differential expression that support our EST results. Clones whose expression was expected to remain stable during the regeneration process, such as NADH and cytochrome b, behaved as expected. In addition, differences in expression profiles were found with two sequences that were identified only in the regenerating libraries and that are of interest for future studies. Melanotransferrin ESTs were found only in the 3dpe and 7dpe libraries. This sequence is similar to a gene associated with angiogenesis (53), eosinophil differentiation (38), and tumor progression (15). Interestingly, the expression analysis shows that this sequence is overexpressed in the 3dpe and 7dpe stages of regeneration, supporting our ESTs results. Similarly, centaurin, a recently described gene associated mainly with cell signaling control (31), actin reorganization (60), Alzheimer’s disease (47, 48), and downregulation of NF-kB (64), was found only in the 7dpe library. PCR results showed that centaurin is overexpressed at the 7dpe stage compared with normal and 3dpe, which is in accordance with our EST analysis. Moreover, the temporal expression of centaurin peaking at 14dpe and returning to levels similar to the normal animals at 28dpe clearly suggests that this gene might play an important role during intestinal regeneration.

Development, homeostasis, and repair are complex processes that require precise orchestration of cell proliferation, migration, differentiation, and death. All of these events take place during intestinal regeneration. Our EST project not only has yielded new insights into the molecular basis of gastrointestinal regeneration but provides evidence for the existence of a differential expression profile during regeneration. Future studies will focus on the characterization of those sequences that were found among the most frequent in the libraries and that lack similarity to known sequences, in the hope that they will provide important information to the process of intestinal tract repair and regeneration. The analyses of regenerating and normal intestine cDNA libraries of H. glaberrima have served to increase the annotation of genes in these organisms and underscored the utility of H. glaberrima as a model system for the study of repair and regeneration of the digestive tract.

ACKNOWLEDGMENTS

The authors thank Dr. Maria E. Perez for invaluable help in the statistical analysis. In addition we thank Dr. Sheila Ward for editorial comments.

GRANTS

This study was supported by National Institutes of Health (NIH)-Minority Biomedical Research Support (S06GM-08102). Partial support was also provided by DEI-PBDD (doctoral dissertation fellowship to E. C. Suarez-Castillo), NSF-Integrative Biology and Neuroscience (0110692), NIH-Research Centers in Minority Institutions (RRO-3641-01), and the University of Puerto Rico. This project made use of the computational resources at the High Performance Computing Facility of the University of Puerto Rico supported by
REFERENCES


