nature genetics

Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*

Sergio Verjovski-Almeida[1], Ricardo DeMarco[1], Elizabeth A L Martins[2], Pedro E M Guimarães[3], Elida P B Ojopi[3], Apuã C M Paquola[4], João P Piazza[5], Milton Y Nishiyama Jr.[4], João P Kitajima[5,15], Rachel E Adamson[6], Peter D Ashton[6], Maria F Bonaldo[7], Patricia S Coulson[6], Gary P Dillon[6], Leonardo P Farias[2], Sheila P Gregorio[1,3], Paulo L Ho[2], Ricardo A Leite[8], L Cosme C Malaquias[9], Regina C P Marques[8], Patricia A Miyasato[10], Ana L T O Nascimento[2], Fernanda P Ohlweiler[10], Eduardo M Reis[1,4], Marcela A Ribeiro[11], Renata G Sá[12], Gaëlle C Stukart[3], M Bento Soares[7,13], Cybele Gargioni[14], Toshie Kawano[10], Vanderlei Rodrigues[12], Alda M B N Madeira[11], R Alan Wilson[6], Carlos F M Menck[8], João C Setubal[5], Luciana C C Leite[2] & Emmanuel Dias-Neto[3]

***Schistosoma mansoni* is the primary causative agent of schistosomiasis, which affects 200 million individuals in 74 countries. We generated 163,000 expressed-sequence tags (ESTs) from normalized cDNA libraries from six selected developmental stages of the parasite, resulting in 31,000 assembled sequences and 92% sampling of an estimated 14,000 gene complement. By analyzing automated Gene Ontology assignments, we provide a detailed view of important *S. mansoni* biological systems, including characterization of metazoa-specific and eukarya-conserved genes. Phylogenetic analysis suggests an early divergence from other metazoa. The data set provides insights into the molecular mechanisms of tissue organization, development, signaling, sexual dimorphism, host interactions and immune evasion and identifies novel proteins to be investigated as vaccine candidates and potential drug targets.**

Schistosomiasis is a public health problem in many developing countries, and *Schistosoma mansoni* is the most widespread species of the causative trematode parasite[1]. Parasite eggs laid in the hepatic portal vasculature are the principal cause of morbidity, and the ensuing pathology may prove fatal[2]. Control of the disease by chemotherapy has relied heavily on praziquantel, potentially allowing drug-resistant parasites to emerge[3]. Protective immune mechanisms in humans that might form the basis for a vaccine have proven difficult to characterize[4] owing to effective immune evasion by parasites. Nevertheless, the successful vaccination of both rodents and primates with attenuated larvae[5] indicates that the goal is feasible.

As representatives of the platyhelminths, schistosomes are the lowest group of bilateria that diverged early from the metazoan lineage[6]. With a blind-ending gut and no body cavity, their body plan seems simple, but tissues corresponding to the main organ systems of higher animals are present. Schistosomes have a complex life cycle, and they are among the first animals to develop sexual dimorphism and heteromorphic sex chromosomes. They are intimately associated with the gastropod mollusk intermediate and the mammalian final host, perhaps relying on host signals for development. Active transmission between hosts and internal migrations show their capacity for sophisticated neuromuscular coordination.

The large size (270 Mb; ref. 7) and complexity of the *S. mansoni* genome have previously deterred full-scale sequencing (see The Institute for Genomic Research and The Sanger Institute websites). Current knowledge of expressed genes is limited to a set of 163 full-length cDNAs and approximately 16,000 ESTs, 75% derived from adult worms[8,9]. We report here a multicenter effort to obtain and

[1]Departamento de Bioquimica, Instituto de Quimica, Universidade de São Paulo, 05508-900 São Paulo, SP, Brazil. [2]Centro de Biotecnologia, Instituto Butantan, 05503-900 São Paulo, SP, Brazil. [3]Laboratory of Neurosciences (LIM27), Instituto de Psiquiatria, HCFM, Universidade de São Paulo, 05403-010 São Paulo, SP, Brazil. [4]Laboratorio de Bioinformatica, Departmento de Bioquimica, Instituto de Quimica, Universidade de São Paulo, 05508-900 São Paulo, SP, Brazil. [5]Instituto de Computacao, C.P. 6176, Universidade Estadual de Campinas, 13084-971 Campinas, SP, Brazil. [6]Department of Biology, University of York, P.O. Box 373, York YO10 5YW, UK. [7]Department of Pediatrics, University of Iowa, Iowa City, Iowa 52242, USA. [8]Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Av. Prof. Lineu Prestes, 1374, 05508-900, São Paulo, SP, Brazil. [9]Universidade Vale do Rio Doce, 35030-390 Governador Valadares, MG, Brazil. [10]Laboratório de Parasitologia, Instituto Butantan, Av. Vital Brasil, 1500, 05503-900, São Paulo, SP, Brazil. [11]Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, 05508-900 São Paulo, SP, Brazil. [12]Departamento de Bioquímica e Imunologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, 14049-900 Ribeirão Preto, SP, Brazil. [13]Departments of Biochemistry, Orthopaedics, Physiology and Biophysics, University of Iowa, Iowa City, Iowa 52242, USA. [14]Departamento de Parasitologia, Instituto Adolfo Lutz, Av. Dr. Arnaldo nº 351, 8º andar, 01246-902 São Paulo, SP, Brazil. [15]Present address: Alellyx Applied Genomics, TechnoPark, rod. Anhanguera km 104, 13067-850 Campinas, SP, Brazil. Correspondence should be addressed to S.V.-A. (verjo@iq.usp.br).

Published online 14 September 2003; doi:10.1038/ng1237

Table 1 ***S. mansoni*** **transcriptome features and gene complement**

	Number of reads	
Total sequenced reads[a]		163,586
Total analyzed reads[b]		124,640
Adults[c]	33,180	
Eggs	19,077	
Miracidia	18,638	
Germ balls	16,715	
Cercariae	10,014	
Cultured day-7 schistosomula	27,016	
Average EST size (bp after trimming)		385.4
Total number of SmAE sequences		30,988
Number of contigs	12,322	
Number of orphan sequences (singlets)	18,666	
Average contig size (bp)		505
Total SmAEs matching known *S. mansoni* sequences		7,086 (23%)
Match to *S. mansoni* known genes from GenBank	639 (2%)	
Match to *S. mansoni* known ESTs from dbEST	6,447 (21%)	
Total SmAEs with putative new *S. mansoni* gene fragments		23,902 (77%)
Match to *S. mansoni* known proteins (new paralogs)	449 (1%)	
Match to genes of other organisms (new orthologs)	6,274 (20%)	
No-match in GenBank (fragments with unknown function)	17,179 (55%)	
Estimated gene complement[d]		13,960–14,205

[a]All libraries were closely monitored for redundancy; sequencing was halted when redundancy reached 50% in a given library. Sequenced reads are those with more than 100 bp with a Phred value higher than 15. [b]The analyzed data set excludes 15,226 reads that are putative contaminants derived from bacteria, mouse or human and another 26,702 reads of repetitive DNA sequences, mitochondrial, transposon or ribosomal origin. [c]21,605 reads from ORESTES libraries and 11,575 from a poly-dT-primed normalized adult worm library. [d]Two different methods were used to estimate the number of *S. mansoni* genes (see **Supplementary Methods** online).

annotate extensive transcriptome data for *S. mansoni*, using both a normalized cDNA library[10] from adults and ORESTES minilibraries from six life-cycle stages (**Supplementary Fig. 1** online). This approach, based on arbitrary primers and low-stringency RT–PCR[11], preferentially amplifies the central, function-defining coding regions of messages[12]. This first large-scale database for a bilaterian acoelomate should enhance our understanding of the evolution, biology and adaptation to parasitism of these animals and identify novel proteins to be exploited as drug targets and vaccine candidates.

Results

Transcriptome features and gene complement

We obtained 163,586 EST reads from the *S. mansoni* transcriptome: 151,684 using ORESTES minilibraries and 11,902 from a normalized adult worm library. All our results are from a filtered data set of 124,681 analyzed reads, which resulted in 30,988 assembled EST sequences (**Table 1**), called *Schistosoma mansoni* assembled EST sequences (SmAEs). Newly identified *S. mansoni* genes are listed by product in **Supplementary Table 1** online. The SmAE data set is estimated to sample 92% of the *S. mansoni* transcriptome. Comparison of SmAEs with publicly available sequences shows that 77% represent new *S. mansoni* gene fragments, either novel paralogs (1%), new orthologs (20%) or fragments with unknown function (no match in GenBank; 55%; **Table 1**). An average SmAE sequence provides around 32% coverage of a matching gene in GenBank (**Supplementary Fig. 2** online); nevertheless, 359 novel orthologs have their entire coding region fully sequenced (**Supplementary Table 2** online).

The total number of genes in the parasite was predicted by two different methods to be around 14,000 (**Table 1**), comparable to the 14,000–19,000 predicted genes of other fully sequenced invertebrates[13–15]. Extrapolation from nonredundant bases acquired from adult worm ESTs indicates that 7,200 genes are expressed in this stage (**Supplementary Fig. 3** online). We obtained 58,846 tags from serial analysis of gene expression (SAGE), and the number of unique tags reached a clear plateau at 6,263 (**Supplementary Fig. 3** online), suggesting that almost all adult transcripts were sampled. Thus, about 50% of all *S. mansoni* genes are expressed in adult worms.

Functional classification of transcripts

We assigned Gene Ontology classifications to 8,001 SmAEs (Gene Ontology browser is available at the project website). The distribution of SmAEs among the main categories is shown in **Supplementary Table 3** online. Protein metabolism was the most frequently identified

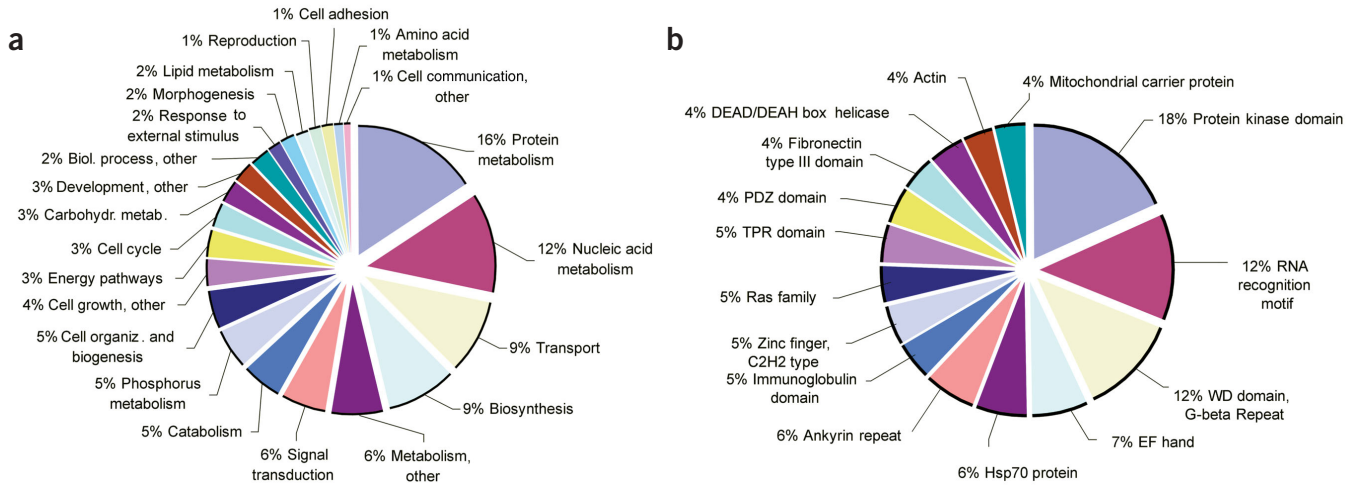

Figure 1 Gene Ontology classification and frequently encountered Pfam domains in SmAEs. (**a**) Percentage of *S. mansoni* SmAEs in each of the biological process categories of Gene Ontology classification. A total of 5,463 distinct SmAEs were assigned to 9,497 different biological processes (individual SmAEs can have multiple Gene Ontology assignments). (**b**) Fifteen Pfam domains occurred most frequently in *S. mansoni* SmAEs. Multiple Pfam domains on the same SmAE were counted only once.

of the biological process categories (**Fig. 1a**). Searching for conserved domains (in the Pfam database) showed that protein kinases were the most abundant (**Fig. 1b**) proteins, with 180 identified, suggesting that *S. mansoni* has a more compact set of protein kinases than any of the fully sequenced metazoa[16]. Most of the top 15 Pfam domains were from proteins involved in either intercellular communication or transcriptional regulation, which is expected for a parasite with multiple tissues and organs.

Being a metazoan

It has been proposed that the platyhelminth acoelomates, represented by *S. mansoni*, diverged from other eubilaterian metazoa more than a billion years ago[6]. As such, they lie somewhere between the unieukaryotes *Saccharomyces cerevisiae* and *Plasmodium falciparum* and the more advanced invertebrates *Caenorhabditis elegans*, *Drosophila melanogaster* and *Ciona intestinalis*. Phylogenetic analyses (ref. 6 and **Supplementary Fig. 4** online) support the ancient and independent divergence of acoelomates from other metazoa, which may explain the high fraction (55%) of SmAEs with no significant matches to sequences in GenBank. Thus, *S. mansoni* sequences should make an important contribution to understanding early metazoan evolution.

Metazoa-specific and eukarya-conserved sequences

We selected SmAEs that encode proteins that have been conserved among either the eukarya or the metazoa by comparison with known proteomes of organisms whose genomes have been completely sequenced. We built a metazoa-specific base set with the SmAEs that had orthologs only in each of the multicellular eukaryotes, *Homo sapiens*, *D. melanogaster*, *C. elegans* and *C. intestinalis*, but no matches with the unicellular eukaryotes, *S. cerevisiae* and *P. falciparum*, or with prokaryotes. The base set contains 1,598 sequences (~645 genes) that may be essential to the more complex metazoan cell functions. The eukarya-conserved sequences had at least one ortholog in all of the eukaryotes listed above. This data set contains 3,194 SmAEs (~1,443 genes), representing *S. mansoni* genes that would be important for eukaryotic cell functions.

Table 2 Cell adhesion and tissue structure orthologs

Adhesion molecules	Two protocadherins
	Two pannexins/innexins
	α3 and β2 integrins
Adherens junctions linking proteins	β-catenin
	Vinculin
	VASP homology protein (homer)
Actin polymerization	Small G proteins (Rho, Rac, Ras)
	Afadin
Tight junction proteins	Oap/Tspan3
	Several Maguk orthologs including Zo2
Extracellular matrix	Ten collagens
	Four laminins
	X and C tenascin
Integrin-cytoskeleton links	Talin
	Focal adhesion kinase
	Vinculin
	Actinin
Apoptosis components	Four caspases
	Two death-associated protein kinases
	Apoptosis-inducing factor (AIF)
	Bcl-2-interacting protein (beclin-1)
	Bax inhibitor
Autophagy components	Apg proteins 2–9 and 16
	Apg1p
	Autophagins Aut-1, Aut-2, Aut-3
	Target of rapamycin (Tor)

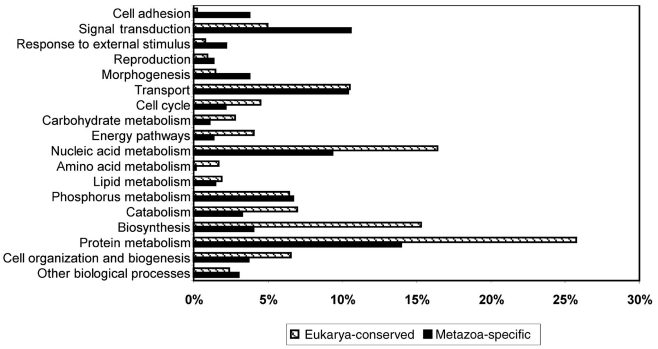

Figure 2 Category distribution of eukarya-conserved and metazoa-specific SmAEs. The metazoa-specific sequences (solid bars) have orthologs in each of the multicellular eukaryotes *H. sapiens*, *D. melanogaster*, *C. elegans* and *C. intestinalis* but not in the unicellular eukaryotes *S. cerevisiae* and *P. falciparum*. The essential and conserved eukarya SmAEs (striped bars) have orthologs in all of the eukaryotes listed above.

The relative distribution of SmAEs in Gene Ontology categories for the eukarya-conserved and metazoa-specific data sets (**Fig. 2**) shows that the latter set contains higher proportions of sequences in a few categories (cell-to-cell interactions, developmental processes, response to external stimulus and signal transduction). In general, the metazoa-specific sequences that have diverse roles in the tissues of a complex organism are overrepresented relative to the eukarya-conserved sequences.

Cell adhesion and tissue structure

As triploblastic acoelomates, schistosomes have three germ layers, bilateral symmetry, dorso-ventral patterning and rudimentary organs, for which intercellular adhesion mechanisms were an evolutionary prerequisite. The occurrence of homotypic cell adhesion is indicated by transcripts for protocadherins and the proteins that link them to the actin cytoskeleton in adherens junctions (**Table 2**). The small G proteins involved in actin polymerization are all present. The existence of organized tight junctions, important in maintaining the integrity of epithelia, can also be inferred, and evidence for gap junctions is provided by two pannexins/innexins. The extracellular matrix is represented by collagens, laminins and tenascins to which cells may attach by a potential integrin heterodimer; the intracellular links between integrins and the actin cytoskeleton are also evident.

The ability to undergo remodeling is a feature of organized tissues, but evidence for apoptosis is fragmentary. Some orthologs of this pathway were found (**Table 2**) whereas others (Bax, Bcl-2 family, endonuclease G) were not. In contrast, numerous components of autophagy were identified, apart from Apg13p and initiator Apg12p. This situation probably reflects the absence of wandering phagocytes to eliminate redundant cells.

Antero-posterior axis differentiation

S. mansoni has several axis-determining components in common with other metazoa. The presence of nanos, pumilio and the knirps gap-gene strongly suggests parallels with the mechanism used by *D.*

Table 3 Novel ortholog and paralog genes for transporters identified in *S. mansoni*

Sugar transporters	Ribose, hexoses, maltose and inositol
Lipid uptake transporters	Short chain fatty acid transporter
	ATP-driven phospholipid transporter
Amino acid transporters	Three ATP-driven (ABC cassette present): two uncertain, one sodium-dependent and one for oligopeptides
	Amino acid neurotransmitter:sodium symporters: GABA, glutamate, serotonin and dopamine transporters
Nucleotide transporters	Cytosine transporter and equilibrative nucleoside transporter
Ion transporters (ATP-driven Na^+-dependent or cotransporters)	Na^+/K^+
	K^+
	Ca^{2+}
	H^+
	K^+/Cl^- cotransporter
	Na/bicarbonate cotransporter
	Ion transporters for Fe, Ca, Zn, Mg, Mn, Cu, Co, SO_4, PO_4 and carboxylate
Ion transporters (channel/pore type transporters)	Voltage-gated chloride, potassium and calcium channels
	Voltage-dependent anion channel 1 (VDAC-1) and 2 (VDAC-2)
	Glutamate-gated ion channel (excitatory): NMDA or kainate selective
	Ligand-gated ion channel: nicotinic acetylcholine receptor
	Ryanodine-sensitive calcium-release channel
	Inositol 1,4,5-triphosphate-sensitive calcium-release channel
	Cyclic nucleotide-gated cation channel

melanogaster, in which maternal factors segregate to one pole of the egg and determine the antero-posterior axis. We detected the polycomb group transcripts, enhancer of zeste, polyhomeotic distal and extra sex combs, responsible for the maintenance of pattern, but none of the archetypal *Hox* cluster sequences. Orthologs of putative *S. mansoni* homeotic transcription factors included LIM-homeodomain, double homeobox protein 4 and homeotic protein Msx1.

Dorso-ventral patterning

Dorso-ventral patterning may be dictated by an analog of the TGF-β pathway. We identified activin/TGF-β receptor orthologs, Smad4, Smad8 and Medea as well as the known Smad1 and Smad2 (ref. 17). The R-Smads (Smad1, Smad2 and Smad8) are anchored to the plasma membrane by SARA, also newly identified. Specification of the dorso-ventral axis may also involve the Wnt pathway; we identified two Wnts and their transmembrane receptor frizzled as well as the cytosolic components of the intracellular signaling cascade dishevelled, axin, Gsk3 and β-catenin.

Epithelia

Adult schistosomes have three epithelia, surface tegument, gastrodermis and protonephridial canals, which control the transport of material into and out of their bodies. We found transcripts of villin family members supervillin and archvillin, which may cap and bundle actin filaments to provide an internal scaffold for cellular extensions cross-braced at their base by spectrin, also present. Functional studies have identified mediated transport of sugars, amino acids and nucleotides[18]. At least nine SmAEs for sugar transporters (some ATP-driven) can be added to the already cloned *Sgtp1*, *Sgtp2* and *Sgtp4* (ref. 19). We identified several transporters for lipids, amino acids, nucleotides and ions (**Table 3**).

Endocytosis is prominent in the gastrodermis but caveolin-type lipid rafts have also been postulated in the tegument surface[20]. We did not identify caveolin transcripts but did find the raft-associated flotillin. Transcripts for components of clathrin-mediated endocytosis included the clathrin heavy chain, assembly protein Ap180 and adaptor complex Ap2, which together encode all the functions to select cargo and form a vesicle. Dynamin, the master regulator of endocytosis, was present, along with phospholipid-interacting endophilin, Eps15 and epsin. In addition to low density lipoprotein–binding proteins[21], transcripts for serotransferrin, low density lipoprotein and very low density lipoprotein receptors attest to the importance of receptor-mediated endocytosis.

Motility and the nervous system

All life-cycle stages have an extensive and intricately organized musculature comprised of smooth fibers[22], and only the cercarial tail has a form of striated muscle. We identified transcripts for several myosins, two actins, tropomyosin, paramyosin and troponins C, I and T, involved in the regulation of contraction, the filament attachment proteins, α-actinin, vinculin and titin, many of which are novel paralogs. We found no transcripts encoding specific striated muscle proteins.

Platyhelminths are the first metazoan group to possess a central nervous system[23] and have a variety of sensory structures[24] that transduce a wide range of stimuli. Notch receptor, its transcription factor partner (suppressor of hairless) and membrane-bound ligand (delta) suggest a role for Notch signaling in *S. mansoni* neurogenesis. Transcripts for axon guidance molecules to direct nerves to their synaptic partners (netrin and its membrane receptor Unc5, two semaphorin-like and two plexin-like molecules) document the presence of a molecular repertoire for sophisticated neural circuitry. Regarding sensory structures, we identified components of the light detection system (a rhodopsin paralog of that previously described[8,25], rhodopsin kinase, arrestin and transducin), the first two in eggs and germ balls, respectively, consistent with the responsiveness of miracidia and cercariae to light.

Signaling

Transcriptome analysis identifies the molecular basis for some elements of schistosome neurotransmitter/receptor systems. We found ligand-gated channels, including three versions of the nicotinic acetylcholine receptor, choline o-acetyltransferase for synthesis and acetylcholine esterase for breakdown of this inhibitory neurotransmitter. We also found a glutamate receptor and transcripts for the γ-amino butyric acid (GABA) transporter and GABA receptor–associated protein but not the inhibitory GABA receptor itself.

We found G-protein-coupled receptors for glutamate and the excitatory transmitter serotonin along with its transporter, as well as a putative muscarinic acetylcholine receptor. Although *S. mansoni* has been reported to respond to catecholamine[26], we found no transcripts for the relevant receptors. Primitive neuroendocrine processes are known to be mediated by FaRP-type peptides[27], but we found a transcript only for allatostatin precursor protein. Nevertheless, orthologs of hormone proprotein convertase 2, which processes the precursors of bioactive peptides, and its regulatory neuroendocrine protein 7B2 were present, as was glycine peptidyl α-amide monooxygenase, required for the C-terminal amidation of the resulting peptides. Proprotein convertase 2 generates the opioid peptides and enkephalin in higher animals and might have the same function in schistosomes, as these peptides have previously been reported[28].

It is difficult to envisage how hormone signaling might operate in acoelomates, except over a short distance or through the neuroendocrine route. Nevertheless, two members of the nuclear receptor superfamily (retinoid-X and fushi tarazu factor 1) have been characterized[29], and SmAEs for a retinoic acid receptor (RAR-γ), a thyroid hormone receptor family member, a nuclear receptor 1 and a nuclear orphan receptor Tr2/4 can be added. But detection of transcripts for thyroid hormone interactor proteins 4, 12, 13 and 15 and thyroid hormone receptor–associated proteins Trap240 and Trap80, together with the reported effect of thyroid hormone on schistosome development[30], suggests that at least one nuclear orphan receptor may have a functional ligand. An ortholog of thyroid peroxidase, required to synthesize thyroid hormone, is present, but thyroglobulin, its vertebrate substrate, is not. If there is endogenous thyroid hormone, perhaps *S. mansoni* uses an alternative tyrosine-rich protein as a precursor.

The presence of transcripts for a series of cytochrome P450 enzymes, testosterone 6-β-hydroxylase and 17b-hydroxysteroid dehydrogenase suggests that schistosomes synthesize steroid hormones from cholesterol. They also seem to have some receptor elements (progesterone receptor membrane component 2 and estrogen-related receptor), which could bind endogenous steroids or mediate the supposed action of exogenous steroids on their maturation. Identification of other receptors for insulin and FGF, but not their ligands, reinforces the concept that host molecules act on parasite receptors. The presence of SmAEs encoding neurotensin and natriuretic peptide receptors is notable but more difficult to place in context.

Sex determination and sexual maturation

Most platyhelminths are hermaphrodites, but sexual dimorphism seems to have evolved separately on at least eight occasions, arguing for a relatively simple underlying mechanism[31]. Determination of sex is inherent whereas envelopment by the male is a prerequisite for female maturation[32], showing the need for cross-talk. We detected orthologs of *fox-1*, *mog-1*, *mog-4*, *tra-2* and *fem-1*, involved in the determination of sex in *C. elegans*. We also found the ortholog of mago-nashi, which in *C. elegans* (*mag-1*) specifies female development by inhibiting the hermaphrodite phenotype. The presence of the above transcripts in *S. mansoni* confirms their evolutionarily ancient role in sex determination, but it is unclear how they contribute to the dioecious state.

Being a parasite

Schistosomes have a prolonged association with their hosts and should therefore possess specific adaptations to the parasitic way of life. Adult worms are bathed in, and feed on, host blood, and we found transcripts for echicetin-like molecules that affect hemostasis and prevent thrombosis. Adult worms also expressed apyrase (CD39/ATP-diphosphohydrolase), an enzyme involved in platelet aggregation and thromboregulation that has been localized to the tegument[33], possibly indicating the capacity to inhibit platelet activation.

Longevity

In contrast to the short lifespan of *C. elegans* or *D. melanogaster*, schistosomes have predicted lifespan of 6–10 years[34]. In yeast and *C. elegans*, an extra copy of *Sir2* or *sir-2.1*, implicated in chromatin silencing, can increase lifespan, and we identified orthologs to *sir-2.1*, *sir-2.2*, *sir-2.5*, *sir-2.6* and *sir-2.7* in *S. mansoni*. We identified SmAEs from the insulin-signaling pathway, associated with longevity in *C. elegans*, including Daf2, an insulin-like receptor, Age1, a phosphatidylinositol-3-OH kinase and Daf16. Daf16 is a transcription factor that regulates many genes that affect lifespan, including enzymes that protect against or repair oxidative damage[35]. We also identified Pdk1 and PTEN, proteins that regulate the Daf2 pathway.

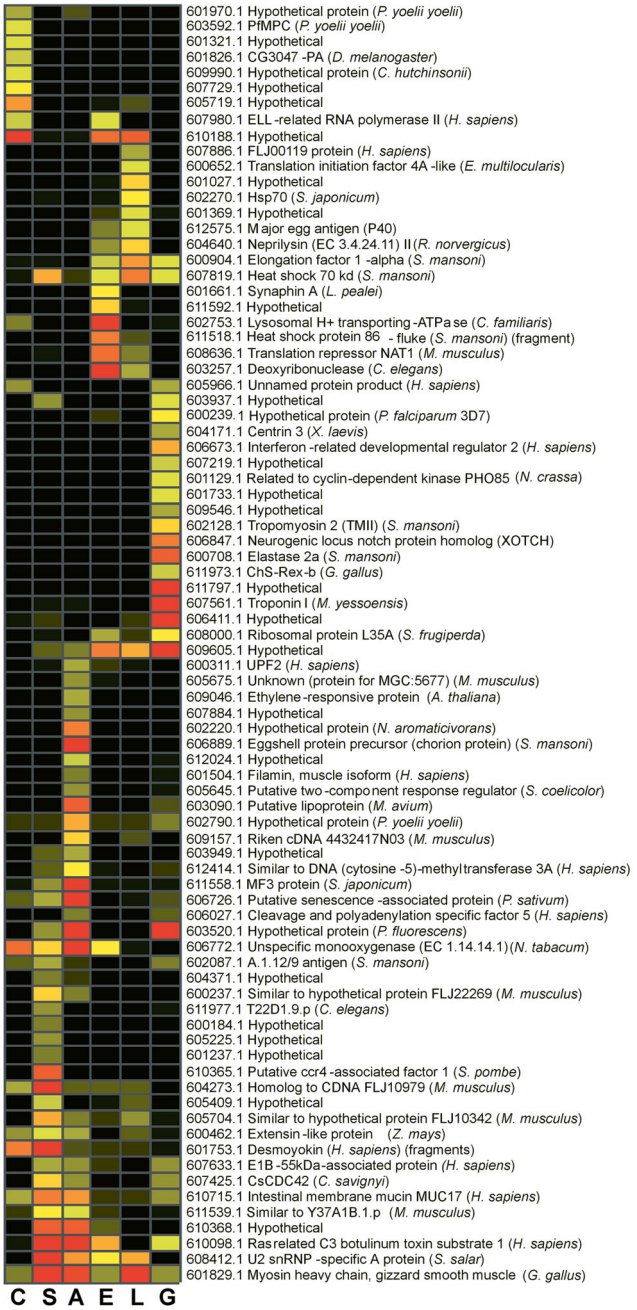

Figure 3 Frequency of sequenced transcripts in life-cycle stages. Hierarchical clustering of SmAEs using relative expression inference, estimated from the count of reads in a SmAE obtained with the same primer from each stage: C, cercaria; S, schistosomula; A, adults; E, eggs; L, miracidia; G, germ balls. The SmAE number and annotation of each gene are shown. Color scale indicates the number of counts with black representing no count and red representing a count above 20. *Cytophaga hutchinsonii*, *Loligo pealei*, *Canis familiaris*, *Neurospora crassa*, *Gallus gallus*, *Mizuhopecten yessoensis*, *Spodoptera frugiperda*, *Neuorospora aromaticivorans*, *Streptomyces coelicolor*, *Mycobacterium avium*, *Pisum sativum*, *Pseudomonas fluorescens*, *Neurospora tabacum*, *Zea mays*, *Ciona savignyi* and *Salmo salar* are the full names of species not previously mentioned.

Stress responses

S. mansoni undergoes rapid transitions between environments that are accompanied by temperature and osmotic stresses. We extended the list of previously described heat shock genes (23 SmAEs, 12 possibly new), which includes an HtrA ortholog, a stress-regulated serine protease. Uroplakin is believed to limit the permeability of membranes to water and small non-electrolytes[36]; we found an ortholog in egg, miracidia and cercaria stages. Parasites also encounter oxidative stress during host immune attack, which is dealt with by antioxidant enzymes, both previously characterized (superoxide dismutases, thioredoxin and glutathione reductases and peroxidases) and novel, including mitochondrial thioredoxin 2, a PKC-interacting thioredoxin, thioredoxin-like 2, an ortholog of *Plasmodium yoelii* thioredoxin, and glutaredoxin 3.

The innate immune response comprises primitive mechanisms used by metazoa in defense against infection[14,15]. The Toll pathway has an important role in this, and we identified several components including Tollip, pellino and NF-κB kinase (NEMO), implying that *S. mansoni* can respond to extracellular pathogens. The presence of transcripts for adenosine deaminase, Dicer and Piwi/argonaute indicates that *S. mansoni* can also deal with intracellular attack mediated by viral dsRNA. By extension, the last two genes indicate that post-transcriptional gene silencing could occur, and the use of RNA interference to suppress schistosome gene function was recently reported[37,38].

Evasion of host immune responses

S. mansoni has been proposed to use several strategies to evade host immune responses, including protection of the tegument surface by a secreted membranocalyx[39], molecular mimicry, antigenic variation and immunomodulation. As an example of molecular mimicry, the convergent evolution of *S. mansoni* and *Biomphalaria glabrata* (snail intermediate host) tropomyosins 1 and 2, has been suggested[40] on the basis of immunological cross-reactivity and amino acid sequence identity (~63%). We detected a new isoform, tropomyosin 3, in adults, eggs and germ balls with only 35% amino acid identity to *B. glabrata*, suggesting a different tissue location not subjected to the same selective pressure.

In the context of antigenic variation, we found no evidence of highly variable gene families (compared with *Plasmodium*), but our database identified 449 putative novel paralogs to known *S. mansoni* genes (**Table 1**); 33 of these had high identity and >30% coverage (**Supplementary Table 4** online). This multiplicity of isoforms would allow the parasite to use paralogs of an essential enzyme targeted by the immune system to avoid loss of function, thus making vaccine development more difficult. Indeed, we identified several paralogs of previously investigated vaccine candidates (**Supplementary Table 5** online).

Non-synonymous single-nucleotide polymorphisms (SNPs) are another source of variation. Analysis of redundant EST coverage of genes encoding vaccine candidates identified eight putative polymorphisms, two of which could be validated (see **Supplementary Methods** online) in isolates from different regions of the world. We detected alternative splicing in several genes, including a recently identified exon skipping in Sm14 (ref. 41) present in germ balls, schistosomula and adults.

Modulation of mammalian host immune responses by a schistosome infection is well documented, but the agents and mechanisms are not yet fully defined. The presence of transcripts for pro-inflammatory phospholipase A2-activating protein supports the documented effect of lyso-phosphatidylserine as an inducer of T-regulatory cells and Th2 polarization[42]. *S. mansoni* eggs and adults induce a characteristic allergic response[43,44]. The identification of a family of orthologs to wasp venom allergen 5 raises the question of how the parasite benefits from amplifying such a response.

Stage-associated frequency of sequences

The frequency of reads in a SmAE cluster obtained from different life cycle stages can reflect differential gene expression when the same set of primers is used for generating ORESTES minilibraries. We validated this approach experimentally by semi-quantitative RT–PCR (**Supplementary Fig. 5** online). We analyzed 5,172 sequences obtained with the same set of primers, generating 2,058 SmAEs. We found that 82 of these had conspicuously different patterns of distribution among stages (with 99.8% confidence), several being predominant in one stage only (**Fig. 3** and **Supplementary Table 6** online). In particular, germ balls overexpressed elastase 2a (secretion for host invasion[45]), troponin I and tropomyosin 2 (muscle development), and centrin3 and S-rex/Nsp (differentiation).

Potential drug targets and multidrug-resistance genes

One main benefit from our project should be the identification of novel proteins amenable to rational drug design. Selected examples of potential molecular targets are detailed in **Table 4**. Existing anthelminthics[46] that disrupt neurotransmission provide the rationale for one group. Paralogs of calcium channel subunits, the targets of praziquantel, and cyclophillins, which mediate the antischistosomal effect of cyclosporin, are also listed. Molecules proposed as targets in other systems include innexins (connexins of vertebrates) and DNA polymerase. We identified transcripts for several multidrug resistance transporters, however, which could complicate the development of new drugs.

Potential vaccine candidates

Potential vaccine candidates should include proteins that are preferentially surface-exposed or exported and that are expressed in intramammalian stages. These properties can be searched for using Gene Ontology categorization. Thus, orthologs of secreted toxins and surface proteins involved in cell adhesion both warrant investigation (**Table 5**). Three orthologs of *Plasmodium* circumsporozoite protein, expressed in schistosomula and adults, and an ortholog of the *S. cerevisiae* threonine-rich cell-wall protein may be surface-exposed. Likewise, receptors that potentially bind host hormones should be accessible to the immune system. Targeting glycosyl phosphatidyl inositol–anchored proteins or receptors for nutrients could impair vital functions in the parasite and thus provide another avenue for vaccine development.

DISCUSSION

Our study of the *S. mansoni* transcriptome increases tenfold the number of ESTs available to define the gene complement of this blood fluke and will be an essential resource for annotation of its genome. Our overall impression of this member of one of the simplest extant bilaterian groups is that most, if not all, of the cellular and physiological systems of higher animals were established before the divergence of the platyhelminths. Thus, components required for tissue organization and smooth muscle function were present at an early stage of metazoan evolution. An extensive range of neurotransmitter systems and enzymes for the generation of neuropeptides and opioid peptides indicates substantial capacity for neurosecretory control of physiology. Potential components of thyroid and steroid hormone systems were identified; it will be pertinent to establish the source of ligands for the relevant receptors. Apoptosis seems to be a later evolutionary development, however, with autophagy the predominant means of removing unwanted cells.

Features of the transcriptome that can be associated with the parasitic way of life are more difficult to define. One probable reason for this is that we found no similarity for 55% of SmAEs. A singular advantage of parasitism is the ready access to a supply of nutrients, uptake of which is facilitated by a wide variety of transporters and

receptors for lipids and cholesterol. With respect to immune evasion, the paucity of mechanisms for antigenic variation, compared with *Plasmodium* or *Trypanosoma*, is notable. Immune evasion by secretion of an inert bilayer masking the parasite-host interface can now be investigated by combining the transcriptome database with proteomics techniques to elucidate the architecture of the tegument surface. A similar approach should allow identification of protein immunomodulators known to be released by cercariae, adult worms and eggs.

We should not forget that *S. mansoni* is an important human pathogen with no vaccine and a single drug for treatment. Mining the SmAE database for drug targets and vaccine candidates should therefore be a priority. By analogy with other systems, we have singled out a number of chemotherapeutic possibilities from a potentially long list. The prediction of vaccine candidates from sequence information alone is highly speculative, but key antigens should now be identifiable by immunological studies in experimental animals and humans.

Table 4 Chemotherapy in schistosomiasis: potential new drug targets

SmAEs	Gene	Similar to	Remarks
C710243.1, C719264.1	Nicotinic acetylcholine receptor	*Felis catus*, *H. sapiens*	Levamisole and pyrantel bind nematode nAChR
C705718.1	Choline O-acetyltransferase	*D. melanogaster*	In filarial worms the enzyme has a key role in motility; ChAT is inhibited by low doses of ethacrynic acid
C713648.1	Acetylcholinesterase	*Schizaphis graminum*	Metrifonate inhibits acetylcholinesterase
C708367.1	Muscarinic acetylcholine receptor	*D. melanogaster*	Levamisole and pyrantel possibly bind nematode mAChR
C603771.1, C711869.1	Glutamate transporter *Glut2* and *AmEAAT*	*C. elegans*, *Apis mellifera*	Removes glutamate, an excitatory neurotransmitter, and permits normal neurotransmission; putative drug target if significantly different from the mammalian protein
C610861.1, C705975.1, C716672.1, C719080.1	Glutamate receptors	*Mus musculus*, *D. melanogaster*, *Rattus norvegicus*, *Lymnaea stagnalis*	Ivermectin is believed to exert its anthelmintic effects by binding to glutamate-gated chloride channels
C714193.1	GABA transporter	*R. novergicus*	Piperazine binds to *Ascaris* GABA receptors
C702111.1	Serotonin receptor	*Anopheles gambiae*	
C609540.1, C718443.1	L-type calcium channel alpha subunit	*Porcellio scaber*, *Stylophora pistillata*	Praziquantel is believed to act through tegument Ca^{++} channels
C601467.1, C609572.1, C602142.1	Cyclophilin-like and matrin cyclophilin	*D. melanogaster*, *R. norvegicus*	Cyclosporin binds to cyclophilin and has an antiparasitic effect against helminths and protozoa
C605281.1, C610889.1, C608660.1	Innexins (*Unc7*, *Unc9* and *Inx1* Gap junction proteins)	*C. elegans*	Neuromuscular ion channel exclusive from invertebrates; proposed as targets in cancer chemotherapy
C600095.1, C717578.1, C703546.1	DNA polymerase delta	*D. melanogaster*, *H. sapiens*, *A. gambiae*	Target for antiviral drugs
C604319.1, C601691.1, C605154.1, C706943.1, C707248.1, C714828.1	*Smdr2* paralogs	*S. mansoni*	Paralogs of the previously known *S. mansoni* SMDR2
C600192.1, C703117.1	MDR7	*H. sapiens*	Drug resistance
C605069.1, C605742.1, C707555.1, C706898.1, C703117.1	ATP-binding cassette protein (*Cftr/Mrp*), sub-family C, multidrug resistance-associated protein *MRP2*	*A. gambiae*, *M. musculus*, *R. norvegicus*	Drug resistance, prevents amphiphilic organic anions accumulation, transports glutathione conjugates
C715202.1, C711423.1	RND multidrug efflux transporter	*Nostoc* sp.	
C609997.1	Breast cancer resistance protein (ABC G2)	*H. sapiens*	Drug resistance, prevents anthracycline accumulation
C604844.1	Phosphoglycerate mutase	*Schistosoma japonicum*	Clorsulon is a selective antagonist of fluke phosphoglycerate kinase and mutase, and the enzyme is important to maintain parasitic infection
C608696.1	Toll Interacting Protein (*Tollip*)	*Danio rerio*	Inflammatory response, IL18 receptor complex; negative regulation of TLR-1
C609382.1, C610315.1	Adenylate cyclase	*C. elegans*, *H. sapiens*	Synthesis of cyclic AMP from ATP; putative drug target if significantly different from their mammalian counterpart
C606856.1, C612634.1	Stomatin	*A. gambiae*, *C. elegans*	Interaction with anti-malarial drugs; mechanoreception or lipid anchorage; uptake of exogenous phospholipid, binds to HDL

Table 5 Novel *S. mansoni* genes to be investigated as vaccine candidates

Category	SmAE	Orthologous protein (size in amino acids)	Organism (identity, coverage)	Possible function
Toxins	C607733.1	Wasp venom allergen 5 (202 aa)	*Vespa mandarinia* (33%, 202 aa)	Exotoxin, allergen
	C602160.1	Wasp venom allergen 5 (205 aa)	*Vespula squamosa* (35%, 143 aa)	Exotoxin, allergen
	C708986.1	Wasp venom allergen 5 (206 aa)	*Vespula vidua* (38%, 167 aa)	Exotoxin, allergen
	C600509.1	Wasp venom allergen 5 (204 aa)	*Vespula pensylvanica* (31%, 161 aa)	Exotoxin, allergen
	C712286.1	Echicetin-α subunit (177 aa)	*Echis carinatus* (36%, 74 aa)	Exotoxin, sugar binding; inhibits binding of von Willebrand factor and alboaggregins to platelet glycoprotein Ib
	C607255.1	Sphingomyelin phosphodiesterase 2 (419 aa)	*M. musculus* (35%, 101 aa)	Esterase secreted to effect target cell lysis
Cell surface adhesion, receptors	C600716.1	CD36 / scavenger receptor class 3 (509 aa)	*B. taurus* (30%, 509 aa)	Cell adhesion; scavenger receptor class B type 1; platelet and leukocyte adhesion; evidence also suggests a role in signal transduction
	C611319.1	CD36 / Lysosyme membrane protein II (531 aa)	*D. rerio* (35%, 236 aa)	Cell adhesion; scavenger receptor class B type 1; platelet and leukocyte adhesion; evidence also suggests a role in signal transduction
	C603064.1	CD18 / β-integrin (771 aa)	*M. musculus* (40%, 289 aa)	Cell surface adhesion glycoprotein, leukocyte adhesion protein, complement receptor C3
	C602256.1	CD18 / β-integrin (771 aa)	*M. musculus* (32%, 251 aa (N-terminal))	Cell surface adhesion glycoprotein, leukocyte adhesion protein, complement receptor C3
Surface-exposed membrane proteins	C604900.1	Circumsporozoite protein CSP (386 aa)	*P. simium* (44%, 182 aa)	Main *Plasmodium* outer membrane protein, proline-rich
	C607313.1	Circumsporozoite protein CSP (388 aa)	*Plasmodium reichnowi* (36%, 122 aa)	Main *Plasmodium* outer membrane protein, proline-rich
	C706735.1	Circumsporozoite Protein CSP (513 aa)	*Plasmodium chabaudi* (44%, 100aa)	Main *Plasmodium* outer membrane protein, proline-rich
	C600436.1	Cell wall protein delayed anaerobic (1,161 aa)	*S. cerevisiae* (42%, 116 aa)	Membrane protein, response to stress, threonine-rich
Receptors for host factors	C600934.1	VLDL receptor (869 aa)	*Xenopus laevis* (43%, 97 aa)	Fatty acid metabolism, transmembrane protein
	C606856.1	Stomatin, erythrocyte band 7 (356 aa)	*H. sapiens* (62%, 201 aa)	Mechanoreception or lipid anchorage; involved in calcium transport through lipid membranes; uptake of exogenous phospholipids, binds to HDL; interaction with anti-malarial drugs
	C703516.1	FGF receptor (877 aa)	*Halocynthia roretzi* (32%, 112 aa)	Fibroblast growth factor receptor, member of immunoglobulin superfamily, cell surface recognition
	C602729.1	Activin IIB /TGF-β (512 aa)	*B. taurus* (38%, 80 aa)	Serine/threonine protein kinase, signal transduction, localized to tegument
	C700977.1	Activin IIB /TGF-β (504 aa)	*Carassius auratus* (32%, 110 aa)	Serine/threonine protein kinase, signal transduction, localized to tegument
	C600474.1	Insulin receptor (1,749 aa)	*Echinococcus multilocularis* (34%, 109 aa)	Receptor for insulin, surface exposed by analogy with *E. multilocularis*
	C611659.1	Insulin receptor (1,749 aa)	*E. multilocularis* (27%, 459 aa)	Receptor for insulin, surface exposed by analogy with *E. multilocularis*
	C611659.1	Insulin receptor (1,749 aa)	*E. multilocularis* (68%, 172 aa)	Receptor for insulin, surface exposed by analogy with *E. multilocularis*
Surface exposed enzymes	C710539.1	Leishmanolysin-like peptidase (640 aa)	*H. sapiens* (37%, 243 aa)	Metalopeptidase, endopeptidase; membrane bound by GPI anchor; most abundant cell surface protein in *Leishmania* promastigotes
	C602834.1	Carboxipeptidase N (458 aa)	*H. sapiens* (41%, 58 aa)	Metalopeptidase regulating biological activity of kinins and anaphyltoxins (human plasma)
	C608649.1	Carboxipeptidase N (458 aa)	*M. musculus* (45%, 176 aa)	Metalopeptidase regulating biological activity of kinins and anaphyltoxins (human plasma)
	C609556.1	Esterase, β-lactamase (429 aa)	*C. elegans* (34%, 210 aa)	Penicillin binding protein, cell envelope biogenesis, outer membrane
	C607243.1	Alkaline phosphatase (524 aa)	*M. musculus* (36%, 332 aa)	Phosphate ester hydrolysis; glycoprotein attachedto membrane by GPI anchor; anti-AP antibodies detected in sera of infected individuals
	C608449.1	Apyrase, ecto-ATP diphosphohydrolase2 (306 aa)	*H. sapiens* (37%, 132 aa)	Extracellular ATP hydrolysis signaling; integral membrane protein; localized to tegument

METHODS

Parasites. We maintained the BH and PR isolates of *S. mansoni* in the laboratory by routine passage through mice and snails and recovered parasite life cycle stages as described in **Supplementary Methods** online. We concentrated cercaria, schistosomula and adults by centrifugation and stored them at –20 °C in RNAlater (Ambion) according to the manufacturer's recommendations before extracting mRNA. We used freshly isolated parasites from the other stages (eggs, miracidia and germ balls) for immediate extraction of mRNA.

Construction of cDNA libraries and sequencing. We obtained DNase-treated mRNA with MACs mRNA isolation kits (Miltenyi Biotec) and used it to construct cDNA and SAGE libraries. We carried out cDNA synthesis and amplification using the ORESTES protocol with modifications[12,47] (see **Supplementary Methods** online). We prepared normalized poly-dT-primed cDNA libraries as previously described[10] using the abundantly available mRNA from adult worms. We sequenced cDNA using standard fluorescence-labeling dye-terminator protocols. To analyze differential gene expression, we used a set of six primers to construct ORESTES cDNA minilibraries from all stages. Sequencing of at least two 96-well plates per library resulted in at least 140 sequences per stage per primer (see **Supplementary Methods** online).

EST processing pipeline and annotation. We stored, processed and trimmed EST sequence chromatograms through a web-based service[48] and accepted sequences with at least 100 bp with phred-15 or higher for further evaluation. We filtered sequences using BLASTN analysis with a local copy of GenBank NT database and the BlastMachine (Paracel) to eliminate those that matched non-*S. mansoni* sequences with $E \leq 10^{-15}$ and had at least 98% identity along at least 75 nucleotides. We also excluded reads that matched *S. mansoni* ribosomal or mitochondrial sequences and transposon sequences with $E \leq 10^{-15}$ and at least 85% identity along at least 75 nucleotides or that matched bacterial sequences with $E \leq 10^{-20}$ and at least 95% identity along at least 75 nucleotides. We filtered further transposon and bacterial sequences by comparing with BLASTX against the set of transposon and bacterial sequences from GenBank NR and eliminating those with matching $E \leq 10^{-4}$ and at least 30% identity along at least 75 amino acids with transposons or matching $E \leq 10^{-6}$ and at least 95% identity along at least 75 amino acids with bacteria. We clustered and assembled ESTs using CAP3 (ref. 49). We assigned putative protein products to SmAEs based on BLASTX hits to National Center for Biotechnology Information's NR database. We assigned Gene Ontology terms to SmAEs based on BLASTX hits against a database locally built from public sequences associated with Gene Ontology terms. The public Gene Ontology annotated data sets used were from *H. sapiens*, *D. melanogaster*, *Arabidopsis thaliana*, *Oryza sativa*, *C. elegans*, *S. cerevisiae*, *Schizosaccharomyces pombe* and *Vibrio cholerae* plus a curated sequence database (Gene Ontology Annotation at EBI) available at the Gene Ontology Consortium website. In both cases, we used $E \leq 10^{-6}$ as the BLASTX cut-off. We used ESTscan to deduce amino acid sequences and used them as queries against the Pfam database 7.8.

SAGE. We constructed a SAGE library with mRNA derived from adult worms (males and females) using the I-SAGE Kit (Invitrogen). We treated poly(A)$^+$ mRNA with DNase before extraction with oligo-dT. We cloned and sequenced concatamers and derived tags from high-quality sequence segments. To determine the relative abundance of transcripts in adult worms, we compared the SAGE tag list with the complete SmAE data set and with all full-length cDNA sequences from *S. mansoni*.

Phylogeny inferences. We aligned protein sequences using the ClustalX multiple sequence alignment program. Only unambiguous positions were used in the phylogenetic analysis. We generated phylogenetic trees using the Phylip program as described in **Supplementary Methods** online.

Differential expression analysis. To evaluate differential expression, we assembled the ORESTES sequences derived from six primers along all six life cycle stages and considered the number of reads per stage for each cluster as an indirect inference of the expression level in the stage. Sequences with a differential frequency of reads by stage (99.8% confidence) when analyzed by a randomization test[50] are discussed. Hierarchical clustering of these data was done using correlation distance UPGMA as provided in the Spotfire for Functional Genomics software (Spotfire). We carried out semi-quantitative RT–PCR to confirm differential expression of three selected genes (see **Supplementary Methods** online).

SNP analysis. We identified putative SNPs in *S. mansoni* genes using Polybayes as described in **Supplementary Methods** online. We selected a fraction of the putative SNPs in vaccine candidates for experimental validation using DNA derived from pooled adult worms (see **Supplementary Methods** online).

URLs. Project website including *Schistosoma* Gene Ontology browser, BLAST server and SmAEs search tools, http://bioinfo.iq.usp.br/schisto/; The Institute for Genomic Research *S. mansoni* genome project, http://www.tigr.org/tdb/e2k1/sma1/; The Sanger Institute *S. mansoni* genome project, http://www.sanger.ac.uk/Projects/S_mansoni/; The Phred/Phrap/Consed System Home Page, http://www.phrap.org/; National Center for Biotechnology, http://www.ncbi.nlm.nih.gov/BLAST/; Gene Ontology Consortium, http://www.geneontology.org/; ESTScan2 server, http://www.ch.embnet.org/software/ESTScan2.html; Pfam server, http://www.sanger.ac.uk/Software/Pfam/.

Accession numbers. Sequences were deposited in GenBank under accession numbers CD059164–CD088507, CD088510–CD120734, CD120740–CD150744 and CD151578–CD202980. SNPs identified in this study were deposited in dbSNP at National Center for Biotechnology Information under the accession numbers ss8486502–ss8486509.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

E.D.N. thanks Associação Beneficente Alzira Denise Hertzog da Silva for financial support, D. Rollinson for providing schistosome isolates from Africa and Lebanon and M.G. dos Reis and N. Lucena for providing isolates from northeast Brazil. This project was financed by Fundação de Amparo a Pesquisa do Estado de Sao Paulo and by the Brazilian Ministry of Science and Technology , Conselho Nacional de Desenvolvimento Científico e Tecnológico. The York schistosomiasis group received support from the Biology and Biotechnology Science Research Council, Wellcome Trust and the European Commission Research for Development Programme, Sector Health.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 20 May; accepted 17 August 2003
Published online at http://www.nature.com/naturegenetics/

1. World Health Organization. *TDR Strategic Direction for Research: Schistosomiasis* (World Health Organization, Geneve, 2002).
2. King, C.L. Initiation and regulation of disease in schistosomiasis. in *Schistosomiasis* (ed. Mahmoud, A.A.F.) 213–264 (Imperial College Press, London, 2001).
3. Doenhoff, M.J., Kusel, J.R., Coles, G.C. & Cioli, D. Resistance of Schistosoma mansoni to praziquantel: is there a problem? *Trans. R. Soc. Trop. Med. Hyg.* **96**, 465–469 (2002).
4. Dunne, D. & Mountford, A. Resistance to infection in humans and animal models. in *Schistosomiasis* (ed. Mahmoud, A.A.F.) 133–211 (Imperial College Press, London, 2001).
5. Coulson, P.S. The radiation-attenuated vaccine against schistosomes in animal models: paradigm for a human vaccine? *Adv. Parasitol.* **39**, 271–336 (1997).
6. Hausdorf, B. Early evolution of the bilateria. *Syst. Biol.* **49**, 130–142 (2000).
7. Simpson, A.J., Sher, A. & McCutchan, T.F. The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequences. *Mol. Biochem. Parasitol.* **6**, 125–137 (1982).
8. Santos, T.M. *et al.* Analysis of the gene expression profile of *Schistosoma mansoni* cercariae using the expressed sequence tag approach. *Mol. Biochem. Parasitol.* **103**, 79–97 (1999).
9. Williams, S.A. & Johnston, D.A. Helminth genome analysis: the current status of the filarial and schistosome genome projects. Filarial Genome Project. Schistosome Genome Project. *Parasitology* **118 Suppl**, S19–S38 (1999).
10. Soares, M.B. *et al.* Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* **91**, 9228–9232 (1994).
11. Dias-Neto, E. *et al.* Minilibraries constructed from cDNA generated by arbitrarily primed RT–PCR: an alternative to normalized libraries for the generation of ESTs from nanogram quantities of mRNA. *Gene* **186**, 135–142 (1997).
12. Dias-Neto, E. *et al.* Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. USA* **97**, 3491–3496 (2000).
13. Adams, M.D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
14. Dehal, P. *et al.* The draft genome of *Ciona intestinalis*: insights into chordate and

vertebrate origins. *Science* **298**, 2157–2167 (2002).
15. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
16. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
17. Osman, A., Niles, E.G. & LoVerde, P.T. Identification and characterization of a *Smad2* homologue from *Schistosoma mansoni*, a transforming growth factor-β signal transducer. *J. Biol. Chem.* **276**, 10072–10082 (2001).
18. Pappas, P.W. Membrane transport in helminth parasites: a review. *Exp. Parasitol.* **37**, 469–530 (1975).
19. Skelly, P.J., Kim, J.W., Cunningham, J. & Shoemaker, C.B. Cloning, characterization, and functional expression of cDNAs encoding glucose transporter proteins from the human parasite *Schistosoma mansoni*. *J. Biol. Chem.* **269**, 4247–4253 (1994).
20. Racoosin, E.L., Davies, S.J. & Pearce, E.J. Caveolae-like structures in the surface membrane of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **104**, 285–297 (1999).
21. Xu, X. & Caulfield, J.P. Characterization of human low density lipoprotein binding proteins on the surface of schistosomula of *Schistosoma mansoni*. *Eur. J. Cell Biol.* **57**, 229–235 (1992).
22. Mair, G.R., Maule, A.G., Day, T.A. & Halton, D.W. A confocal microscopical study of the musculature of adult *Schistosoma mansoni*. *Parasitology* **121**, 163–170 (2000).
23. Halton, D.W. & Gustafsson, M.K.S. Functional morphology of the platyhelminth nervous system. *Parasitology* **113**, S47–S72 (1996).
24. Dorsey, C.H., Cousin, C.E., Lewis, F.A. & Stirewalt, M.A. Ultrastructure of the *Schistosoma mansoni* cercaria. *Micron* **33**, 279–323 (2002).
25. Hoffmann, K.F., Davis, E.M., Fischer, E.R. & Wynn, T.A. The guanine protein coupled receptor rhodopsin is developmentally regulated in the free-living stages of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **112**, 113–123 (2001).
26. Pax, R.A. & Bennett, J.L. Neurobiology of parasitic platyhelminths: possible solutions to the problems of correlating structure with function. *Parasitology* **102 Suppl**, S31–S39 (1991).
27. Smart, D. *et al.* Peptides related to the *Diploptera punctata* allatostatins in nonarthropod invertebrates: an immunocytochemical survey. *J. Comp. Neurol.* **347**, 426–432 (1994).
28. Pryor, S.C. & Elizee, R. Evidence of opiates and opioid neuropeptides and their immune effects in parasitic invertebrates representing three different phyla: *Schistosoma mansoni*, *Theromyzon tessulatum*, *Trichinella spiralis*. *Acta Biol. Hung.* **51**, 331–341 (2000).
29. de Mendonca, R.L., Escriva, H., Bouton, D., Laudet, V. & Pierce, R.J. Hormones and nuclear receptors in schistosome development. *Parasitol. Today* **16**, 233–240 (2000).
30. Saule, P. *et al.* Early variations of host thyroxine and interleukin-7 favor *Schistosoma mansoni* development. *J. Parasitol.* **88**, 849–855 (2002).
31. Snyder, S.D., Loker, E.S., Johnston, D.A. & Rollinson, D. The Schistosomatidae: Advances in Phylogenetics and Genomics. in *The Interrelationships of Platyhelminthes* (eds. Littlewood, D.T.J. & Bray, R.A.) 194–199 (Taylor and Francis, London, 2000).
32. Basch, P.F. *Schistosoma mansoni*: nucleic acid synthesis in immature females from single-sex infections, paired *in vitro* with intact males and male segments. *Comp. Biochem. Physiol. B* **90**, 389–392 (1988).
33. DeMarco, R., Kowaltowski, A.T., Mortara, R.A. & Verjovski-Almeida, S. Molecular characterization and immunolocalization of *Schistosoma mansoni* ATP-diphosphohydrolase. *Biochem. Biophys. Res. Commun.* **307**, 831–838 (2003).
34. Fulford, A.J., Butterworth, A.E., Ouma, J.H. & Sturrock, R.F. A statistical approach to schistosome population dynamics and estimation of the life-span of *Schistosoma mansoni* in man. *Parasitology* **110 (Pt 3)**, 307–316 (1995).
35. Murphy, C.T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424**, 277–283 (2003).
36. Hu, P. *et al.* Role of membrane proteins in permeability barrier function: uroplakin ablation elevates urothelial permeability. *Am. J. Physiol. Renal Physiol.* **283**, F1200–F1207 (2002).
37. Skelly, P.J., Da'dara, A. & Harn, D.A. Suppression of cathepsin B expression in *Schistosoma mansoni* by RNA interference. *Int. J. Parasitol.* **33**, 363–369 (2003).
38. Boyle, J.P., Wu, X.J., Shoemaker, C.B. & Yoshino, T.P. Using RNA interference to manipulate endogenous gene expression in *Schistosoma mansoni* sporocysts. *Mol. Biochem. Parasitol.* **128**, 205–215 (2003).
39. Wilson, R.A. & Barnes, P.E. The formation and turnover of the membranocalyx on the tegument of *Schistosoma mansoni*. *Parasitology* **74**, 61–71 (1977).
40. Dissous, C. & Capron, A. Convergent evolution of tropomyosin epitopes. *Parasitol. Today* **11**, 45–46 (1995).
41. Ramos, C.R. *et al.* Gene structure and M20T polymorphism of the *Schistosoma mansoni* Sm14 fatty acid-binding protein. Molecular, functional, and immunoprotection analysis. *J. Biol. Chem.* **278**, 12745–12751 (2003).
42. van der Kleij, D. *et al.* A novel host-parasite lipid cross-talk. Schistosomal lyso-phosphatidylserine activates toll-like receptor 2 and affects immune polarization. *J. Biol. Chem.* **277**, 48122–48129 (2002).
43. Cutts, L. & Wilson, R.A. Elimination of a primary schistosome infection from rats coincides with elevated IgE titres and mast cell degranulation. *Parasite Immunol.* **19**, 91–102 (1997).
44. Damonneville, M., Pierce, R.J., Verwaerde, C. & Capron, A. Allergens of *Schistosoma mansoni*. II. Fractionation and characterization of *S. mansoni* egg allergens. *Int. Arch. Allergy Appl. Immunol.* **73**, 248–255 (1984).
45. Salter, J.P. *et al.* Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes. *J. Biol. Chem.* **277**, 24618–24624 (2002).
46. Mansour, T.E. *Chemotherapeutic Targets in Parasites* (Cambridge University Press, Cambridge, 2002).
47. Fietto, J.L., DeMarco, R. & Verjovski-Almeida, S. Use of degenerate primers and touchdown PCR for construction of cDNA libraries. *Biotechniques* **32**, 1404–1411 (2002).
48. Paquola, A., Nishiyama, M. Jr., Reis, E.M., daSilva, A.M. & Verjovski-Almeida, S. ESTWeb: bioinformatics services for EST sequencing projects. *Bioinformatics* **19**, 1587–1588 (2003).
49. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999).
50. Stekel, D.J., Git, Y. & Falciani, F. The comparison of gene expression from multiple cDNA libraries. *Genome Res.* **10**, 2055–2061 (2000).