The Schmidtea mediterranea database as a molecular resource for studying 
platyhelminthes, stem cells and regeneration

Alejandro Sánchez Alvarado*†§, Phillip A. Newmark*‡, Sofia M. C. Robb† and Réjeanne Juste

Carnegie Institution of Washington, Department of Embryology, Baltimore, MD 21210, USA
*These authors contributed equally to this work
†Present address: University of Utah School of Medicine, Department of Neurobiology and Anatomy, Salt Lake City, UT 84132-3401, USA
‡Present address: University of Illinois at Urbana-Champaign, Department of Cell and Structural Biology, Urbana-Champaign, IL 61801, USA
§Author for correspondence (e-mail: sanchez@neuro.utah.edu)

Accepted 11 August 2002

SUMMARY

Platyhelminthes are excellent models for the study of stem cell biology, regeneration and the regulation of scale and proportion. In addition, parasitic forms infect millions of people worldwide. Therefore, it is puzzling that they remain relatively unexplored at the molecular level. We present the characterization of ~3000 non-redundant cDNAs from a clonal line of the planarian Schmidtea mediterranea. The obtained cDNA sequences, homology comparisons and high-throughput whole-mount in situ hybridization data form part of the S. mediterranea database (SmedDb; http://planaria.neuro.utah.edu). Sixty-nine percent of the cDNAs analyzed share similarities with sequences deposited in GenBank and dbEST. The remaining gene transcripts failed to match sequences in other organisms, even though a large number of these (~80%) contained putative open reading frames. Taken together, the molecular resources presented in this study, along with the ability of abrogating gene expression in planarians using RNA interference technology, pave the way for a systematic study of the remarkable biological properties displayed by Platyhelminthes.

Key words: Regeneration, Stem cells, Plasticity, Database

INTRODUCTION

The phylum Platyhelminthes (flatworms) consists of approximately 50,000 different species that populate a remarkable variety of niches (Littlewod and Bray, 2001). In addition to free-living forms, it encompasses parasitic organisms responsible for inflicting debilitating diseases upon hundreds of millions of people throughout the world (see World Health Organization fact sheet 115 at http://www.who.int/inf-fs/en/fact115.html). Platyhelminthes are considered by many to occupy an important position in the evolution of the Metazoa (Adoutte et al., 1999; Henry et al., 2000; Tyler, 2001; Willmer, 1994), and the panoply of developmental properties displayed by these organisms has attracted the attention of generations of biologists (Newmark and Sánchez Alvarado, 2002). For example, the ability of freshwater planarians to regenerate completely from small body fragments has been known for over two centuries (Morgan, 1898; Randolph, 1897), and the life cycles of some digenetic trematodes involve as many as three different hosts as well as both sexual and asexual strategies for their reproduction (Brusca and Brusca, 1990; Hyman, 1951). Yet, as important, abundant and diverse as platyhelminthes are, little is known about the molecular events that guide their sophisticated and often plastic biological properties.

Moreover, many members of this phylum possess large populations of undifferentiated mesenchymal stem cells, the study of which could contribute significantly to fundamental biomedical research in the areas of tissue regeneration, stem cell maintenance and degenerative disorders. In most free-living species these stem cells, which are often referred to as neoblasts, are used for the regeneration of missing body parts and/or the replacement of cells that are lost during the course of physiological turnover (Gschwentner et al., 2001; Ladurner et al., 2000; Newmark and Sánchez Alvarado, 2000). Similarly, free mesenchymal cells in parasitic flukes are known to produce complete larval forms (Brusca and Brusca, 1990; Hyman, 1951), and in the cestode Taenia crassiceps complete cysts can be reconstituted from individual cells (Toledo et al., 1997). Thus, platyhelminthes also provide a unique opportunity for studying the mechanisms that underlie the control of cellular pluripotentiality.

To address many of these unsolved problems, we and others (Agata and Watanabe, 1999) have chosen to reintroduce the freshwater planarian as an experimental model. We report the establishment of a clonal line of a diploid, asexual form of the planarian Schmidtea mediterranea (Turbellaria, Tricladida), along with the isolation and sequence characterization of ~3000 non-redundant, expressed sequence tags (ESTs) from this organism. Furthermore, we show the suitability of using planarians for high-throughput mapping of gene expression patterns in the whole animal, and introduce the S.
GenBank, SmedDb has been programmed to update the BLAST number of new sequences being continuously deposited into BLASTn or translated (BLASTx) searches. In addition, dbEST comparisons using BLASTcl3 running either nucleotide-nucleotide management and internet browser accessibility of the data, unique proportion of the complexity of the libraries (50-55%; see that the non-redundant clones identified represent a significant analysis of the frequency distribution of unique sequences indicates internal redundancy and to identify unique clones. Statistical obtained sequences were compared against one another using DNA sequencer. The sequencing strategy is outlined in Fig. 1A.chemistry and the resulting products were run on an ABI Prism 377 Sequencing reactions were performed using Big Dye Terminator Isolation were performed using a MiniPrep24 machine (MacConnell RI and XhoI restriction enzymes were used for the initial cloning of the cDNAs. The plasmids were then digested with HindIII and EcoRI and cloned into the pBluescript II SK (+) expression vector. The resulting plasmids were then transformed into DH10B cells. Some of the transformed clones were used for sequencing, while others were used for the expression of the cDNA inserts in bacteria. The expression of the cDNA inserts was then verified by SDS-PAGE analysis.

RESULTS

The S. mediterranea database

Two tissue-specific cDNA libraries were generated, one from the head and another from the body. The head library contained 564 cDNAs, while the body library contained 542 cDNAs. The cDNAs were sequenced and analyzed using the following bioinformatics tools: BLAST, ClustalW, and Gene Ontology. The results showed that the cDNAs from the head library were enriched in genes involved in the immune response, while the cDNAs from the body library were enriched in genes involved in growth and development. The cDNA sequences were then uploaded to the SmedDb database, which is accessible through a user-friendly web interface.

Whole-mount in situ hybridization

A total of 500-600 planarians were used for in situ hybridization experiments. The planarians were fixed and embedded in paraffin, and then sectioned and hybridized with a digoxigenin-labeled cDNA probe specific for a gene of interest. The sections were then stained with an antibody against digoxigenin, which binds to the cDNA probe and allows for visualization of the gene expression pattern. The results showed that the gene of interest was expressed in a specific cell type, indicating that the gene is involved in a specific biological process.

GenBank accession numbers

The GenBank accession numbers for the cDNA sequences are as follows: AY066058-AY066070; AY066071-AY066080; AY066081-AY066090; and AY066091-AY066100.
corresponding BLAST results linked to Entrez-PubMed (see http://planaria.neuro.utah.edu). Examples of SmedDb entries placed into functional categories are shown in Table 1. At least 77 transcription factors, 130 DNA replication/modification molecules and 97 receptors, channels and other membrane-associated proteins were putatively identified.

Interestingly, when the planarian ESTs with significant homologies to GenBank are ranked by lowest expectancy value, we find that 64% of the entries in SmedDb have highest overall similarities to vertebrate rather than to invertebrate sequences (Fig. 1C). When comparative BLASTx analyses between SmedDb and the proteomes of C. elegans, D. melanogaster and H. sapiens were performed, a set of 124 S. mediterranea ESTs with significant similarity only to proteins found in the human genome were revealed. Sixty-three of these are similar to human genes encoding proteins of unknown function. Noteworthy is the presence in S. mediterranea of thymidine phosphorylase/endothelial cell growth factor 1 (BLASTx E=5×10–30), acyl-CoA dehydrogenase (BLASTx E=2×10–21), epoxide hydrolase (BLASTx E=5×10–29) and formiminotransferase cyclodeaminase (BLASTx E=4×10–42). These genes were recently postulated to be present in the human genome as a result of direct horizontal gene transfer (HGT) between bacteria and vertebrates based on their absence in the genomes of C. elegans and D. melanogaster (Lander et al., 2001). However, the presence of these transcripts in planarians suggests that these loci are most probably not shared by bacteria and vertebrates via HGT, but rather by descent through common ancestry (Kyrpides and Olsen, 1999; Stanhope et al., 2001).

High-throughput in situ hybridization
The ~3000 independent ESTs available in SmedDb provide a wealth of material for studying the flatworms. One such use will be for identifying cell type- and region-specific markers. Thus, we have used whole-mount in situ hybridization to begin to determine the spatial expression patterns of SmedDb entries; to date, results from nearly 300 clones have been deposited in SmedDb, and more are being added regularly as they become available. The analysis has revealed some surprising complexities in the spatial expression patterns of many of the genes represented in the EST collection (Fig. 2). We find, for example, that the morphologically simple cephalic ganglia of flatworms display a diverse array of expression domains, some of which are depicted in Fig. 2A (see figure legend for explanation). In addition, other organ-system-specific genes have been identified that label the gastrovascular system, the dorsal epithelium, the excretory system and the pharynx (Fig. 2B from top to bottom). We also find transcripts expressed in various subsets of cells, including the planarian neoblasts in which piwi, a transcript found in many metazoan stem cells (Benfey, 1999), can be detected (Fig. 2C, bottom picture). Striking expression patterns defining both dorsal and ventral boundaries have been observed as well. This is illustrated by the lateral view of in situ hybridization results using clone H.8.1f, which has no known homolog in the available databases (Fig. 2D).

Cell loss during de-growth
The identification of cell type-specific markers from the large-scale in situ hybridization screen provides new tools for

---

**Fig. 1.** Bioinformatics and categorization of S. mediterranea sequences. (A) Organigram depicting the steps performed to produce a non-redundant collection of S. mediterranea cDNA sequences (SmedDb). Bioinformatic analyses compared the database against itself to identify and remove redundant clones before sending sequences to the public databases. The GenBank protein and nucleotide collections as well as the EST database (dbEST) were queried using the BLAST algorithm. Sequences returning significant matches (E<10–4) were subjected to annotation into functional categories based on the identity/function of the GenBank match. (B) Distribution of informative sequences by functional categories. For simplicity the cell signaling category includes the cell/cell communication and internal signaling categories. Metabolism is an amalgamation of the general metabolism, mitochondria and protein metabolism categories. Visit http://planaria.neuro.utah.edu for a detailed categorization profile and access to sequences. (C) Distribution by percentage of planarian genes displaying highest similarities with members of either the vertebrates or invertebrates.
Fig. 2. Representative results of high-throughput, whole-mount in situ hybridization. (A-C) Probe identity corresponding to the images is given from top to bottom with each clone ID in parentheses as follows. (A) Gene expression patterns within the nervous system of *S. mediterranea*: synaptotagmin (H.6.7h); quinoid dihydropteridine reductase (H.9.5b); pax6 (H.109.7h) and degenerin (H.112.3c). (B) Gene expression patterns in organ systems: gastrovascular system (D.14; unknown function), dorsal epithelium (H.7.1e; gp25L/p24 family), excretory system (H.14.9d; carbonic anhydrase); and pharynx (H.14.11f; unknown function). (C) Gene expression in discrete cell types: matrix metalloproteinase (A115) in central secretory cells; epithelial cells (H.12.11a; intermediate filament); subepidermal marginal adhesive gland cells (H.1.3b; zonadhesin); and free-mesenchymal cells (neoblasts) expressing piwi (H.2.12c). (D) Ventral view (left) of clone H.8.1f (unknown function), and lateral view of the same specimen (right) demonstrate a dorsoventral segregation of differentiation. The red asterisk indicates the position of the pigmented photoreceptor. Anterior is to the left in all panels.

Table 1. Examples of *S. mediterranea* sequences placed into functional categories in SmedDb

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Clone ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA metabolism</td>
<td>Transcription factors</td>
<td>H.119.4D</td>
<td>Class IV POU-homeodomain protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.17.9E</td>
<td>Smad4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.38.3(T3)</td>
<td>LIM/homeobox protein LIM (HRLIM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.8.6C</td>
<td>Homeobox protein DTH-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.110.1c</td>
<td>Pre B-cell leukemia transcription factor 2 (Pbx-2)</td>
</tr>
<tr>
<td>DNA replication/modification</td>
<td>Chromosome/nuclear structure</td>
<td>H.90.1e(T3)</td>
<td>Sirtuin 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HB.19.8F</td>
<td>Histone acetyltransferase MORF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.25.11e(T3)</td>
<td>Maleless gene product</td>
</tr>
<tr>
<td></td>
<td>Apoptosis</td>
<td>H.105.11H</td>
<td>Apoptosis inhibitor 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.8.7G</td>
<td>Caspase 6 precursor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.57.1d</td>
<td>Probable Bax inhibitor 1</td>
</tr>
<tr>
<td>Cell-cell communication</td>
<td>Receptors</td>
<td>E-99</td>
<td>FGF homologous factor receptor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.103.12E</td>
<td>Cysteine-rich fibroblast growth factor receptor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.111.10F</td>
<td>GABA&lt;sub&gt;A&lt;/sub&gt; receptor-associated protein</td>
</tr>
<tr>
<td></td>
<td>Other membrane proteins</td>
<td>H.110.2E</td>
<td>Mechanosensory protein 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.119.4E</td>
<td>Endothelin converting enzyme</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.44.6a(T3)</td>
<td>Serrate 2</td>
</tr>
<tr>
<td>Intracellular signaling</td>
<td>Channels/transporters</td>
<td>H.102.1B</td>
<td>Delayed rectifier K&lt;sup&gt;+&lt;/sup&gt; channel</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.90.5b(T3)</td>
<td>Cu&lt;sup&gt;2+&lt;/sup&gt; transporter (Menkes disease-associated protein)</td>
</tr>
<tr>
<td></td>
<td>Transduction</td>
<td>H.16.11G</td>
<td>Na&lt;sup&gt;+&lt;/sup&gt;/K&lt;sup&gt;+&lt;/sup&gt;-ATPase α-subunit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.2.10h(T3)</td>
<td>Rab GDP-dissociation inhibitor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.31.11B</td>
<td>GTP-binding regulatory protein G&lt;sub&gt;α&lt;/sub&gt; chain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.56.3A</td>
<td>cAMP-dependent protein kinase catalytic subunit</td>
</tr>
</tbody>
</table>

Complete lists for each category and their respective subcategories can be found and searched in SmedDb (http://planaria.neuro.utah.edu).

POU, Pit, Oct, Unc DNA-binding domain; Smad, similar to mothers-against decapentaplegic; LIM, Lim11, Isl1, Mec3 protein-binding domain; DTH, Dugesia tigrina homeobox; Pbx, postbithorax; MORF, monocytic leukemia zinc finger protein-related factor; Bax, Bcl2 associated X gene; FGF, fibroblast growth factor; GABA, gamma amino-butryic acid; GDP, guanosine diphosphate; GTP, guanosine triphosphate; cAMP, cyclic adenosine monophosphate.
studying morphallaxis, a classic problem first defined by Morgan in 1898 (Morgan, 1898). Morphallaxis refers to the remodeling that occurs when small fragments of planarians (or other organisms, like Hydra) restore their appropriate proportion and pattern without adding additional tissue. In addition to this remodeling during regenerative events, planarians show a high degree of plasticity in their ability to either grow or de-grow, depending upon environmental conditions. During periods of prolonged starvation, planarians will shrink (Lillie, 1900; Schultz, 1904; Berninger, 1911; Child, 1911; Abeloos, 1930): a 20 mm long worm can be reduced to less than 1 mm over the course of several months. This change in body size is due to an overall reduction in total cell number, as opposed to a reduction in cell size (Baguñà and Romero, 1981; Romero and Baguñà, 1991). Previous studies of this phenomenon have used techniques in which planarians are macerated into a suspension of individual cells. Using this method, roughly 13 different cell types from organisms in varying stages of growth and de-growth were classified and quantitated (Baguñà and Romero, 1981; Romero and Baguñà, 1991). Because the flatworms were dissociated into single cells in these studies, the distribution of the cells could not be monitored in the whole animal as it changed in size. Furthermore, the morphological criteria alone underestimated the true number of different cell types in the planarian.

cDNA clone H112.3c shows weak sequence similarity to degenerin 1 from C. elegans and is expressed in a subset of cells near the anterior margin of the planarian (Fig. 3A); these cells are likely to be involved in chemoreception through ciliated pits that lie at the ciliated anterior margin in this genus (Farnesi and Tei, 1980). The number of H112.3c-expressing cells can be counted easily in organisms of different sizes after whole-mount in situ hybridization. Remarkably, the number of these cells increases linearly with length (Fig. 3B), suggesting that even for cell types comprising a small percentage of the body (~0.03%), their total numbers are regulated as the animal grows and shrinks. How these organisms can ‘count’ different cell types relative to total body size remains a complete mystery.

**DISCUSSION**

Considering that flatworms comprise the fourth largest phylum on Earth (Brusca and Brusca, 1990) and that many of its members have challenged scores of biologists and biomedical researchers, it is puzzling that the molecular biology of the Platyhelminthes has remained largely unexplored. The problems of regeneration, de-growth and proportion regulation remain as puzzling today as they were over 200 years ago. Furthermore, diseases such as Shistosomiasis, which is caused by members of this phylum, continue to be global public health problems with no signs of abating. Thus, deciphering the molecular bases underlying the biology of these organisms should not only improve our knowledge of the phylum, but also contribute to the fields of developmental biology and biomedicine.

The establishment of a clonal line (CIW4) of the freshwater planarian S. mediterranea and the identification of nearly 3000 non-redundant cDNAs from this line will aid the molecular study of the most salient biological properties of this taxon.

Nearly 70% of all S. mediterranea clones share significant homologies to sequences deposited in GenBank (Fig. 1B), and a large number of these have highest similarity to the deuterostome branch of the metazoans (Fig. 1C). These results indicate either a closer proximity of the phylum to the deuterostome lineage as recently proposed by Tyler (Tyler, 2001), or are more likely a reflection of the poor representation of invertebrate sequences in current databases. The latter possibility is illustrated by the identification in planarians of cDNAs encoding proteins that until recently were ascribed to be present only in bacteria and vertebrates based on a comparative analysis of the human, fly and nematode genomes (Lander et al., 2001). The presence of Thymidine phosphorylase/endothelial cell growth factor 1, acyl-CoA dehydrogenase, epoxide hydrolase and formimidotransferase cycloeaminase in S. mediterranea suggests that these loci reached the vertebrates by common ancestry and not by horizontal gene transfer as originally proposed (Lander et al., 2001). Therefore, even though the proteomes of both C. elegans and D. melanogaster have been deposited in GenBank, limiting sequence comparisons to these two invertebrates is not sufficient to draw sound phylogenetic conclusions, especially on the basis of BLAST results alone. Only rigorous
phylogenetic analyses can most closely approximate phyletic relationships and we expect that the sequences in SmedDb will contribute to the production of higher resolution intra- and inter-phyletic metazoan relationships.

In addition to sharing a large number of genes with the human, fly and nematode genomes, it should be noted that several planarian cDNAs with significant similarities to human sequences were not identified in the C. elegans or D. melanogaster genomes by BLAST searches. At least 63 of these cDNAs are similar to human genes encoding proteins of unknown function. Therefore, S. mediterranea is likely to expand and complement the repertoire of organisms used for the study of genes and pathways involved in various aspects of human biology and disease.

The high-throughput in situ hybridization analyses reported here will serve as a first step in deciphering the roles of genes encoding proteins of unknown function. The tissue- or cell type-specific expression patterns of these genes may provide hints as to their function. For example, cDNA clones H14.5b and H12.6g share similarity with human genes for which no function is known (hypothetical protein XP_044953.1: E=5e–9 and unnamed protein product AK022687; E=1e–12, respectively), and are expressed in neurons of both the planarian central and peripheral nervous system (see http://planaria.neuro.utah.edu). Our previous demonstration that double-stranded RNA can be used to inhibit gene expression in planarians (Sánchez Alvarado and Newmark, 1999) provides the means for testing gene function on a large scale, thus allowing the functional characterization of novel, evolutionarily conserved gene products.

Furthermore, cell type-specific markers identified by large-scale in situ screens provide useful reagents for examining the processes of patterning, differentiation and remodeling in intact and regenerating planarians. We have shown the use of such a marker (H112.3c) to quantify cell number changes as planarians alter their size, and found that these animals also regulate accordingly the numbers of a specific cell type (Fig. 3). This maintenance of pattern and proportion is a fascinating corollary to the regenerative abilities displayed by these organisms. In addition, little is known about the heterogeneity of the stem cell population in planarians and markers such as piwi (H2.12c) will provide necessary reagents for analyzing the processes by which neoblasts differentiate to give rise to the ~30 cell types in the animal. The tools described make these daunting problems more amenable to molecular dissection.

Finally, BLASTn and BLASTx queries also revealed that ~31% of the cDNAs obtained do not share sequence similarities with the available databases. This lack of similarities with GenBank and dbEST is not due to the divergences commonly found in untranslated sequences, because only ~20% of these cDNAs lack a putative ORF. These results suggest that some of these sequences may correspond to Platyhelminth-specific genes. Therefore, in addition to its obvious advantages for studying the problem of regeneration, the easily manipulable planarian provides a free-living counterpart likely to complement current research efforts on the parasitic forms, in particular Schistosoma mansoni and S. japonicum, for which abundant sequence data are being obtained (Snyder et al., 2001). Given that the parasitic flatworms are difficult experimental subjects, the ability to identify flatworm-specific genes through comparisons to S. mediterranea sequences should help identify candidate molecules for therapeutic intervention. Furthermore, the in situ hybridization data being generated in S. mediterranea will help identify genes expressed in cell types unique to the platyhelminthes, providing additional potential therapeutic targets. The combination of sequence comparisons, gene expression patterns, and RNAi technology provide new experimental possibilities for studying the free-living and parasitic members of this phylum. Thus, the SmedDb resources will be useful to a wide gamut of developmental and biomedical endeavors.

We thank our many colleagues at the Carnegie Institution, Department of Embryology for their support and encouragement during the early phases of this project. We are also indebted to Ms Allison Pinder at the Carnegie’s sequencing core for overseeing the initial stages and training of personnel for automated sequencing procedures; and to Ms Jennie Bentz (bentz_jennie@yahoo.com) for the design of the initial version of our web page and for suggesting the use of Cold Fusion to manage SmedDb. Our thanks are also extended to Mr Nestor Oviedo for in situ work. P. A. N. acknowledges the Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation for supporting his study of planarian biology. This work was also supported by NIH NIGMS RO-1 GM57260 to A. S. A. and NIH F32-GM19539 to P. A. N.

REFERENCES


