

Database

Open Access

## SjTPdb: integrated transcriptome and proteome database and analysis platform for *Schistosoma japonicum*

Feng Liu\*<sup>†1,2</sup>, Ping Chen<sup>†1</sup>, Shu-Jian Cui<sup>1</sup>, Zhi-Qin Wang<sup>1</sup> and Ze-Guang Han\*<sup>1,2</sup>

Address: <sup>1</sup>Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, PR China and <sup>2</sup>Shanghai Jiao Tong University School of Medicine, Shanghai, PR China

Email: Feng Liu\* - liuf@chgc.sh.cn; Ping Chen - baobaocp121@hotmail.com; Shu-Jian Cui - cuijs@chgc.sh.cn; Zhi-Qin Wang - wangzq@chgc.sh.cn; Ze-Guang Han\* - hanzg@chgc.sh.cn

\* Corresponding authors †Equal contributors

Published: 26 June 2008

Received: 17 October 2007

BMC Genomics 2008, 9:304 doi:10.1186/1471-2164-9-304

Accepted: 26 June 2008

This article is available from: <http://www.biomedcentral.com/1471-2164/9/304>

© 2008 Liu et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** *Schistosoma japonicum* is one of the three major blood fluke species, the etiological agents of schistosomiasis which remains a serious public health problem with an estimated 200 million people infected in 76 countries. In recent years, enormous amounts of both transcriptomic and proteomic data of schistosomes have become available, providing information on gene expression profiles for developmental stages and tissues of *S. japonicum*. Here, we establish a public searchable database, termed SjTPdb, with integrated transcriptomic and proteomic data of *S. japonicum*, to enable more efficient access and utility of these data and to facilitate the study of schistosome biology, physiology and evolution.

**Description:** All the available ESTs, EST clusters, and the proteomic dataset of *S. japonicum* are deposited in SjTPdb. The core of the database is the 8,420 *S. japonicum* proteins translated from the EST clusters, which are well annotated for sequence similarity, structural features, functional ontology, genomic variations and expression patterns across developmental stages and tissues including the tegument and eggshell of this flatworm. The data can be queried by simple text search, BLAST search, search based on developmental stage of the life cycle, and an integrated search for more specific information. A PHP-based web interface allows users to browse and query SjTPdb, and moreover to switch to external databases by the following embedded links.

**Conclusion:** SjTPdb is the first schistosome database with detailed annotations for schistosome proteins. It is also the first integrated database of both transcriptome and proteome of *S. japonicum*, providing a comprehensive data resource and research platform to facilitate functional genomics of schistosome. SjTPdb is available from URL: <http://function.chgc.sh.cn/sj-proteome/index.htm>.

### Background

Schistosomiasis remains one of the most prevalent and serious parasitic diseases, and there are ~200 million patients in 76 countries and territories predominantly in

tropical and subtropical regions. Schistosomiasis is caused by the multi-cellular parasite schistosomes, which include three major species – *Schistosoma japonicum*, *S. mansoni*, and *S. haematobium* [1]. *S. japonicum* is endemic

in China, the Philippines and some other sites in East Asia. Schistosomes have a complex life cycle, including free-living life stages (cercaria and miracidium) and life stages parasitizing in the snail hosts (sporocyst) or definitive mammalian hosts (egg, schistosomulum and adult). The eggs, produced by adult females that are deposited in the liver, intestines, and other host organs, are the major contributors to the pathology and morbidity associated with schistosomiasis. However, schistosome biology, host-parasite interactions and molecular mechanisms of immunopathology of schistosomiasis are not well understood. Comprehensive genomic, transcriptomic, and proteomic analyses will shed light on these aspects and facilitate the development of novel intervention strategies for the control and treatment of schistosomiasis.

In transcriptomic analysis, expressed sequence tags (ESTs) are useful resources for cataloguing expressed genes. For schistosomes, more than 43,000 and 163,000 ESTs from various life stages of *S. japonicum* [1] and *S. mansoni* [2,3] have been acquired and analysed by our group and others, representing the first gene-discovery program and an initial step towards sequencing the complete genome sequence of this parasite. We identified and characterized 611 *S. japonicum* EST clusters with complete open reading frames (ORFs) in our earlier study [1]. Although ESTs are useful resources for monitoring gene expression, they represent the stretches or fragments of transcripts and usually cover only part of full-length genes. Furthermore, complementary DNAs (cDNAs) are limited in providing expression features, because they do not indicate the subcellular localization and post-translational modifications of proteins. On the other hand, proteomic strategies represent a feasible method to monitor protein profiles and complement transcriptomic strategies. Recently, our group reported an integrated comprehensive transcriptomic and proteomic survey of *S. japonicum* [4]. In that study, we identified ~15,000 EST clusters and 8,420 protein-coding genes including ~3,000 transcripts with entire ORFs. The transcriptomic data of *S. japonicum* were collected from schistosomulum, adult worm (including mix-sex adult worm, male worm and female worm), egg and miracidium (minor EST data) [4]. Moreover, we verified the expression states of ~3,200 genes by proteomics throughout different life stages of *S. japonicum* (all developmental stages except for the sporocyst), tegument samples from mix-sex adult worm, male worm, female worm and schistosomulum, and eggshell. The findings represent a comprehensive transcriptomic/proteomic view of *S. japonicum* and should lead to a more profound understanding of schistosome biology and the host-parasite relationship [4].

However, it should be pointed out that other transcriptomic approaches, such as oligonucleotide microarray,

were also used in *S. japonicum* and *S. mansoni* studies [5-11]. Several aspects of schistosome biology have been investigated by these approaches, including differential expressions of genes across different life cycles in *S. mansoni* [5,6] and *S. japonicum* [7], gender-specific expression in *S. mansoni* [8,9] and *S. japonicum* [10], as well as species-specific transcription [11]. These studies provide comprehensive information on growth, development, sex differentiation and maturation of this pathogen. Furthermore, in recent studies, the transcriptome of adult worm of *S. mansoni* has been investigated by serial analysis of gene expression (SAGE) technology [12,13]. Indeed, the report by McKerrow and co-workers perhaps represents the most comprehensive survey to date on the transcriptome of the developmental stages of *S. mansoni* including cercariae, juvenile liver-stage worms, adult worms, miracidia, mother sporocysts and eggs [13].

Recently, the *S. japonicum* genome has been sequenced by the Chinese National Human Genome Center at Shanghai, and the sequence datasets, including assembled supercontigs and predicted genes, are publicly available [14]. The sequencing of *S. mansoni* genome was also performed by a genome sequencing consortium of The Institute for Genomic Research (TIGR) and the Wellcome Trust Sanger Institute [15], with sequence data available from their website [16]. The latest assembled version 3.1 of the *S. mansoni* genome has approximately 19,000 supercontigs and ~13,000 predicted full-length genes [17,18]. These genomic data of *S. japonicum* and *S. mansoni* certainly represents a new invaluable resource. However, we feel that for best use in the future of these data the massive information from *S. japonicum* needs to properly organized and curated to enable its most efficient access and investigation. The aim of the present study was to establish a comprehensive database dedicated to *S. japonicum*, encompassing data derived from sequence analysis as well as links to external online resources. The database integrates transcriptomic and proteomic resources of *S. japonicum*, as well as shortcuts linking to online genomic and transcriptomic datasets of *S. japonicum* and *S. mansoni*. We consider that SjTPdb will represent a valuable tool and provide more effective access to a wealth on information on schistosome development, evolution and host-parasite interplay.

## Construction and content

### Data collection

SjTPdb contains (1) 84,499 ESTs of several, discrete developmental stages of *S. japonicum*; (2) 14,962 EST clusters; (3) 8,420 protein-coding genes [4], where these *S. japonicum* genes were annotated in detail in SjTPdb; (4) the proteomic dataset of *S. japonicum*. The draft genomic data of *S. japonicum*, including contig sequences and supercontig sequences, the EST clusters [3] and predicted genes

derived from draft genomic data (version 4.0) of *S. mansoni* [16], are publicly available through their website gateways [14].

### EST analysis

All ESTs from *S. japonicum* were first filtered to remove low quality sequences. The remaining ESTs were masked at the both ends for contaminating regions, including vector sequences, adaptor, poly (A/T) and restriction sites. The masked regions of ESTs were changed, not trimmed, into the equal length of poly "N", as this processing would ensure that the length of EST sequences remains the same as their corresponding quality files, so that users can check conveniently the quality of specific regions of selected ESTs. The processes of EST assembly, protein coding sequence (CDS) prediction and the primary annotation have been reported previously [4]. Finally, the available 14,962 EST clusters contain 10,434 contigs and 4,528 singletons, where 8,420 were considered as CDS-containing.

### Annotation methods

The functional annotations of these CDS-containing EST clusters were mainly based on sequence similarity to known genes in public databases [4]. The 8,420 translated peptides were compared by BLASTP with sequences from UniProtKB/Swiss-Prot Release 55.0 [19] and the non-redundant (NR) database from NCBI [20] (downloaded on March-12-2008), respectively. In general, for a given query from these *S. japonicum* proteins or peptides, its BLASTP hit with the lowest E value in the both databases must be accepted for functional annotation. The public databases contain *S. japonicum* protein sequences submitted previously by us. Therefore, to avoid the self-annotations, we discarded all self BLASTP hits and chose the next best BLASTP hit for annotation. However, if the BLAST hit is to a sequence of uncertain function - indicated by descriptions like "hypothetical", "unknown" and "uncharacterized", a meaningful hit was selected. We also performed BLASTX searches using EST clusters against SwissProt database in addition to BLASTP searches. We found a few EST clusters had better hits in reverse strand, suggesting an error in ORF prediction. In these cases, we reversed the sequence and chose the correct ORF according to BLASTX results. We also found some ORFs with apparent frame-shift or truncation mutations involving single nucleotide insertion or deletion. Moreover, we found some EST clusters with hidden intron sequences that usually introduce stop codons. In these cases, the EST clusters were re-assembled and checked manually to eliminate errors. After obtaining the correct protein sequences, we performed BLASTP searches using these new sequences against NR and SwissProt databases and annotated these proteins as above.

The *S. japonicum* protein will be annotated as "hypothetical protein" if its BLASTP E value is higher than E-5. All of the *S. japonicum* genes were annotated as "SJCHGCxxxxx protein, xxx aa, BLASTP (against SwissProt/NR) similarity to xxx, E = xxx, Identities = xxx" or "SJCHGCxxxxx, xxx aa, hypothetical protein".

The functional classification of *S. japonicum* proteins was performed using terms from the Molecular Function and Biological Process aspects of the Gene ontology (GO) system [21]. To examine homologous protein domains, the peptide sequences of *S. japonicum* were used to search the InterPro database [22] by software InterProScan [23], the Pfam database [24] by HMMER 2.4i [25], and the PROSITE database [26] by ScanProsite [27], respectively. Potential signal peptides of *S. japonicum* proteins with full length CDS were identified by SignalP 3.0 [28,29]. Transmembrane helices were predicted with the TMHMM 2.0 server [30] and HMMTOP 2.0 [31]. Putative subcellular localizations of *S. japonicum* proteins were identified by TargetP 1.1 [32] and PSORT II [33].

In the present report, expression levels of *S. japonicum* proteins across different developmental stages, sexes, and within tegument and eggshell were calculated based on our EST and proteomic data, as previously reported [1,3]. In the future, we plan to re-calculate expression levels based on *S. japonicum* gene microarray analyses. To address this issue, we will (1) integrate microarray data reported by others [7,10,11], (2) perform a comprehensive microarray study covering most life stages of *S. japonicum*, and (3) accept submitted microarray data. The expression level calculations will be updated at least once per year to reflect the advances in the study of *S. japonicum* molecular genetics. Genetic variations, including single nucleotide polymorphism (SNP), deletion and insertion (INDEL) of *S. japonicum* genes, and associated and microsatellites have been described previously [4], and in like fashion we will aim to include new information on these phenomena as it becomes available.

### Database implementation

SjTPdb was built on the Windows XP operating system, and the data are parsed into a MySQL database that is accessible publicly via the Apache web server. The PHP and Perl programming languages are used to produce dynamic web pages in response to users' queries. A local BLAST program is incorporated into SjTPdb, including BLASTP, BLASTN, BLASTX, TBLASTN and TBLASTX. A pull-down menu for database selection is included, with a "NR/NT" option where users can perform BLAST searches against the current non-redundant (NR) protein database or the nucleotide sequence database of NR (NT) of GenBank. When the option is selected, the web page will be transferred automatically to the BLAST interface in NCBI.

In addition, we have embedded links in SjTPdb for BLAST searching against *S. mansoni* EST data and genomic sequences of *S. japonicum* and *S. mansoni*.

## Utility and discussion

### Database overview

*S. japonicum* is a multi-cellular parasite of mammals including humans and also represents a model member of the phylum Platyhelminthes. Previous comprehensive 'omics surveys have produced valuable information for schistosome biology [4]. In SjTPdb we have integrated different types of expression data of *S. japonicum* including transcriptomic and proteomic profiles across the developmental stages of this flatworm. Although the ESTs and some *S. japonicum* gene sequences are accessible in public databases, there is only minimal biological annotation available. Herein, SjTPdb provides the biological annotation for these nucleotide sequences in much more detail, which will facilitate further data mining. This can be anticipated to lead to better understanding of schistosome biology.

An important feature of SjTPdb is the incorporation of proteomic data. The proteomic data contains the expression profiles of cercaria, miracidium, tegument and egg-shell that complements transcriptomic data [1,3]. It is the first genomic database of schistosome that integrates both transcriptomic and proteomic datasets.

The schema of SjTPdb is shown in Figure 1.

### Annotated pages for *S. japonicum* genes/proteins

In SjTPdb, each gene sequence entry of *S. japonicum* encoding proteins is well-annotated. A typical annotation page contains the following information:

**Entry information:** basic entry information such as protein name, accession number(s) (protein) and gene identifier (gi) with a clickable link to GenBank.

**Name and origin of the protein:** this section lists a brief biological annotation of this protein based on the similarity to known proteins, as well as the mRNA accession number, CDS region and taxonomy. For a given annotation, the details of BLAST similarity are shown in this section as an indication of the confidence evaluation of the annotated gene. Users can check the detailed BLAST result files via the links. The similarity comparison, however, could become outdated when the public database is updated, leading to false annotations. To re-evaluate the similarities, we offer users five clickable BLAST links, which under clicking will submit the protein or DNA sequence automatically to BLAST engines in NCBI (Figure 2A) and users can perform BLAST searches against the latest public databases.

**References:** citations of available publications reporting the generation and analysis of the *S. japonicum* gene/protein are provided.

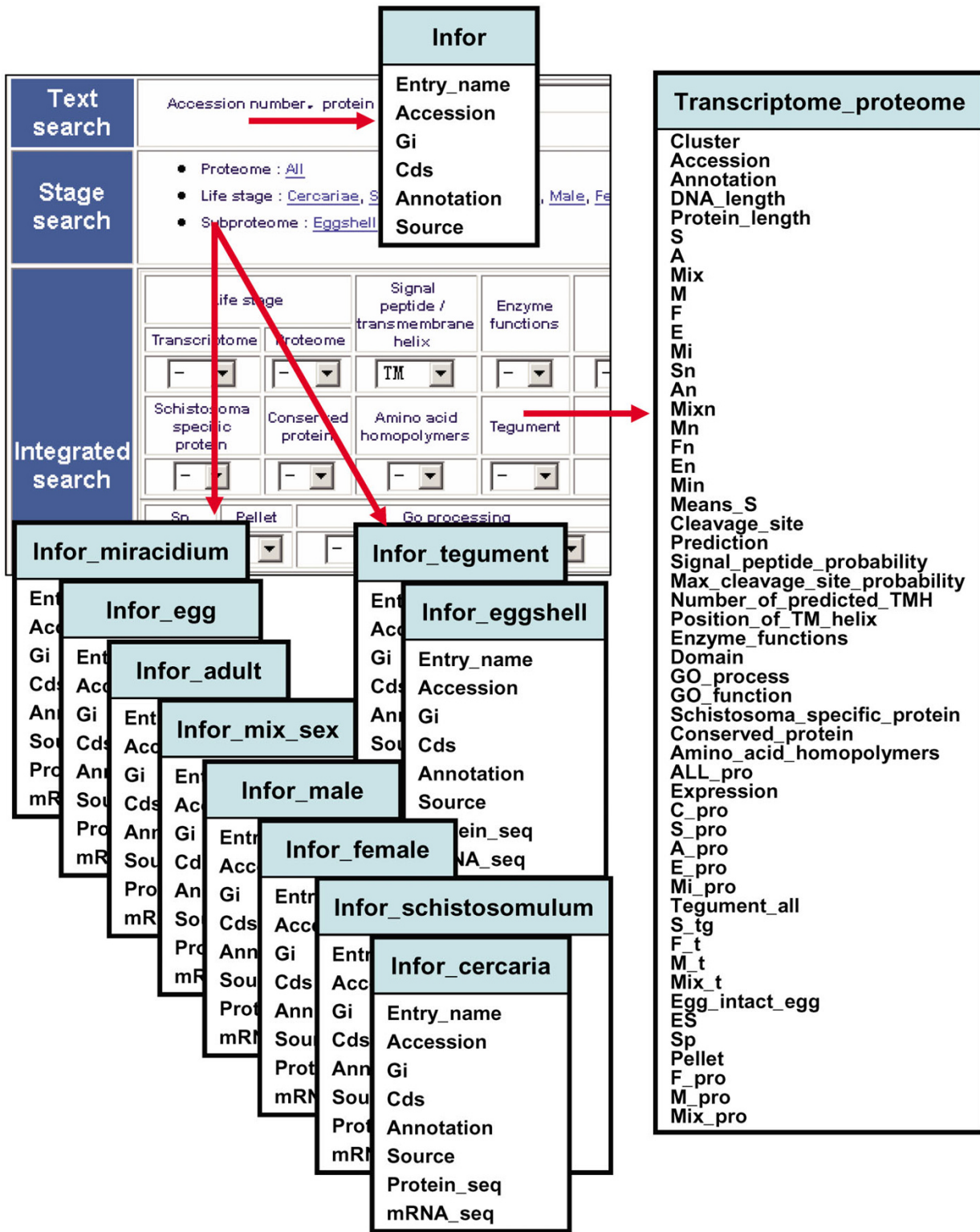
**Gene Ontology:** functional categories by Gene Ontology associations are listed here if assigned. Again, users can re-analyze the GO assignments by clicking the quick link embedded in the page.

**Family and domain:** this section records conserved protein motifs or domains of the *S. japonicum* protein, obtained by comparing with Interpro, Pfam and Prosite databases.

**Molecular characteristics:** this section provides information on predictions on whether the *S. japonicum* protein is secreted, anchored to the membrane, or located in other subcellular organelles, based on predictions by SignalP, TMHMM, HMMTOP, TargetP and PSORT programs using artificial neural networks, hidden Markov methods (HMM), k-nearest neighbours classifier [34] or other methods. Secreted proteins released by *S. japonicum* may function as enzymes in invasion [35] or feeding process [36], or as immune modulators [37], while the membrane proteins may contribute to signal transduction, immune evasion, antigen mimic, and host-parasite interaction [38]. For brevity, only the summarized prediction result is displayed; detailed descriptions can be obtained through clickable internal links.

**Genetic polymorphisms:** genetic diversity of schistosome populations, determined as SNPs, INDELS and microsatellite allelism may lead to the variations in infectivity, drug sensitivity, pathogenicity, immunogenicity, host species range and so on. Genetic diversity is apparent among 1,496 of 8,420 *S. japonicum* genes, involving at least 6,038 known SNPs, with 2,272 synonymous and 1,552 nonsynonymous substitutions found in the CDS regions [4]. Furthermore, a small subset nonsynonymous SNPs were verified by mass spectrometry analysis [4]. The genetic polymorphisms could significantly confound the host immune responses targeting these antigens, which will need to be considered during development of new vaccines and drug targets.

**Expression:** schistosomes are dioecious and the sexually mature adults display major dimorphism, with distinct male and female forms. Investigations on gender-enriched expression can be expected to elucidate mechanisms of development, sexual maturation and egg production in schistosomes [39,40]. To obtain the expression differences between sexes, we calculated the expression levels of mixed-sex, male and female adult worms based on both transcriptomic and proteomic analyses (Figure 2B). Transcript expression levels were calculated using relevant EST numbers, while the protein abundance were

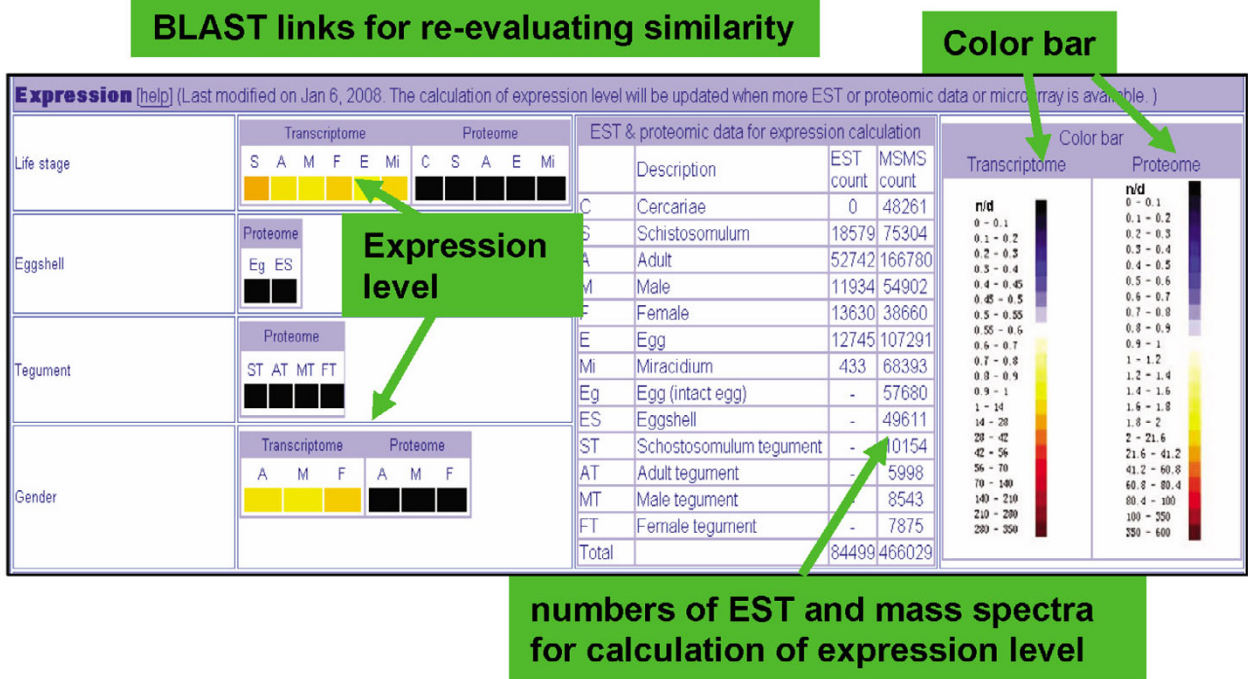


**Figure 1**  
**Schema of SjtPdb, an integrated transcriptome and proteome database of *S. japonicum*.** The field names of each MySQL databases in SjtPdb are shown. The database "infor" and "transcriptome and proteome" can be queried through text search and integrated search, respectively. The other databases can be queried by stage search.

A

Entry information [help]	
Entry name	SJCHGC03237
Accession number	<a href="#">AAW26004.1</a>
GI	<a href="#">56755651</a>
Entry was last modified on	April-01-2008
Name and origin of the protein	
Protein annotation	SJCHGC03237 protein, 224 aa, BLASTP (against NR) similarity to chaperonin containing TCP1, subunit 2 (beta) [Xenopus tropicalis], Expect=4E-82, Identities=159/224 (70%). (Check the BLASTP results: BLASTP against <a href="#">Swissprot v55.0</a> or <a href="#">NR (20080312)</a> )
Re-analyze Similarity [help]	<a href="#">Blastp</a> ; <a href="#">Blastn</a> ; <a href="#">Blastx</a> ; <a href="#">tBlastn</a> ; <a href="#">tBlastx</a> (Click the links and perform BLAST searches against last databases at NCBI.)
mRNA	<a href="#">AY814272.1</a> CDS 14,688
Source	<a href="#">Schistosoma japonicum</a>
Taxonomy	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma

B



**Figure 2**

**The annotation page of a *S. japonicum* protein.** A, The head of annotation page for a sample protein in SjTPdb, showing five BLAST links for re-analysis of similarity comparison of this protein against external public databases. B, The expression patterns of a *S. japonicum* protein across different life cycle stages, gender and tegument or eggshell. The expression levels are represented by changing colors according to the legend of the color bar (right).

represented by the ratio of the sum of all Mascot peptide scores for the protein to the total number of tandem mass spectra of the protein [4,41]. The results were transformed into colors for intuitive visualization, with the legend displayed in the protein pages.

The schistosomes have complex developmental cycles that have evolved to include adaptation to free-swimming aquatic environment, intermediate host snail in an

aquatic environment and warm-blooded mammalian host environment. Characterization of expression patterns across these life stages of *S. japonicum* will provide further insight into schistosome biology, molecular mechanisms of immunopathology of schistosomiasis and host-parasite interplay [6,42]. Here, we further calculated and displayed the expression patterns of the *S. japonicum* cercaria, schistosomulum, egg and miracidium with the integrated information from both transcriptomic and

proteomic datasets (Figure 2B). While the transcriptome expression profile of the cercaria was not included because the ESTs of this life stage were not collected, the expression pattern of cercariae is represented here at least by proteomic data.

The outer covering of the schistosome, the tegument, is a living syncytial layer; it is evident on the schistosomulum and the adult, stages that inhabit the human blood vessels. The tegument plays a pivotal role in nutrient uptake [43-45] and locates at the interface of the parasite and its host. Thus tegumental proteins represent potential drug targets and vaccine candidates. Here, we presented the expression patterns of predicted tegumental proteins on each annotation page based on our earlier studies of the protein expression profiles in schistosomula, mixed-sex adults, and male and female adults of *S. japonicum* [4].

The egg of schistosome is covered by a sclero-proteinaceous eggshell within which the miracidium develops [46]. Eggshell proteins may contribute to initiation of host immune response [47]. For each eggshell protein, we present its eggshell expression pattern on the annotation page, with its expression in intact eggs shown for comparison.

*Sequence information:* this section shows both the protein and the encoding DNA sequence of a given gene/protein, which are downloadable as compressed files. Shown on the top of the sequences are the length, deduced molecular weight (MW) and isoelectric point (pI) of the protein.

#### Querying the database

SjTPdb offers users four types of web-based tools to query and download information from the database – text-based, life stage, integrated, and BLAST search programs. The text-based tool is used to retrieve the annotation pages for a given gene/protein by keywords such as protein name, accession number, EST cluster name, or any words that exist in the annotation line (Figure 3A). The query result can be displayed in a table form with each hit listed in a row. However, because the output page is configurable, users can set the display options to show the mRNA or protein sequences in fasta format. Also, users can specify the number of rows for display when the querying result contains more than 20 rows (default number). The output table includes six columns that provide links to the relevant gene or protein annotation pages in SjTPdb, in addition to outer links to related proteins and mRNAs deposited in GenBank. Furthermore, the query result and the related protein or DNA sequences are downloadable as plain tables and fasta files (Figure 3B), respectively. Users can also choose to download part of the result by checking the boxes in the front of each row.

The life stage search provides users with a shortcut to retrieve the specialized expression profiles of a given developmental life stage of *S. japonicum*, as represented in the proteomics datasets [4]. This search also offers direct access to profiling differential gender-associated proteins and sub-proteomes of the parasite's eggshell and tegument.

Unlike text-based or life stage searches, the integrated search provides a convenient and flexible way to retrieve specific information on particular topics of *S. japonicum* by employing multiple constrains. For example, when "TM" option is selected within "Signal peptide/transmembrane helix" section, the output will display the records of *S. japonicum* proteins with putative transmembrane domains. When both "TM" and "ES" options within "Eggshell" section are checked, the query result will exhibit the putative transmembrane proteins detected in eggshell samples. Similarly, users can retrieve a specific subset of information of *S. japonicum* by applying more constrains. This function is expected to be especially useful for in-depth data-mining and to obtain more fundamental insights into schistosome biology. Furthermore, the output format of the query result by this search tool has more detailed information than that of the text-based and life stage searches. Here, the output tables and the related DNA or protein sequences are downloadable freely.

BLAST search tools are powerful and used widely for sequence similarity comparison. We implemented BLAST algorithms in SjTPdb (Figure 4). Herein, the searchable sequences in SjTPdb include EST/EST cluster sequences, protein sequences deduced from EST clusters and the predicted genes of *S. japonicum* and *S. mansoni*. If users fail to locate significant hits in schistosome databases, they can search against the more comprehensive NR or NT databases (Figure 4).

All *S. japonicum* sequences in SjTPdb are freely downloadable, along with the EST sequences of *S. japonicum* from the discrete developmental stages and the corresponding quality files of EST reads. The information also includes names of ESTs incorporated into each 14,962 EST cluster and tables with the association links between EST accession numbers and EST clones.

#### Future prospects

The databases will be regularly updated to reflect future progress in genome, transcriptome and proteome analyses of *S. japonicum* and other schistosomes. The expression patterns of genes across developmental stages of *S. japonicum* will be reconstructed using microarray data of published studies or ongoing projects. We are also planning to launch a comprehensive SNP scanning analysis of *S. japonicum* genome. New data will be incorporated into the

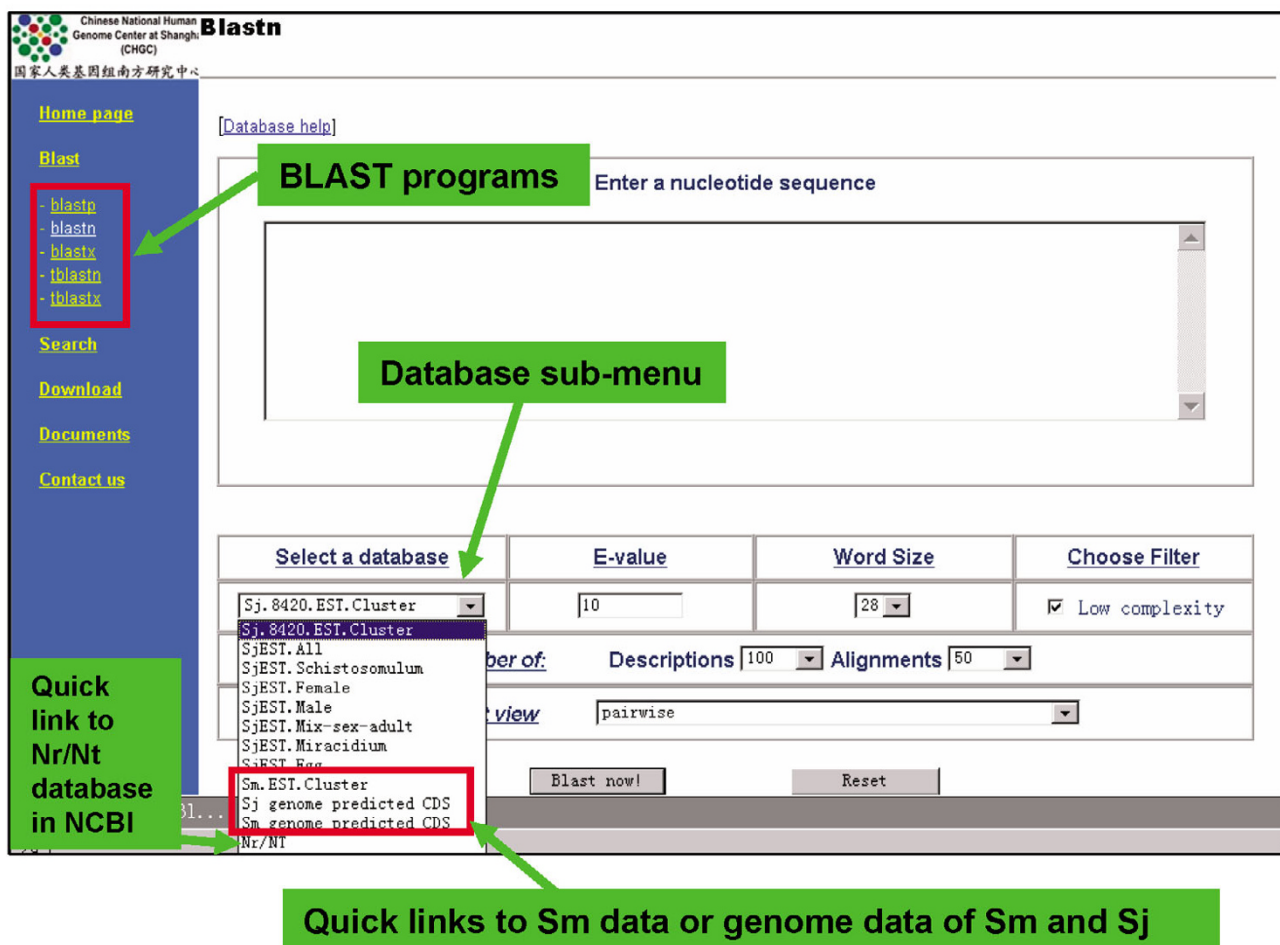
A

B

Entry	ACCESSION	GI	mRNA	ANNOTATION	SOURCE
<input type="checkbox"/> <a href="#">SJCHGC00146</a>	<a href="#">AAW25871.1</a>	<a href="#">56755383</a>	<a href="#">AY814139.1</a>	<a href="#">SJCHGC00146 protein, similar to NM_022756 hypothetical protein FLJ11730 in Homo sapiens; hypothetical protein FLJ11730 in Homo sapiens [Schistosoma japonicum]</a>	<a href="#">Schistosoma japonicum</a>
<input type="checkbox"/> <a href="#">SJCHGC00176</a>	<a href="#">AAW25156.1</a>	<a href="#">56753927</a>	<a href="#">AY813424.1</a>	<a href="#">SJCHGC00176 protein, similar to CG11624-PA (Drosophila melanogaster)</a>	<a href="#">Schistosoma japonicum</a>

**Figure 3**  
**Search interface and the search results.** A, The search interface of SjTPdb is illustrated, showing search tools by the text, life cycle developmental stage and integrated selections. The red arrows indicate the selections of a combined search for the putative secreted proteins with the detectable transcripts in the schistosomulum stage of the parasite. B, an example of a query result, with options for downloading and re-phrasing.





**Figure 4**  
**BLAST search tools implemented in SjtPdb.** We implemented five kinds of BALST programs in this database. To perform similarity comparison, users can choose various sequence databases of schistosome and external links to NR/NT databases at NCBI within the database sub-menu.

SjtPdb, providing an even more comprehensive data-mining platform for investigation on schistosome biology and, hopefully, to facilitate development of new interventions for schistosomiasis.

**Conclusion**

In SjtPdb, we have integrated comprehensive transcriptomic and proteomic datasets of a human blood fluke, *S. japonicum*, and implemented convenient tools for querying and data retrieving. We consider that SjtPdb represents an important contribution to the genomics of *S. japonicum*, and we expect it will facilitate fundamental research on schistosome evolution, development, host-parasite interplay and so forth.

**Availability and requirements**

The database is located at <http://function.chgc.sh.cn/sj-proteome/index.htm> and is suitable for most graphical web browsers. All sequences can be freely downloaded from SjtPdb (see Download section).

**Authors' contributions**

FL conceived the study, designed and optimized the website, and drafted the manuscript. PC constructed the database and performed the computational analysis. S-JC and Z-QW contributed data to the database. Z-GH supervised the work and revised the manuscript.

**Acknowledgements**

We warmly thank Paul J. Brindley of George Washington University Medical Center and Dr. Guo-An Zhang of New York University School of Medicine for advice and editorial assistance. We also thank Dr. Jin-Yan Huang and Zhi-Wei Cao for their important suggestions. This work was sup-

ported by grants from the Chinese High-Tech Research and Development Program (863) (2006AA02Z318), Chinese National Key Program on Basic Research (973) (2006CB708510 and 2007CB513106), National Foundation for Excellence Doctoral Project, Shanghai Commission for Science and Technology (06JC14059 and 05DZ22201), China Postdoctoral Science Foundation and Shanghai Rising-Star Program (07QA14043).

## References

- Hu W, Yan Q, Shen DK, Liu F, Zhu ZD, Song HD, Xu XR, Wang ZJ, Rong YP, Zeng LC, Wu J, Zhang X, Wang JJ, Xu XN, Wang SY, Fu G, Zhang XL, Wang ZQ, Brindley PJ, McManus DP, Xue CL, Feng Z, Chen Z, Han ZG: **Evolutionary and biomedical implications of a *Schistosoma japonicum* complementary DNA resource.** *Nat Genet* 2003, **35**:139-147.
- Verjovski-Almeida S, DeMarco R, Martins EA, Guimarães PE, Ojopi EP, Paquola AC, Piazza JP, Nishiyama MY Jr, Kitajima JP, Adamson RE, Ashton PD, Bonaldo MF, Coulson PS, Dillon GP, Farias LP, Gregorio SP, Ho PL, Leite RA, Malaquias LC, Marques RC, Miyasato PA, Nascimento AL, Ohlweiler FP, Reis EM, Ribeiro MA, Sá RG, Stukart GC, Soares MB, Gargioni C, Kawano T, Rodrigues V, Madeira AM, Wilson RA, Menck CF, Setubal JC, Leite LC, Dias-Neto E: **Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*.** *Nat Genet* 2003, **35**:148-157.
- Schistosoma mansoni* EST Genome Project** 2003 [<http://bio.info.iq.usp.br/schisto/>].
- Liu F, Lu J, Hu W, Wang SY, Cui SJ, Chi M, Yan Q, Wang XR, Song HD, Xu XN, Wang JJ, Zhang XL, Zhang X, Wang ZQ, Xue CL, Brindley PJ, McManus DP, Yang PY, Feng Z, Chen Z, Han ZG: **New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*.** *PLoS Pathog* 2006, **2**:e29.
- Verjovski-Almeida S, Venancio TM, Oliveira KC, Almeida GT, Demarco R: **Use of a 44k oligoarray to explore the transcriptome of *Schistosoma mansoni* adult worms.** *Exp Parasitol* 2007, **117**(3):236-245.
- Jolly ER, Chin CS, Miller S, Bahgat MM, Lim KC, DeRisi J, McKerrow JH: **Gene expression patterns during adaptation of a helminth parasite to different environmental niches.** *Genome Biol* 2007, **8**(4):R65.
- Chai M, McManus DP, McInnes R, Moertel L, Tran M, Loukas A, Jones MK, Gobert GN: **Transcriptome profiling of lung schistosomula, in vitro cultured schistosomula and adult *Schistosoma japonicum*.** *Cell Mol Life Sci* 2006, **63**(7-8):919-929.
- Fitzpatrick JM, Johnston DA, Williams GW, Williams DJ, Freeman TC, Dunne DW, Hoffmann KF: **An oligonucleotide microarray for transcriptome analysis of *Schistosoma mansoni* and its application/use to investigate gender-associated gene expression.** *Mol Biochem Parasitol* 2005, **141**(1):1-13.
- Francis M, Lobo FP, Cerqueira GC, Passos LK, Carvalho OS, Franco GR, El-Sayed NM: **Microarray analysis of gene expression induced by sexual contact in *Schistosoma mansoni*.** *BMC Genomics* 2007, **8**:181.
- Moertel L, McManus DP, Piva TJ, Young L, McInnes RL, Gobert GN: **Oligonucleotide microarray analysis of strain- and gender-associated gene expression in the human blood fluke, *Schistosoma japonicum*.** *Mol Cell Probes* 2006, **20**(5):280-289.
- Gobert GN, McInnes R, Moertel L, Nelson C, Jones MK, Hu W, McManus DP: **Transcriptomics tool for the human *Schistosoma* blood flukes using microarray gene expression profiling.** *Exp Parasitol* 2006, **114**(3):160-172.
- Ojopi EP, Oliveira PS, Nunes DN, Paquola A, DeMarco R, Gregório SP, Aires KA, Menck CF, Leite LC, Verjovski-Almeida S, Dias-Neto E: **A quantitative view of the transcriptome of *Schistosoma mansoni* adult-worms using SAGE.** *BMC Genomics* 2007, **8**:186.
- Williams DL, Sayed AA, Bernier J, Birkeland SR, Cipriano MJ, Papa AR, McArthur AG, Taft A, Vermeire JJ, Yoshino TP: **Profiling *Schistosoma mansoni* development using serial analysis of gene expression (SAGE).** *Exp Parasitol* 2007, **117**(3):246-258.
- Schistosoma* subject database at Shanghai Center for Life Science & Biotechnology information** 2006 [<http://life.center.sgst.cn/en/schistosomaDispatch.do?disName=intro>].
- El-Sayed NM, Bartholomeu D, Ivens A, Johnston DA, LoVerde PT: **Advances in schistosome genomics.** *Trends Parasitol* 2004, **20**:154-157.
- The *Schistosoma mansoni* Genome Project** 2006 [[http://www.sanger.ac.uk/Projects/S\\_mansoni/](http://www.sanger.ac.uk/Projects/S_mansoni/)].
- Wilson RA, Ashton PD, Braschi S, Dillon GP, Berriman M, Ivens A: **'Omning in on schistosomes: prospects and limitations for post-genomics.** *Trends Parasitol* 2007, **23**:14-20.
- Haas BJ, Berriman M, Hirai H, Cerqueira CG, Loverde PT, El-Sayed NM: ***Schistosoma mansoni* genome: closing in on a final gene set.** *Exp Parasitol* 2007, **117**(3):225-228.
- UniProtKB/Swiss-Prot Release 55.0** [[ftp://ftp.expasy.org/data\\_bases/uniprot/current\\_release/knowledgebase](ftp://ftp.expasy.org/data_bases/uniprot/current_release/knowledgebase)]
- Non-redundant (NR) database from NCBI** [<ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz>]
- Gene Ontology Consortium: **The Gene Ontology (GO) project in 2006.** *Nucleic Acids Res* 2006, **34**:D322-D326.
- Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Bullard V, Cerutti L, Copley R, Courcelle E, Das U, Daugherty L, Dibley M, Finn R, Fleischmann W, Gough J, Haft D, Hulo N, Hunter S, Kahn D, Kanapin A, Kejariwal A, Labarga A, Langendijk-Genevaux PS, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Nikolskaya AN, Orchard S, Orengo C, Petryszak R, Selengert JD, Sigrist CJ, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C: **New developments in the InterPro database.** *Nucleic Acids Res* 2007, **35**:D224-D228.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier.** *Nucleic Acids Research* 2005, **33**:W116-W120.
- Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A: **Pfam: clans, web tools and services.** *Nucleic Acids Res* 2006, **34**:D247-D251.
- Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14**:755-763.
- Hulo N, Bairoch A, Bullard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJ: **The PROSITE database.** *Nucleic Acids Res* 2006, **34**:D227-D230.
- de Castro E, Sigrist CJ, Gattiker A, Bullard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N: **ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.** *Nucleic Acids Res* 2006, **34**:W362-W365.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *J Mol Biol* 2004, **340**:783-795.
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Protein Eng* 1997, **10**:1-6.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**:567-80.
- Tusnády GE, Simon I: **The HMMTOP transmembrane topology prediction server.** *Bioinformatics* 2001, **17**:849-850.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, **300**:1005-1016.
- Nakai K, Horton P: **PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization.** *Trends Biochem Sci* 1999, **24**:34-35.
- Horton P, Nakai K: **Better prediction of protein cellular localization sites with the k nearest neighbors classifier.** *Proc Int Conf Intell Syst Mol Biol* 1997, **5**:147-152.
- Knudsen GM, Medzihradsky KF, Lim KC, Hansell E, McKerrow JH: **Proteomic analysis of *Schistosoma mansoni* cercarial secretions.** *Mol Cell Proteomics* 2005, **4**:1862-1875.
- Delcroix M, Sajid M, Caffrey CR, Lim KC, Dvorak J, Hsieh I, Bahgat M, Dissous C, McKerrow JH: **A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite.** *J Biol Chem* 2006, **281**:39316-39329.
- Schramm G, Gronow A, Knobloch J, Wippersteg V, Grevelding CG, Galle J, Fuller H, Stanley RG, Chiodini PL, Haas H, Doenhoff MJ: **IPSE/alpha-1: a major immunogenic component secreted from *Schistosoma mansoni* eggs.** *Mol Biochem Parasitol* 2006, **147**:9-19.
- Braschi S, Borges WC, Wilson RA: **Proteomic analysis of the schistosome tegument and its surface membranes.** *Mem Inst Oswaldo Cruz* 2006, **101**(Suppl 1):205-212.
- Hoffmann KF, Johnston DA, Dunne DW: **Identification of *Schistosoma mansoni* gender-associated gene transcripts by cDNA**

- microarray profiling.** *Genome Biol* 2002, **3**:research0041.1-research0041.12.
40. Fitzpatrick JM, Hoffmann KF: **Dioecious *Schistosoma mansoni* express divergent gene repertoires regulated by pairing.** *Int J Parasitol* 2006, **36**:1081-1089.
  41. Schirle M, Heurtier M-A, Kuster B: **Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry.** *Mol Cell Proteomics* 2003, **2**:1297-1305.
  42. Curwen RS, Ashton PD, Johnston DA, Wilson RA: **The *Schistosoma mansoni* soluble proteome: a comparison across four life-cycle stages.** *Mol Biochem Parasitol* 2004, **138**:57-66.
  43. Dalton JP, Skelly P, Halton DW: **Role of the tegument and gut in nutrient uptake by parasitic plathelminths.** *Can J Zool* 2004, **82**:211-232.
  44. Jones MK, Gobert GN, Zhang L, Sunderland P, McManus DP: **The cytoskeleton and motor proteins of human schistosomes and their roles in surface maintenance and host-parasite interactions.** *Bioessays* 2004, **26**(7):752-765.
  45. Gobert GN, Stenzel DJ, McManus DP, Jones MK: **The ultrastructural architecture of the adult *Schistosoma japonicum* tegument.** *Int J Parasitol* 2003, **33**(14):1561-1575.
  46. Ashton PD, Harrop R, Shah B, Wilson RA: **The schistosome egg: Development and secretions.** *Parasitology* 2001, **122**:329-338.
  47. Pearce EJ: **Priming of the immune response by schistosome eggs.** *Parasite Immunol* 2005, **27**:265-270.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

