

The *Schistosoma japonicum* genome reveals features of host–parasite interplay

The *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium*

Schistosoma japonicum is a parasitic flatworm that causes human schistosomiasis, which is a significant cause of morbidity in China and the Philippines. Here we present a draft genomic sequence for the worm. The genome provides a global insight into the molecular architecture and host interaction of this complex metazoan pathogen, revealing that it can exploit host nutrients, neuroendocrine hormones and signalling pathways for growth, development and maturation. Having a complex nervous system and a well-developed sensory system, *S. japonicum* can accept stimulation of the corresponding ligands as a physiological response to different environments, such as fresh water or the tissues of its intermediate and mammalian hosts. Numerous proteases, including cercarial elastase, are implicated in mammalian skin penetration and haemoglobin degradation. The genomic information will serve as a valuable platform to facilitate development of new interventions for schistosomiasis control.

Schistosomiasis is an ancient scourge of mankind, depicted graphically in papyri from Pharaonic Egypt and known from human remains over 2,000 years old from China^{1,2}. Blood-dwelling trematodes (phylum Platyhelminthes) of the genus *Schistosoma* cause this chronic and debilitating disease, which afflicts more than 200 million people in 76 tropical and subtropical countries. Morbidity is high and schistosomiasis contributes to several hundreds of thousands of deaths annually^{3–5}. Three principal species can infect humans: *Schistosoma japonicum*, *Schistosoma mansoni* and *Schistosoma haematobium*. The first of these is prevalent in the Philippines and parts of Indonesia, and is a major disease risk for 66 million people living in southern China². It remains a major public health concern in China despite over 50 years of concerted campaigns for its control^{2,6}. Approximately one million people in China, and more than 1.7 million bovines and other mammals, are currently infected². Control measures include community-based praziquantel chemotherapy, health education, improved sanitation, environmental modification and snail control. However, additional approaches, such as the development and deployment of new drugs and anti-schistosome vaccines are urgently needed to meet the prevailing challenges, which include the spectre of praziquantel-resistant parasites^{7,8}.

During their complex developmental cycle, schistosomes alternate between a mammalian host and a snail host through the medium of fresh water. After burrowing out of the snail host, free-swimming cercariae penetrate the skin of the mammalian host, travel through the blood to the liver via the lungs, and transform into schistosomula. These mature in the hepatic portal vein, mate and, in the case of *S. japonicum*, migrate to their final destination in the mesenteric venous plexus. Female worms release thousands of eggs daily, which are discharged in the faeces after a damaging passage through the intestinal wall. If they reach fresh water, eggs hatch to release free-swimming ciliated miracidia, which, guided by light and chemical stimuli, seek amphibious snails of the genus *Oncomelania*. Within the hemocoel of the snail, miracidia give rise asexually to numbers of sporocysts, in which further asexual propagation produces numerous cercariae.

Eggs deposited by adult female schistosomes embolize in the liver, intestines and other tissue sites and are the key contributors to the

pathology and associated morbidity of schistosomiasis. Notably, the highly adapted relationship between schistosomes and their snail intermediate and mammalian definitive hosts appears to involve exploitation by the parasite of host endocrine and immune signals^{9,10}. The evasion strategies that underpin avoidance of the host immune system, allowing schistosomes to survive for years despite strong host immune responses, have long interested investigators intent on development of an efficacious vaccine.

Unlike most other platyhelminths, schistosomes are dioecious. The genome is arrayed on eight pairs of chromosomes, seven pairs of autosomes and one pair of sex chromosomes. Females are the heterogametic sex (ZW); males are homogametic (ZZ)^{11,12}. No other lophotrochozoan¹³ has yet been sequenced.

Genome features and evolution

General information. The whole-genome shotgun (WGS) sequencing strategy was used to decode the 397-megabase-pair (Mb) sequences, covering most (>90%) of the *S. japonicum* genome (Supplementary Tables 1 and 2 and Supplementary Fig. 1). A total of 13,469 protein-coding genes were identified, comprising about 4% of the draft *S. japonicum* genome (Supplementary Figs 2 and 3). Of the protein-coding genes, 6,972 (52%) were mapped to categories established by the Gene Ontology project (Fig. 1a and Supplementary Fig. 4) and an orthologue relationship existed between 2,516 (19%) of them and 1,546 Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology groups (Supplementary Fig. 5). *Schistosoma japonicum* has a relatively large genome and low gene density in comparison with other invertebrates, including *Brugia malayi* (Table 1). On the basis of the outbred source of the genomic libraries, high-quality discrepancies found during assembly were used to identify 557,739 single nucleotide polymorphisms (SNPs) (Supplementary Table 3), with an average density of 1.4 SNPs per kilobase pair, and the insertion and deletion (indel) rates were much lower.

Repeat sequences. A total of 657 different repeat families/elements, constituting 159 Mb (40.1%) of the *S. japonicum* genome were revealed by comparing known repetitive sequences and using the software REPEATSCOUT (version 1.0.3)¹⁴ (Fig. 2 and Supplementary Table 4). Among them, 29 kinds of retrotransposon were found,

*Lists of participants and their affiliations appear at the end of the paper.

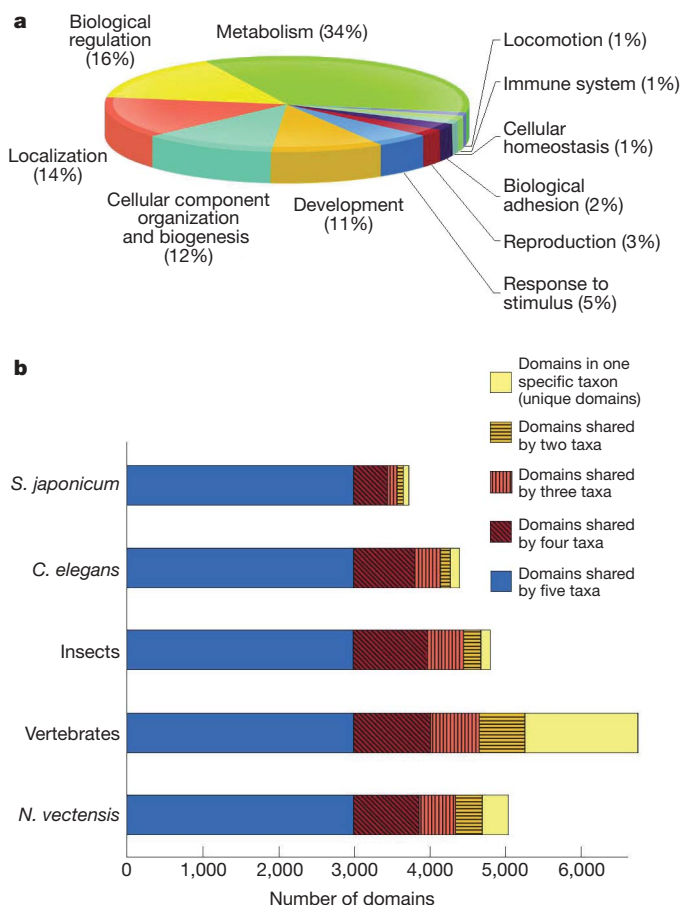


Figure 1 | Functional categorization of *S. japonicum* genes and protein-domain analysis. **a**, Proportion of the 6,972 *S. japonicum* proteins with functional information in different Gene Ontology categories. **b**, In *S. japonicum*, vertebrates (*H. sapiens*, *G. gallus* and *D. rerio*), insects (*D. melanogaster* and *A. gambiae*), *C. elegans* and *Nematostella vectensis*, a total of 7,562 domains were detected. The majority of *S. japonicum* domains are shared with other taxa, having the fewest unique domains, whereas vertebrates evolved significant numbers of unique protein domains.

including known *Gulliver*, *SjR1*, *SjR2* and *Sj-pido* elements as well as 25 novel elements, together constituting 19.8% of the genome (Supplementary Table 5). Of the 25 novel retrotransposons, 18 are long terminal repeat (LTR) forms, four are non-LTR forms and three are *Penelope*-like elements—enigmatic retroelements that retain introns¹⁵. Each type of retrotransposon was represented by one to 793 intact copies or hundreds to thousands of partial copies. The non-LTR retrotransposons have significantly higher copy numbers, constituting 12.6% of the genome.

Gene loss/duplication. It was intriguing to observe that schistosomes share more orthologues with the vertebrates (Supplementary Table 6), such as *H. sapiens* (4,324 pairs), than they do with the ecdysozoans, for example *C. elegans* (3,292), despite the Ecdysozoa and Lophotrochozoa being phylogenetically adjacent¹³. Similarly, the cnidarians and vertebrates have been shown to share more orthologue genes with each other than either does with the ecdysozoans¹⁶. One possible reason for this is that a higher evolutionary rate in the Ecdysozoa causes an apparently larger orthologue divergence, although the scenario of functional selection of orthologue patterns in the context of parasite–host interplay is also worth consideration.

To test possible consequences of parasitism at the genome level, we investigated gene family and domain variations between schistosomes and other metazoans. It is clear that there was minor variation in total numbers of protein families among *S. japonicum* (6,322) and the other species, such as *C. elegans* (6,669), *D. melanogaster* (5,184) and *H. sapiens* (6,877) (Supplementary Table 7). However, a major

reduction in number, or even the elimination, of protein domains was apparent in the *S. japonicum* genome, in that the great majority (3,654) of the 3,728 protein domains from the flatworm were shared with other species (Fig. 1b) and can thus be considered ubiquitous among metazoans, whereas 3,834 domains found in at least one of the other species were not detected in schistosomes. Of these 3,834 domains, 1,140 were shared by more than three taxa of vertebrates, insects, a nematode and sea anemones (Supplementary Fig. 6). Notably, domain-loss events seem to be more widespread in *S. japonicum* than in any other species studied so far, including *C. elegans*, a model organism well known for rapid evolutionary rates and a high degree of gene loss¹⁷. Roughly 1,000 protein domains have been abandoned by *S. japonicum*, including some involved in basic metabolic pathways and defence, implying that loss of these domains could be, at least partly, a consequence of the adoption of a parasitic way of life.

Against the background of extensive gene/domain loss, the finding of expanded gene families in schistosomes might provide clues to the requirements for a parasitic lifestyle. Among the most expanded gene families in schistosomes (Supplementary Tables 8 and 9), that encoding leishmanolysin (a major surface protease, also called gp63), a member of the metallopeptidase M8 family, has 12 putative family members in *S. japonicum*, but there is only one in human, fruit fly and nematode (*C. elegans*), and only three putative counterparts in the free-living flatworm *Schmidtea*¹⁸ (Supplementary Information). In addition to elastase (see later), leishmanolysin-like proteases may contribute to tissue invasion by schistosome cercariae¹⁹.

Development and metabolism

Cellular signalling pathways in development. To investigate regulatory networks involved in embryonic development and organogenesis, we undertook comparative genomics analysis of well-characterized signalling pathways, including those for Wnt, notch, hedgehog and transforming growth factor β (TGF- β). Notably, the *S. japonicum* genome encodes these growth factors, receptors and essential components to regulate many cellular processes during organogenesis and tissue development (Fig. 3 and Supplementary Tables 10 and 11). *Schistosoma japonicum* also encodes endogenous epidermal growth factor (EGF)-like and fibroblast growth factor (FGF)-like peptides (Fig. 3). The intact downstream cascade composed of the Ras→Raf→mitogen-activated protein kinase (MAPK) and TGF- β →SMAD signalling pathways, including FGF- and EGF-receptors, has components sharing high identity with mammalian orthologues, which implies that schistosomes, in addition to using their own pathways, can exploit host growth factors as developmental signals. Indeed, we have identified an insulin receptor with high sequence similarity with those of mammals²⁰, whereas no insulin growth factor or insulin molecules were found, further supporting the notion that schistosomes exploit key signalling pathways of their hosts for growth and metabolism.

Metabolic pathways. Analysis of the KEGG pathways assigned to metabolic process (Supplementary Table 12 and Supplementary Figs 7 and 8) indicates that *S. japonicum* can use carbohydrates as energy/carbon sources. It is unable to *de novo* synthesize fatty acids, sterols, purines, nine human essential amino acids, arginine or tyrosine (Supplementary Figs 9–11). Loss or degeneracy of fatty acid, sterol and purine synthesis pathways in schistosomes is probably a consequence of the adoption of a parasitic lifestyle; notably, the genes encoding all the key enzymes for both the *de novo* fatty acid and purine syntheses are complete in the free-living flatworm, *Schmidtea mediterranea*¹⁸ (Supplementary Information). To obtain essential lipid nutrients, the *S. japonicum* genome indeed encodes many transporters, including apolipoproteins, low-density lipoprotein receptor, scavenger receptor, fatty-acid-binding protein, ATP-binding-cassette transporters and cholesterol esterase (Supplementary Table 13), to exploit fatty acids and cholesterol from host blood and plasma.

Table 1 | Summary of *S. japonicum* genomic features in comparison with other organisms

Genome features	Lophotrochozoa		Ecdysozoa			Deuterostomia		
	<i>Schistosoma japonicum</i>	<i>Caenorhabditis elegans</i>	<i>Brugia malayi</i>	<i>Drosophila melanogaster</i>	<i>Anopheles gambiae</i>	<i>Gallus gallus</i>	<i>Danio rerio</i>	<i>Homo sapiens</i>
Total genome size (Mb)	398	100	88 [†]	169	265	1,051	1,527	3,255
Total GC content (%)	34.1	35.4	30.5 [†]	40.2	42.2	41.2	34.4	36.2
GC content in coding regions (%)	36	42.7	39.6 [†]	53.3	44.3	42.9	34.4	37.5
GC content in intron regions (%)	33.8	32.1	27.6 [†]	39.5	43.9	41.3	34.1	36.6
GC content in intergenic regions (%)	34.7	35.4	30.9 [†]	40.2	42.2	41.2	34.4	36.2
Repeat rate (%)	40.1	18.3	~15 [†]	24.7	16.6	10.8	46.6	44
Total coding size (Mb)/ratio (%)	15.9/4	24.8/25	12.9/15	21.7/13	14.8/6	25.7/2	47.9/3	35.9/1
Number of coding genes	13,469	20,077	11,515	14,144	12,527	18,107	35,321	25,077
Gene density (genes per Mb)	34	200	130	84	47	17	23	8
Average CDS size (kb)	1.18	1.23	1.12	1.54	1.18	1.42	1.36	1.43
Average gene size ± s.d. (kb)	10.5 ± 16.0	2.8 ± 3.2	2.8 ± 2.9	4.3 ± 10.6	4.6 ± 10.0	21.6 ± 46.8	19.5 ± 35.1	41.3 ± 98.0
Number of transfer RNAs	153	608 (~590*)	~233 [†]	314	450	189	2,010	129
Number of ribosomal RNAs	184	19 (~275*)	~400 [†]	161	N/A	14	N/A	675

All coding sequence (CDS)-related features were calculated from the KEGG database, version 46 (<ftp://ftp.genome.jp/pub/kegg/release/archive/kegg/46/>). *Caenorhabditis elegans* and *B. malayi* genome feature files were downloaded from Wormbase (<ftp://ftp.wormbase.org/pub/>). Other genome-feature files and genome sequences were downloaded from Ensembl (<ftp://ftp.ensembl.org/pub/release-49/>).

* The features calculated from the feature file are different from those in the reference⁵⁰.

[†] Data obtained from previous publications⁵¹.

Nervous system and neuroendocrine system

Platyhelminths possess a central nervous system with a variety of sensory structures that can transduce a wide range of stimuli, and use a neuroendocrine system to regulate growth, metabolism and homeostasis.

Neurotransmitters and receptors. We characterized a number of receptors and transporters of neurotransmitters (Supplementary Table 14) that may be required, for example, by miracidia and cercariae to navigate through water to locate new hosts and for the schistosomula and adult flukes to establish and reproduce within the human vasculature. In addition to known neurotransmitters and receptors, we have identified a receptor for octopamine (Supplementary Table 14) and two key enzymes for synthesis of octopamine (Supplementary Fig. 12).

The nervous systems of flatworms can be considered to be predominantly peptidergic²¹. We found additional putative neuropeptide receptors for opioids, galanin and melatonin. Thus, it appears that schistosomes can accept stimulation of the corresponding ligands as a physiological response to different environments, such as fresh water or the tissues of their snail and mammalian hosts. There are genes encoding receptors predicted to accept gastrointestinal neuropeptide hormone signals including cholecystokinin, secretin, gastric inhibitory polypeptide and xenin, all of which are involved in functions promoting the release of alimentary tract fluids containing digestive enzymes.

We also identified receptors for urotensin, angiotensin II and neuromedin (types U and B), which have an important role in physiological regulation of the cardiovascular system, the hypothalamus and other vertebrate organs. Although schistosomes do not have these organs, these components could have other effects on the cells or tissues of the blood fluke, such as the regulation of cell growth or in tissue remodelling. In addition, we found receptors for hypocretin (orexin), leptin and hypothalamic neuropeptides. Together, these features suggest that schistosomes have many advanced physiological features regarded as more characteristic of higher metazoans.

Unexpectedly, a myokinin-like receptor was also observed (Supplementary Table 14). Myokinins are invertebrate neuropeptides with myotropic and diuretic activities for which a receptor, called lymnokinin receptor, was first identified in the tick *Boophilus microplus*²². The discovery of such a receptor in schistosomes supports the notion that they might synthesize myokinin because their vertebrate hosts do not produce this neuropeptide. Additional examples of receptors found for other invertebrate neuropeptides included FMRFamide and myosuppressin^{23,24}, both belonging to the FMRFamide-like peptide superfamily.

Complex sensory system. Schistosomes have a variety of sensory structures using which they, during their different life stages, presumably respond to a myriad of environmental stimuli. Free-living cercariae and miracidia can sense light, mechanical stimuli and temperature²⁵, facilitating the finding of hosts, whereas the parasitic adult

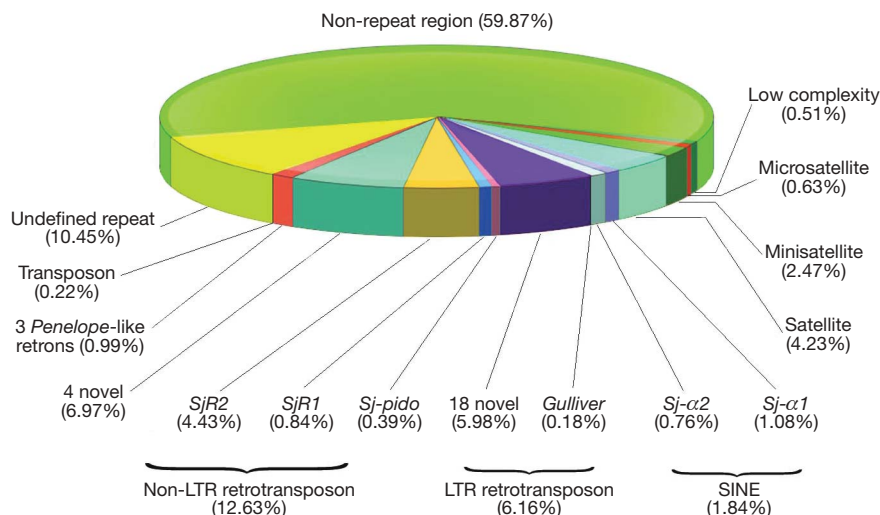


Figure 2 | The distribution of categories and composition of repeat elements in the *S. japonicum* genome. Retrons, retrotransposons; SINE, short interspersed nuclear element.

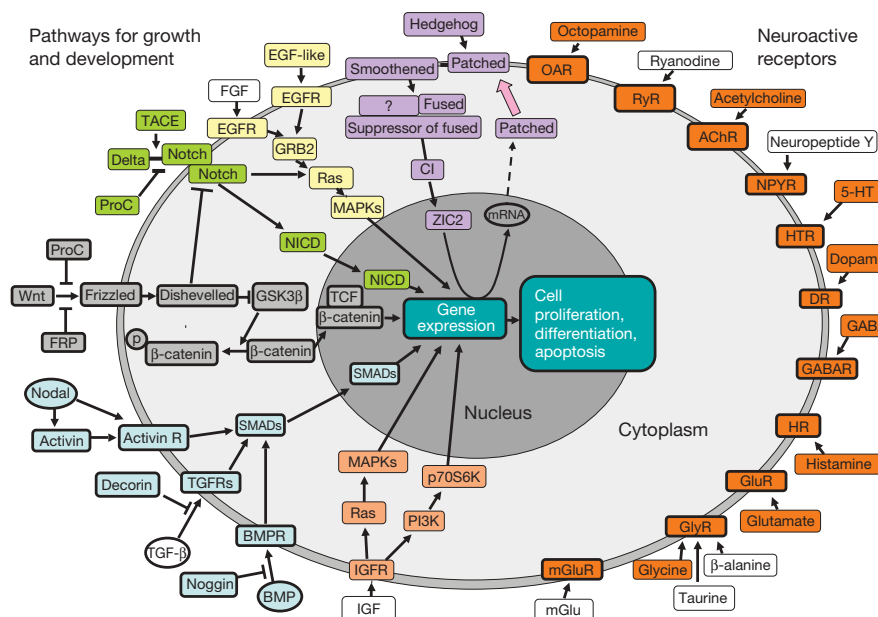


Figure 3 | Putative signalling pathways for growth, development and neuroactive ligand-receptor interaction in *S. japonicum*. The pathways for growth and development (indicated with different colours), and the neuroactive ligand-receptor interactions in *S. japonicum* are shown on the left and right, respectively. TACE, tumour-necrosis-factor- α -converting enzyme; ProC, porcupine homologue (*Drosophila*); NICD, notch intracellular domain; FRP, frizzled-related protein 1; GSK3 β , glycogen synthase kinase 3 β ; TCF, transcription factor 7; 'p' within cycle,

phosphorylation on the proteins indicated; BMP, bone morphogenetic protein; IGF, insulin-like growth factor; mGlu, metabotropic glutamate; GlyR, glycine receptor; GluR, glutamate receptor; HR, histamine receptor; GABA, γ -aminobutyric acid; DR, dopamine receptor; 5-HT, 5-hydroxytryptamine; HTR, 5-hydroxytryptamine receptor; NPYR, neuropeptide Y receptor; AChR, acetylcholine receptor; RyR, ryanodine receptor; OAR, octopamine receptor; ZIC2, Zic family member 2; CI, cubitus interruptus; suffix 'R' denotes receptor.

worms are able to respond to changes in levels of chemicals and nutrients. Using a top-down Gene-Ontology-based strategy to facilitate the gene annotation (Supplementary Fig. 13), we identified 71 genes encoding receptors, membrane channels, enzymes and other components, such as rhodopsins/opsins²⁶, phosrestins/arrestins²⁷, transducins, cyclic nucleotide-gated channel, rhodopsin kinase and guanylate cyclase 2D (Supplementary Table 15). Both *S. japonicum* and *S. mansoni* have only two members of the rhodopsin family, unlike *Drosophila*, which possesses 13 members, and zebrafish, which has at least seven (Supplementary Fig. 14a). Phylogenetic analysis indicated that there are at least four schistosome transducins, each of which could represent a divergent subtype of transducin superfamily across chordates, echinoderms, molluscs and arthropods (Supplementary Fig. 14b), and could therefore mediate distinct responses of sensors to signals.

The genome sequence analysis also revealed an array of genes encoding sensory proteins that could interact with chemical ligands and other stimuli. These included guanine-nucleotide-binding protein, potassium-voltage-gated-channel protein Shaker, the glutamate receptor for umami taste and protein Prospero²⁸ (Supplementary Table 16 and Supplementary Fig. 15a). Notably, the genome encodes most components of four of the five human gustatory sensation pathways: the salty, sour, sweet and umami tastes. We also found several potential sensors for sound perception, a common characteristic of vertebrates and arthropods²⁹, in the genome (Supplementary Table 17).

We discovered an apparently intact olfaction pathway, including cyclic nucleotide-gated olfactory channel, guanine-nucleotide-binding protein and adenylyl cyclase type 3 (Supplementary Table 18 and Supplementary Fig. 15b). Moreover, mechanosensory perception mediated by mechanically gated ion channels represents the basis for the sensing of touch, balance, temperature and sound, and contributes essentially to the development and homeostasis of all Eumetazoa³⁰ (Supplementary Table 19). Putative sensory components for equilibrium/balance, mechanical stimulation, pain and temperature (Supplementary Tables 20 and 21) were also found in the *S. japonicum* genome, including two proteins that have

similarities with the well-known mechanosensory protein, transient receptor potential cation channel³¹, and several receptors such as metabotropic glutamate receptor 3, which participate in the sensory perception of pain, light and taste.

Neuroendocrine system. Schistosomes have receptors that apparently evolved to accept endogenous hormones as well as those of the parasitized mammalian host^{20,32}. By surveying hormones and receptors related to the classical neuroendocrine axis in the genomic sequence of *S. japonicum*, we found (Fig. 4) putative receptors for hypothalamic hormones such as thyrotropin-releasing hormone (TRH), prolactin-releasing hormone, somatostatin, melanin-concentrating hormone and leptin, as well as transmembrane proteins that have some similarities with receptors for gonadotropin-releasing hormone, corticotropin-releasing hormone and growth-hormone-releasing hormone. Moreover, putative receptors are present that show weak similarity with those in mammals for the pituitary hormones thyroid-stimulating hormone (TSH), luteinizing hormone, follicle-stimulating hormone, arginine vasopressin and oxytocin.

Although a hypothalamus–pituitary-like organ has not been described in schistosomes, it is possible that some neurons, similar to those in the hypothalamus and pituitary of vertebrates, could fulfil similar functions in terms of modulating the behaviour of *S. japonicum* through peripheral endocrine tissues and cells. In this regard, it is noteworthy that the genomic information suggests the presence of an integral hypothalamic–pituitary–thyroid axis in *S. japonicum*. In addition to the superior TRH–TSH receptors, an intact system for synthesis of thyroxine and active triiodothyronine, as well as an inactivation mechanism of these hormones using deiodination, was identified. Nuclear receptors for triiodothyronine and thyroxine were revealed with identity to mammalian orthologues. Hence, *S. japonicum* may use an endogenous thyroid hormone/receptor signalling pathway for growth and development (Fig. 4 and Supplementary Table 22).

We confirmed that *S. japonicum* has receptors for steroid hormones such as progesterone, progesterone and oestrogen^{32,33}. In addition, it possesses intricate pathways for processing steroid hormones

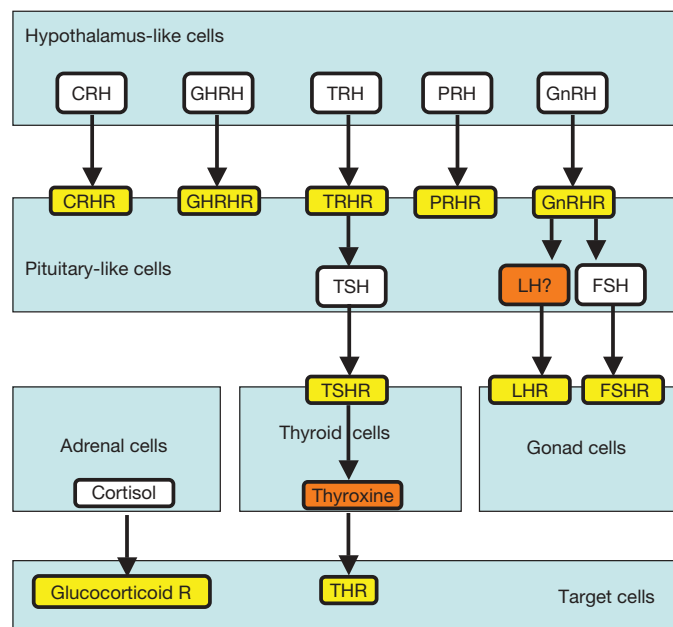


Figure 4 | Putative neuroendocrine system in *S. japonicum*. Structured according to the proposed hypothalamus–pituitary–peripheral-endocrine axis with putative ligands found in *S. japonicum* coloured in orange and *S. japonicum* receptors in yellow. CRH, corticotrophin-releasing hormone; GHRH, growth-hormone-releasing hormone; TRH, thyrotropin-releasing hormone; PRH, prolactin-releasing hormone; GnRH, gonadotropin-releasing hormone; TSH, thyroid-stimulating hormone; FSH, follicle-stimulating hormone; LH, luteinizing hormone; suffix ‘R’ denotes receptor.

to form other sex hormones. For example, there are putative enzymes present that could convert the female hormones progesterone and pregnenolone to estradiol, oestrone, androsterone and testosterone. Hence, schistosomes might use these pathways during their parasitic existence. *Schistosoma japonicum* also encodes enzymes to catabolize excessive or used steroid hormones such as aldosterone.

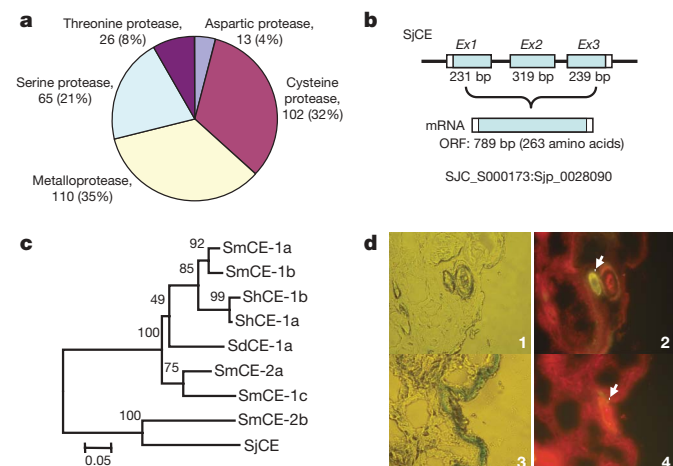


Figure 5 | *S. japonicum* proteases and elastase. **a**, The pie chart shows the distribution of the five kinds of protease. **b**, The genomic structure of *S. japonicum* cercarial elastase (SjCE). **c**, A phylogeny of the elastase family in schistosomes using the neighbour-joining method. Bootstrap values are provided above the branches. SmCE, *S. mansoni* cercarial elastase; ShCE, *S. haematobium* cercarial elastase; SdCE, *Schistosomatium douthitti* elastase. **d**, Immunofluorescence assay showing the presence (white arrow) of SjCE around a schistosomulum following its penetration through mouse skin (panel 2). A naive rabbit serum was used as negative control (panel 4). The location of the cercaria is indicated (white arrow). Panels 1 and 3 show the skin tissue slices under the optical microscope.

With regard to the process of glycolysis for essential energy supply, receptors for adiponectin, an insulin-sensitizing hormone³⁴, and leptin, a suppressor of the secretion of insulin³⁵, are also encoded by the genome of *S. japonicum* (Supplementary Table 22), providing further support for the notion that the blood fluke modulates its energy metabolism in response to either its own insulin-like hormones or those of its mammalian host.

The schistosomulum renews its tegument during maturation into an adult schistosome under the effects of ecdysone^{36,37}. In concordance, we identified an ecdysone-like receptor and its downstream effector ecdysone-induced protein 78C. In addition, allatostatin, a polypeptide hormone that suppresses the secretion of juvenile hormone, was previously reported to be found throughout the schistosome nervous system^{38,39}. An allatostatin-like receptor sequence that has high similarity with that of the cockroach was also identified (Supplementary Table 22).

Disease pathogenesis

Cercarial elastase and protease superfamily. Schistosome proteases have key roles in invasion⁴⁰, migration⁴¹ and feeding/nutrition⁴². We identified 314 putative proteases, including metallo-, cysteine, serine, threonine and aspartic proteases, in the *S. japonicum* genome data set (Fig. 5a and Supplementary Tables 23–27) by searching in the MEROPS database of peptidases. We classified 108 *S. japonicum* metalloproteases into 21 subtypes, 16 belonging to the aminopeptidases (Supplementary Table 23). Notably, the leucine aminopeptidase of the M17 family was reported as a major egg antigen^{43,44} and a putative anti-fluke vaccine⁴⁵. The second largest assemblage comprised the cysteine proteases, of which 102 members were assigned to 17 subtypes (Supplementary Table 24). Among them, the cathepsins B, C, F and L have pivotal roles in schistosome feeding and nutrition⁴², as well as in migration through human tissues⁴¹. The cysteine proteases cathepsins K and S, as well as the cathepsin A serine protease, have not previously been recognized in schistosomes, and may contribute to catabolism of haemoglobin and other host proteins.

Among the 65 serine proteases (Supplementary Table 25), we discovered a *S. japonicum* cercarial elastase (SjCE), an enzyme that in *S. mansoni* is vital in the penetration by cercariae of mammalian skin to initiate infection^{40,46}. The elastase locus predicted from the *S. japonicum* genome spans three exons and two introns, similar to the known *S. mansoni* elastases⁴⁷ (Fig. 5b); however, unlike for *S. mansoni*, only a single elastase was identified in *S. japonicum*. Phylogenetic analysis of available schistosome elastases (Supplementary Table 28) suggested that the elastase genes in *S. mansoni* have expanded through at least two rounds of gene duplication, whereas SjCE is an orthologue of *S. mansoni* cercarial elastase 2b (Fig. 5c). Moreover, by re-examination of mass spectra data that we collected previously³³, we identified a unique peptide (IAFLALSDFDHR) of SjCE in cercariae (Supplementary Fig. 16a). We also confirmed the existence of SjCE gene products in both the sporocyst and cercarial stages of *S. japonicum* by immunoblot and immunofluorescence assays (Fig. 5d and Supplementary Fig. 16b). In addition, the native protease was recognized by anti-recombinant SjCE antibodies in infected mouse skin, indicating that this cercarial elastase is secreted/released by the parasite during invasion of mammalian skin.

Immune system and inflammatory factors. The immune system of *S. japonicum* has to face both invading microbial pathogens and the immune statuses of both its molluscan and mammalian hosts. Although adaptive immune molecules such as immunoglobulin are lacking in *S. japonicum* and a classical Toll-like receptor was not found, putative Toll-interacting protein or proteins containing Toll/interleukin-1 resistance motif or leucine-rich repeats appear to be present (Supplementary Table 29). Therefore, schistosomes, like nematodes, appear to possess a primordial Toll pathway as a first line of defence against microbial infections. The identification of the downstream components of a Toll-related pathway, including putative interleukin-1-receptor-associated kinases, toll-like receptor

adaptors, TNF-receptor-associated factor 6 (TRAF6), inhibitor of nuclear factor κ B kinase subunit epsilon (IKK- ϵ) and p38 MAPK, further support the view that this primitive innate immune system could be crucial for the worm (Supplementary Table 30).

On the other hand, factors and metabolites in *S. japonicum* that could contribute to stimulation and regulation of mammalian immunity were discovered. It is well accepted that glycans and lipids synthesized by adult schistosomes or eggs may regulate secondary signals through corresponding receptors on effector cells and accessory cells of the mammalian host, thus compromising host immunological defences targeting the parasite. We therefore searched for enzymes involved in the metabolism of various glycans or lipid antigens by interrogating this worm genome. It turned out that, with the rare exception of enzymes such as α 1,3-mannosyltransferase, a complete set of enzymatic machinery for biosynthesis and modification of glycans and lipids exists (Supplementary Table 31).

In addition, prostaglandins, which are well-known mediators of inflammation, can be synthesized by *S. japonicum* as a result of arachidonic-acid metabolism. It is feasible that *S. japonicum* synthesizes arachidonate by using lecithin, converting the arachidonate into leukotriene A4 using arachidonate 5-lipoxygenase, followed by the conversion of unstable leukotriene A4 into the active chemical leukotriene B4 through leukotriene A4 hydrolase. The *S. japonicum* genome also encodes putative receptors for leukotriene B4, cysteinyl leukotriene and prostaglandins E2 and F2, suggesting that prostaglandins could have an important role in the physiology of schistosomes and also in the host-parasite interplay. Unexpectedly, *S. japonicum* possesses proteins paralogous to mammalian autoimmune-disease-related autoantigens; these include 69 kDa islet cell autoantigen (ICA1), islet antigen-2 (PTPRN) and glutamate decarboxylase (GAD), known autoantigens related to type-I diabetes in β -cells, which raises the possibility that these autoantigen-mimicking molecules could induce chemokine-receptor-mediated cell migration and initiate leukocyte migration into inflamed tissue, which ultimately contribute to the granuloma formation that promotes parasite survival.

Concluding remarks

Lophotrochozoa, of which *S. japonicum* is a member, is a large taxon that includes ~50% of all animal phyla. Our work provides a model for evaluating the genomic architecture, biology and evolution in this major taxon. Although the genome of *S. japonicum* has undergone significant protein-domain-loss events, a detailed molecular repertoire exists to permit the pathogen to locate and penetrate hosts, nourish itself and interact with the environment and its host. With the release and analysis of the *S. mansoni* genome⁴⁸, a comparative-genomics approach elucidating the similarities and differences between these two closely related parasites will provide more clues regarding these important pathways. Further functional analysis, using approaches such as RNA interference and translational studies are essential to resolve uncertainties in the molecular physiology of schistosomes and to illuminate mechanisms of pathogenesis in schistosomiasis, efforts that may lead to the development of new interventions for its control and eventual elimination.

METHODS SUMMARY

We obtained adult worms and eggs of *S. japonicum* from infected rabbits. The genomic DNA was extracted from ~1,000 mixed, outbred adult male and female *S. japonicum*, perfused from rabbits infected with cercariae released by naturally infected snails. Genomic libraries, including bacterial artificial chromosome (BAC), fosmid and plasmid libraries, were constructed. We performed WGS sequencing on capillary sequencers, and then used a modified PHUSION (version 2.1c) package to assemble the reads. Protein-encoding genes were predicted using EXONHUNTER (version 2.0)⁴⁹. We used a stepwise method to predict the gene functions. The metabolic and regulatory pathway of *S. japonicum* was reconstructed with reference to the KEGG pathway database. Proteins were first clustered using a Markov cluster algorithm and then merged according to

protein-domain information to establish protein-family clusters. We used immunoblot and immunofluorescence assays to detect cercarial elastase.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 February; accepted 8 May 2009.

- Adamson, P. B. Schistosomiasis in antiquity. *Med. Hist.* **20**, 176–188 (1976).
- Zhou, X. N. *et al.* The public health significance and control of schistosomiasis in China - then and now. *Acta Trop.* **96**, 97–105 (2005).
- King, C. H., Dickman, K. & Tisch, D. J. Reassessment of the cost of chronic helminth infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet* **365**, 1561–1569 (2005).
- Steinmann, P., Keiser, J., Bos, R., Tanner, M. & Utzinger, J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infect. Dis.* **6**, 411–425 (2006).
- Finkelstein, J. L., Schleinitz, M. D., Carabin, H. & McGarvey, S. T. Decision-model estimation of the age-specific disability weight for *Schistosomiasis japonica*: a systematic review of the literature. *PLoS Negl. Trop. Dis.* **2**, e158 (2008).
- Utzinger, J., Zhou, X. N., Chen, M. G. & Bergquist, R. Conquering schistosomiasis in China: the long march. *Acta Trop.* **96**, 69–96 (2005).
- Li, Y. S. *et al.* Large water management projects and schistosomiasis control, Dongting Lake region, China. *Emerg. Infect. Dis.* **13**, 973–979 (2007).
- Bergquist, R., Utzinger, J. & McManus, D. P. Trick or treat: the role of vaccines in integrated schistosomiasis control. *PLoS Negl. Trop. Dis.* **2**, e244 (2008).
- Amiri, P. *et al.* Tumour necrosis factor α restores granulomas and induces parasite egg-laying in schistosomiasis-infected SCID mice. *Nature* **356**, 604–607 (1992).
- Davies, S. J. *et al.* Modulation of blood fluke development in the liver by hepatic CD4⁺ lymphocytes. *Science* **294**, 1358–1361 (2001).
- Hirai, H. *et al.* Chromosomal differentiation of the *Schistosoma japonicum* complex. *Int. J. Parasitol.* **30**, 441–452 (2000).
- Moné, H. & Boissier, J. Sexual biology of schistosomes. *Adv. Parasitol.* **57**, 89–189 (2004).
- Halanych, K. M. The new view of animal phylogeny. *Annu. Rev. Ecol. Syst.* **35**, 229–256 (2004).
- Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
- Arkhipova, I. R., Pyatkov, K. I., Meselson, M. & Evgen'ev, M. B. Retroelements containing introns in diverse invertebrate taxa. *Nature Genet.* **33**, 123–124 (2003).
- Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
- Gamulin, V., Muller, I. M. & Muller, W. E. Sponge proteins are more similar to those of *Homo sapiens* than to *Caenorhabditis elegans*. *Biol. J. Linn. Soc.* **71**, 821–828 (2000).
- Robb, S. M., Ross, E. & Sanchez Alvarado, A. SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Res.* **36**, D599–D606 (2008).
- Curwen, R. S., Ashton, P. D., Sundaralingam, S. & Wilson, R. A. Identification of novel proteases and immunomodulators in the secretions of schistosome cercariae that facilitate host entry. *Mol. Cell. Proteomics* **5**, 835–844 (2006).
- Hu, W. *et al.* Evolutionary and biomedical implications of a *Schistosoma japonicum* complementary DNA resource. *Nature Genet.* **35**, 139–147 (2003).
- Mousley, A., Maule, A. G., Halton, D. W. & Marks, N. J. Inter-phyla studies on neuropeptides: the potential for broad-spectrum anthelmintic and/or endectocide discovery. *Parasitology* **131**, S143–S167 (2005).
- Holmes, S. P., Barhoumi, R., Nachman, R. J. & Pietrantoni, P. V. Functional analysis of a G protein-coupled receptor from the southern cattle tick *Boophilus microplus* (Acari: Ixodidae) identifies it as the first arthropod myokinin receptor. *Insect Mol. Biol.* **12**, 27–38 (2003).
- Egerod, K. *et al.* Molecular cloning and functional expression of the first two specific insect myosuppressin receptors. *Proc. Natl Acad. Sci. USA* **100**, 9808–9813 (2003).
- Scholler, S. *et al.* Molecular identification of a myosuppressin receptor from the malaria mosquito *Anopheles gambiae*. *Biochem. Biophys. Res. Commun.* **327**, 29–34 (2005).
- Cohen, L. M., Neimark, H. & Eveland, L. K. *Schistosoma mansoni*: response of cercariae to a thermal gradient. *J. Parasitol.* **66**, 362–364 (1980).
- 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).
- Matsumoto, H. & Yamada, T. Phosrestins I and II: arrestin homologs which undergo differential light-induced phosphorylation in the *Drosophila* photoreceptor *in vivo*. *Biochem. Biophys. Res. Commun.* **177**, 1306–1312 (1991).
- Grosjean, Y., Lacaille, F., Acebes, A., Clemencet, J. & Ferveur, J. F. Taste, movement, and death: varying effects of new prospero mutants during *Drosophila* development. *J. Neurobiol.* **55**, 1–13 (2003).
- Robert, D. & Gopfert, M. C. Acoustic sensitivity of fly antennae. *J. Insect Physiol.* **48**, 189–196 (2002).
- Walker, R. G., Willingham, A. T. & Zuker, C. S. A *Drosophila* mechanosensory transduction channel. *Science* **287**, 2229–2234 (2000).
- Mutai, H. & Heller, S. Vertebrate and invertebrate TRPV-like mechanoreceptors. *Cell Calcium* **33**, 471–478 (2003).

32. Hu, W., Brindley, P. J., McManus, D. P., Feng, Z. & Han, Z. G. Schistosome transcriptomes: new insights into the parasite and schistosomiasis. *Trends Mol. Med.* **10**, 217–225 (2004).
33. Liu, F. *et al.* New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathog.* **2**, e29 (2006).
34. Heilbronn, L. K., Smith, S. R. & Ravussin, E. The insulin-sensitizing role of the fat derived hormone adiponectin. *Curr. Pharm. Des.* **9**, 1411–1418 (2003).
35. Kieffer, T. J., Heller, R. S., Leech, C. A., Holz, G. G. & Habener, J. F. Leptin suppression of insulin secretion by the activation of ATP-sensitive K⁺ channels in pancreatic beta-cells. *Diabetes* **46**, 1087–1093 (1997).
36. Foster, J. M., Mercer, J. G. & Rees, H. H. Analysis of ecdysteroids in the trematodes, *Schistosoma mansoni* and *Fasciola hepatica*. *Trop. Med. Parasitol.* **43**, 239–244 (1992).
37. Basch, P. F. Immunocytochemical localization of ecdysteroids in the life history stages of *Schistosoma mansoni*. *Comp. Biochem. Physiol. Comp. Physiol.* **83**, 199–202 (1986).
38. Smart, D. *et al.* Localization of *Diploptera punctata* allatostatin-like immunoreactivity in helminths: an immunocytochemical study. *Parasitology* **110**, 87–96 (1995).
39. Smart, D. *et al.* Peptides related to the *Diploptera punctata* allatostatins in nonarthropod invertebrates: an immunocytochemical survey. *J. Comp. Neurol.* **347**, 426–432 (1994).
40. Dvorak, J. *et al.* Differential use of protease families for invasion by schistosome cercariae. *Biochimie* **90**, 345–358 (2008).
41. Dvorak, J. *et al.* Multiple cathepsin B isoforms in schistosomula of *Trichobilharzia regenti*: identification, characterisation and putative role in migration and nutrition. *Int. J. Parasitol.* **35**, 895–910 (2005).
42. Koehler, J. W., Morales, M. E., Shelby, B. D. & Brindley, P. J. Aspartic protease activities of schistosomes cleave mammalian hemoglobins in a host-specific manner. *Mem. Inst. Oswaldo Cruz* **102**, 83–85 (2007).
43. Abouel-Nour, M. F., Lotfy, M., El-Kady, I., El-Shahat, M. & Doughty, B. L. Localization of leucine aminopeptidase in the *Schistosoma mansoni* eggs and in liver tissue from infected mice. *J. Egypt. Soc. Parasitol.* **35**, 147–156 (2005).
44. Xu, Y. Z., Shawar, S. M. & Dresden, M. H. *Schistosoma mansoni*: purification and characterization of a membrane-associated leucine aminopeptidase. *Exp. Parasitol.* **70**, 124–133 (1990).
45. Hillyer, G. V. Fasciola antigens as vaccines against fascioliasis and schistosomiasis. *J. Helminthol.* **79**, 241–247 (2005).
46. Newport, G. R. *et al.* Cloning of the proteinase that facilitates infection by schistosome parasites. *J. Biol. Chem.* **263**, 13179–13184 (1988).
47. 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).
48. Berriman, M. *et al.* The genome of the blood fluke *Schistosoma mansoni*. *Nature* doi:10.1038/nature08160 (this issue).
49. Brejova, B. *et al.* Finding genes in *Schistosoma japonicum*: annotating novel genomes with help of extrinsic evidence. *Nucleic Acids Res.* **37**, e52 (2009).
50. Stricklin, S. L., Griffiths-Jones, S. & Eddy, S. R. C. *elegans* noncoding RNA genes. *WormBook* **25**, 1–7 (2005).
51. Ghedin, E. *et al.* Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* **317**, 1756–1760 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This investigation was mainly supported by the Chinese National High-Tech Program (863 Program) (2004AA2Z1010, 2006AA02Z335, 2006AA02Z318, 2007AA02Z153), the Chinese National Key Project for Basic Research (973 Project) (2006CB708510, 2007CB513100), the Chinese Academy of Sciences, the Shanghai Municipal Commission for Science and Technology (04DZ14010, 055407031, 06JC14059, 07QA14043, 07DZ22915), and the National Natural Science Foundation of China. Support from the US National Institute of Allergy and Infectious Diseases (award number AI39461), the National Science and Engineering Research Council of Canada (OGP0046506), a International Collaborative Research Grants award from the National Health and Medical Research Council of Australia, and the Wellcome Trust, UK, is also gratefully acknowledged. The Shanghai Supercomputer Center kindly provided computational facilities for some of the data analysis. The authors wish to thank P. De Jong and his colleagues for the BAC libraries construction of *S. japonicum* and N. M. El-Sayed for his contribution on the collaboration between the *S. japonicum* and *S. mansoni* sequencing consortia.

Author Contributions Y. Zhou, H.Z., F.L., W. Hu, Z.-Q.W., G.L. and S.R. contributed equally to this work.

Author Information The sequences of *S. japonicum* WGS assembly contigs and scaffolds, BACs, full-length complementary DNAs and retrotransposons have

been deposited in the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database (<http://www.ebi.ac.uk/embl/>) and the Shanghai Center for Life Science & Biotechnology Information (LSBI; <http://lifecenter.sgst.cn/schistosoma/en/schistosomaCnIndexPage.do>), and can be freely downloaded. The EMBL accession numbers are CABF01000001–CABF01095265 (contigs) FN330975–FN3356022 (scaffolds), FN293020–FN293041 (BACs), FN313573–FN330973 (full-length cDNAs), FN356203–FN356227 (retrotransposons). The LSBI accession numbers are CNUS00000108051–CNUS00000203315 (contigs), CCON0000096785–CCON00000121832 (scaffolds), CNUS0000095394–CNUS00000108050 (predicted genes), CPRT0000000001–CPRT0000012657 (predicted proteins), CNUS00000203316–CNUS00000203337 (BACs), CNUS00000203338–CNUS00000220738 (full-length cDNAs), CNUS00000220739–CNUS00000220763 (retrotransposons). The sequences of *S. japonicum* integrated protein-coding genes are available on the Chinese National Human Genome Center at Shanghai website (<http://www.chgc.sh.cn/japonicum>). The BAC library (CHORI-108) is available from the laboratory of P. De Jong at the BACPAC Resources Center, Children's Hospital Oakland Research Institute, California (<http://bacpac.chori.org/library.php?id=168>). Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.W. (wangsy@chgc.sh.cn), Z.-G.H. (hanzg@chgc.sh.cn) or Z. Chen (zchen@stn.sh.cn).

The *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium

Genome annotation and evolution analysis Yan Zhou^{1,2}, Huajun Zheng^{1,2}, Yangyi Chen¹, Lei Zhang¹, Kai Wang¹, Jing Guo¹, Zhen Huang¹, Bo Zhang¹, Wei Huang¹, Ke Jin², Tonghai Dou², Masami Hasegawa², Li Wang^{2,3}, Yuan Zhang², Jie Zhou², Lin Tao³, Zhiwei Cao³, Yixue Li³, Tomas Vinar⁴, Brona Brejova⁴, Dan Brown⁴, Ming Li⁴, David J. Miller⁵, David Blair⁵, Yang Zhong (Principal Investigator)^{2,3}, Zhu Chen (Principal Investigator)^{1,6}; **Functional genomics analysis** Feng Liu^{1,2}, Wei Hu¹, Zhi-Qin Wang¹, Qin-Hua Zhang⁸, Huai-Dong Song⁶, Saijuan Chen⁶, Xuenian Xu⁷, Bin Xu⁷, Chuan Ju⁷, Yucheng Huang⁷, Paul J. Brindley⁹, Donald P. McManus¹⁰, Zheng Feng (Principal Investigator)⁷, Ze-Guang Han (Principal Investigator)¹; **Sequencing and assembly** Gang Lu^{1,6}, Shuangxi Ren¹, Yuezhong Wang¹, Wenyi Gu¹, Hui Kang¹, Jie Chen¹, Xiaoyun Chen¹, Shuting Chen¹, Lijun Wang¹, Jie Yan¹, Biyun Wang¹, Xinyan Lv¹, Lei Jin¹, Bofei Wang¹, Shiyin Pu¹, Xianglin Zhang¹, Wei Zhang¹, Qiuping Hu¹, Genfeng Zhu¹, Jun Wang¹¹, Jun Yu¹¹, Jian Wang¹¹, Huanming Yang¹¹, Zemin Ning¹², Matthew Beriman¹², Chia-Lin Wei¹³, Yijun Ruan¹³, Guoping Zhao (Principal Investigator)^{1,2,14}, Shengyue Wang (Principal Investigator)¹; **Paper writing** Feng Liu^{1,2}, Yan Zhou^{1,2}, Zhi-Qin Wang¹, Gang Lu^{1,6}, Huajun Zheng^{1,2}, Paul J. Brindley⁹, Donald P. McManus¹⁰, David Blair⁵, Qin-hua Zhang⁸, Yang Zhong^{2,3}, Shengyue Wang¹, Ze-Guang Han¹, Zhu Chen^{1,6}; **Project leaders** Shengyue Wang¹, Ze-Guang Han¹, Zhu Chen^{1,6}

¹Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China. ²School of Life Science/Institutes of Biomedical Sciences, Fudan University, 220 Han Dan Road, Shanghai 200433, China. ³Shanghai Center for Bioinformatics Technology, 100 Qinzhou Road, Shanghai 200235, China. ⁴Cheriton School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada. ⁵Comparative Genomics Centre/School of Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia. ⁶State Key Laboratory of Medical Genomics and Shanghai Institute of Hematology, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, 197 Rui Jin Road II, Shanghai 200025, China. ⁷National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, 207 Rui Jin Er Road, Shanghai 200025, China. ⁸Shanghai Center for Biochip Engineering, 151 Li Bing Road, Shanghai 201203, China. ⁹Department of Microbiology, Immunology & Tropical Medicine, George Washington University Medical Center, Ross Hall, Room 448, 2300 I Street, NW, Washington DC 20037, USA. ¹⁰Molecular Parasitology Laboratory, Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia. ¹¹Beijing Institute of Genomics, Chinese Academy of Sciences/Beijing Genomics Institute, B-6 Beijing Airport Industrial Zone, Beijing 101300, China. ¹²Pathogen Sequencing Unit, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. ¹³Genome Institute of Singapore, 60 Biopolis Street, Genome #02-01, 138672, Singapore. ¹⁴Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai 200031, China.

METHODS

Schistosoma japonicum genomic and full-length cDNA library construction.

Genomic DNA was extracted from ~1,000 mixed, outbred adult male and female *S. japonicum*, perfused from rabbits infected with cercariae released by naturally infected snails collected from an endemic focus in Anhui Province, as described²⁰. Four genomic libraries with different insert sizes were constructed, one of bacterial artificial chromosomes (inserts, 80–120 kb), one of fosmids (36–42 kb) and two of plasmids (6–10 kb and 1.6–4 kb) (Supplementary Table 1). Total RNAs from *S. japonicum* adults and eggs were isolated using Trizol (Invitrogen), after which mRNA was purified using the Poly(A) Purist mRNA Purification Kit (Ambion). Two full-length cDNA libraries, from adults and eggs were constructed using a modified biotinylated CAP-trapper approach^{52,53}.

WGS sequencing and assembly. After the clone ends of four discrete genomic libraries were sequenced by capillary DNA sequencers ABI3700 (Applied Biosystems) and MegaBACE 1000 or MegaBACE 4000 (General Electric), PHRED (version 0.020425.c)^{54,55} was used for base calling. All reads were qualified by removing clone vector and bacterial host sequences, as well as the host rabbit (*Oryctolagus cuniculus*) DNA sequences (http://www.ensembl.org/Oryctolagus_cuniculus/index.html). A modified PHUSION (version 2.1c) package⁵⁶ was used for assembly.

Repeat and retrotransposon identification. A repetitive sequence library of *S. japonicum* was generated by the method of consensus seed extending using REPEATSCOUT (version 1.0.3)¹⁴, with the k-mer size of 16. Tandem repeats in the genome were identified using TANDEM REPEATS FINDER (version 4.00)⁵⁷ and categorized using the tandem repeats analysis program TRAP (version 1.0)⁵⁸. Microsatellites, minisatellites and satellites are classically defined as repeat units of 1–6 bp, 11–100 bp and more than 100 bp, respectively. Polyprotein and reverse transcriptase from GenBank were used as queries to search genome sequences of *S. japonicum* using tBLASTN (e-value $\leq 10^{-10}$). The best hit sequences were then used to query the genome, and those yielding multiple hits in the genome were categorized as candidate retrotransposons. All candidate retrotransposons were assembled to establish complete CDSs encoding polyprotein or reverse transcriptase. Once the complete CDS was determined, sequences upstream and downstream of this CDS in the genome were analysed to identify LTRs which flank the left and right termini of LTR retrotransposons and retroviruses.

Prediction and integration of protein-coding genes. Protein encoding genes were predicted using EXONHUNTER (version 2.0)⁴⁹. The prediction program combined *ab initio* gene prediction with supporting evidence from *S. japonicum* and *S. mansoni* expressed sequence tags, *S. japonicum* pair-end ditags, the Swiss-Prot protein database⁵⁹ and the Pfam protein-domain database (version 22.0)⁶⁰. Because there were few training sets available for *S. japonicum* or for any other closely related species, we developed an iterative method that started from the distantly related species *C. elegans*, and progressively improved parameters of the gene finder on the basis of well-supported predicted gene fragments. The predicted genes were merged with putative expressed sequence tags and full-length cDNA-derived CDSs (proteins), yielding an integrated protein-coding gene set for further functional analysis. These genes were classified into categories established by the Gene Ontology project through the encoding proteins or domains matched to the Gene Ontology index provided by UniProt⁶¹ and InterPro⁶² (iprscan_DATA_17.0 and iprscan_PTHR_DATA_14.0).

Genome variation analysis. The PHUSION assembler⁵⁶ does not provide alignment information of reads to its contig consensus, so BLASTN was used to relocate reads to contig consensus, with overall identity of over 95%, and to provide alignment information. We established a locally developed SNP pipeline based on neighbourhood quality standard, with the following rules: for each candidate SNP on shotgun reads, the 5-bp flanking sequences should be the same as the contig consensus, the base quality on the SNP site should be no less than 23 and the base quality of the flanking 5 bp should be not less than 15 (refs 63, 64).

Pathway mapping. The metabolic and regulatory pathway of *S. japonicum* was reconstructed on the basis of the KEGG pathway database⁶⁵. The KEGG orthology identifier was used as a linkage between genes and pathways. The assignment of *S. japonicum* genes to KEGG orthologues was implemented with a modified bidirectional-best-BLAST-hits method, which was adjusted using phylogenetic

information. The pathway mapping results for the *S. japonicum* genome are available at <http://chgc.sh.cn/japonicum>.

Gene-family analysis. Proteins of *S. japonicum*, *C. elegans*, *D. melanogaster*, *A. gambiae*, *D. rerio*, *G. gallus*, *H. sapiens* and *N. vectensis* were first clustered using a Markov cluster algorithm⁶⁶ and then merged according to protein-domain information to establish protein-family clusters. The *S. japonicum* protein domains were scanned using INTERPROSCAN⁶². Protein-domain information on other species was sourced from the KEGG database⁶⁵.

Analysis of *S. japonicum* proteases. Putative proteases in the *S. japonicum* data set were identified by comparing *S. japonicum* cDNA and predicted genes with the MEROPS database⁶⁷. The results were manually checked and compared with annotations generated by BLAST searches against more comprehensive databases as above. Results with inconsistent annotations from MEROPS and BLAST were removed. For phylogenetic and evolutionary analyses of gene families, deduced amino-acid sequences were aligned using CLUSTALW (version 1.83)⁶⁸. Phylogenetic trees were generated using MEGA (version 3.1)⁶⁹ with the neighbour-joining method and tested with 1,000 bootstrap replicates.

Immunofluorescence assay of *S. japonicum* cercarial elastase. A mouse anaesthetized with pentobarbital was infected with *S. japonicum* cercariae. After 10 min, the skin was excised, finely diced, and embedded in OCT fixative. The prepared 7- μ m-thick frozen sections were incubated for 30 min in a solution of 20% goat serum in Tris-HCl-buffered saline. The sections were incubated with the rabbit primary antiserum raised against purified recombinant SjCE or normal rabbit serum, followed by a FITC-conjugated second antibody. Fluorescence was visualized using a Leica DM-2500 fluorescence microscope.

52. Seki, M., Carninci, P., Nishiyama, Y., Hayashizaki, Y. & Shinozaki, K. High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated CAP trapper. *Plant J.* **15**, 707–720 (1998).
53. Wei, C. L. *et al.* 5' long serial analysis of gene expression (LongSAGE) and 3' LongSAGE for transcriptome characterization and genome annotation. *Proc. Natl Acad. Sci. USA* **101**, 11701–11706 (2004).
54. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
55. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
56. Mullikin, J. C. & Ning, Z. The phusion assembler. *Genome Res.* **13**, 81–90 (2003).
57. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
58. Sobreira, T. J., Durham, A. M. & Gruber, A. TRAP: automated classification, quantification and annotation of tandemly repeated sequences. *Bioinformatics* **22**, 361–362 (2006).
59. Gasteiger, E. *et al.* ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
60. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **36**, D281–D288 (2008).
61. UniProt Consortium. The universal protein resource (UniProt). *Nucleic Acids Res.* **36**, D190–D195 (2008).
62. Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
63. Mullikin, J. C. *et al.* An SNP map of human chromosome 22. *Nature* **407**, 516–520 (2000).
64. Altshuler, D. *et al.* An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**, 513–516 (2000).
65. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2008).
66. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
67. Rawlings, N. D., Morton, F. R. & Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Res.* **34**, D270–D272 (2006).
68. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
69. Kumar, S., Tamura, K. & Nei, M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**, 150–163 (2004).

The *Schistosoma japonicum* genome reveals features of host-parasite interplay

Content:

Content:	1
Supplementary information	4
<i>S. japonicum</i> isolate	4
WGS sequencing and assembly	4
Construction and sequencing of flicDNA and PET libraries	5
Prediction and integration of protein coding genes	6
Annotation	7
Gene cataloguing	8
Genome variation analysis	9
Pathway mapping	9
Carbohydrate metabolism	10
Amino acid metabolism	10
Purine and pyrimidine metabolism	10
Fatty acid metabolism	11
Steroid biosynthesis	12
Gene loss and duplication	13
References	15
Supplementary Tables	17
Supplementary Table 1. Whole genome sequencing reads of <i>S. japonicum</i>	17
Supplementary Table 2. <i>S. japonicum</i> genome assembly statistics	18
Supplementary Table 3. Sequence variation in the <i>S. japonicum</i> genome	19
Supplementary Table 4. Distribution of repeats in <i>S. japonicum</i>	20
Supplementary Table 5. Statistics for retrotransposons in the <i>S. japonicum</i> genome	21
Supplementary Table 6. <i>S. japonicum</i> ortholog relationship with other species	23
Supplementary Table 7. Summary of gene families in several organisms	24
Supplementary Table 8. Leishmanolysin gene family in <i>S. japonicum</i>	25
Supplementary Table 9. Expanded <i>S. japonicum</i> gene families compared with other taxa	26
Supplementary Table 10. <i>S. japonicum</i> receptors, ligands and modulators of the Hedgehog, Wnt and Notch pathways	27
Supplementary Table 11. Growth factors, cytokines, chemokines and their receptors in <i>S. japonicum</i>	28
Supplementary Table 12. KEGG pathway mapping of <i>S. japonicum</i> genes and ESTs	29
Supplementary Table 13. Functional annotation and mapping of genes associated with lipid metabolism	30

Supplementary Table 14. Neurotransmitter transporters and receptors of <i>S. japonicum</i>	31
Supplementary Table 15. Light sensory molecules of <i>S. japonicum</i>	33
Supplementary Table 16. Chemical sensory molecules of <i>S. japonicum</i>	41
Supplementary Table 17. Sound sensory molecules of <i>S. japonicum</i>	42
Supplementary Table 18. Olfactory sensory molecules of <i>S. japonicum</i>	47
Supplementary Table 19. Mechanical/thermoception molecules of <i>S. japonicum</i>	48
Supplementary Table 20. Equilibrioception molecules of <i>S. japonicum</i>	49
Supplementary Table 21. Pain sensory molecules of <i>S. japonicum</i>	50
Supplementary Table 22. Hormone receptors of <i>S. japonicum</i>	51
Supplementary Table 23. The predicted metallo-proteases of <i>S. japonicum</i>	53
Supplementary Table 24. The predicted cysteine proteases of <i>S. japonicum</i>	63
Supplementary Table 25. The predicted serine proteases of <i>S. japonicum</i>	74
Supplementary Table 26. The predicted threonine proteases of <i>S. japonicum</i>	80
Supplementary Table 27. The predicted aspartic proteases of <i>S. japonicum</i>	83
Supplementary Table 28. Elastase-encoding sequences of schistosomes	85
Supplementary Table 29. Immune defense molecules in <i>S. japonicum</i>	86
Supplementary Table 30. TIR downstream signaling of <i>S. japonicum</i>	87
Supplementary Table 31: Enzymes involved with glycan and lipid synthesis in <i>S. japonicum</i>	88
Supplementary Figure 1. Schematic representation of a portion of the <i>S. japonicum</i> genome illustrated with various feature tracks	96
Supplementary Figure 2. Support for predicted genes by expression datasets	98
Supplementary Figure 3. Integration of <i>S. japonicum</i> protein dataset	99
Supplementary Figure 4. Functional categorization of <i>S. japonicum</i> genes and comparative analysis with other representative organisms	101
Supplementary Figure 5. Stepwise annotation of <i>S. japonicum</i> genes	102
Supplementary Figure 6. Protein domain analysis in metazoan	103
Supplementary Figure 7. An overview of basic metabolic pathways in <i>S. japonicum</i>	104
Supplementary Figure 8. Pathways of carbohydrate metabolism (include five sub-figures)	106
Supplementary Figure 8.1. Glycolysis and gluconeogenesis	107
Supplementary Figure 8.2. Citrate cycle (TCA cycle)	108
Supplementary Figure 8.3. Pentose phosphate pathway	109
Supplementary Figure 8.4. Glycogen metabolism	110
Supplementary Figure 8.5. Oxidative phosphorylation	111
Supplementary Figure 9. Amino acid metabolism (include 10 sub-figures)	112
Supplementary Figure 9.1. Alanine and aspartate/asparagine metabolism	113
Supplementary Figure 9.2. Cysteine metabolism	114
Supplementary Figure 9.3. Glutamate and glutamine metabolism	115
Supplementary Figure 9.4. Glycine, serine and threonine metabolism	116
Supplementary Figure 9.5. Arginine and proline metabolism	117
Supplementary Figure 9.6. Histidine metabolism	118
Supplementary Figure 9.7. Lysine biosynthesis	119
Supplementary Figure 9.8. Methionine metabolism	120

Supplementary Figure 9.9. Phenylalanine, tyrosine and tryptophan metabolism	121
Supplementary Figure 9.10. Valine, leucine and isoleucine metabolism	122
Supplementary Figure 10. Nucleotide metabolism (include two sub-figures)	123
Supplementary Figure 10.1. Purine metabolism	123
Supplementary Figure 10.2. Pyrimidine metabolism.....	124
Supplementary Figure 11. Lipids metabolism (include four sub-figures).....	125
Supplementary Figure 11.1. Fatty acid metabolism.....	125
Supplementary Figure 11.2. Fatty acid biosynthesis.....	126
Supplementary Figure 11.3. Glycerophospholipid metabolism	127
Supplementary Figure 11.4. Cholesterol biosynthesis pathway	128
Supplementary Figure 12. Octopamine synthesis and phylogenetic analysis of octopamine receptors and synthesizing enzymes	130
Supplementary Figure 13. Profile the sensory components of <i>S. japonicum</i> using a top-down strategy.....	132
Supplementary Figure 14. Phylogenetic analysis of light sensory components in <i>S. japonicum</i>	133
Supplementary Figure 15. Putative taste and olfactory pathways in <i>S. japonicum</i>	134
Supplementary Figure 16. Identification of <i>S. japonicum</i> elastase	135

Supplementary information

S. japonicum isolate

Adult worms and eggs for this study were obtained as follows. Snails (*Oncomelania hupensis hupensis*) naturally infected with *S. japonicum* were collected from an endemic focus directly from the field, Guichi County, Anhui Province, China, and maintained in the laboratories of the Centers for Disease Control, National Institute for Parasitic Diseases, Shanghai. Cercariae from the infected snails were used to infect naïve laboratory-raised rabbits, from which adult schistosome worms and eggs were harvested 42 days later. Subsequently, they were washed repeatedly to remove host cells and debris and stored at -80°C.

WGS sequencing and assembly

Over 3.74 million reads were generated by sequencing both ends of four genomic DNA libraries with different insert sizes (Supplementary Table 1). In total, 3.25 million reads were assembled to yield 95,265 contigs and 25,048 scaffolds, spanning 369 and 397 Mbp, respectively (Supplementary Table 2). A modified Phusion package¹ was used to assemble the reads. First, low quality regions (Phred Q≤15) were trimmed. Second, because of the substantial polymorphism of the raw data, derived from an outbreeding population of adult worms, a gap Kmer (K=17) method was used to process read groupings. Repetitive regions were screened out in the grouping process, but related reads were retained for subsequent extensions to contigs according to their

pairing information. Third, read groupings were separately assembled to establish contigs and scaffolds (Supplementary Table 2).

Completeness of the assembled genome was assessed by checking against multiple *S. japonicum* datasets. The coverage of 22 bacterial artificial chromosome (BAC) complete sequences by contigs and scaffolds was 90.1% and 99.3%, respectively. Moreover, 92.3% of 11,825 published expressed sequence tag (EST) clusters and 97.7% of 20,478 full length cDNAs (flcDNAs) (comprising 3,077 flcDNAs from public databases^{2,3} and 17,401 generated in this study), were found in the genome contigs and scaffolds. These results indicated that the draft sequence covered most (>90%) of the *S. japonicum* genome.

Construction and sequencing of flcDNA and PET libraries

Total RNAs from *S. japonicum* adults and eggs were isolated using Trizol (Invitrogen), after which messenger RNA was purified using the Poly(A) Purist™ mRNA Purification Kit (Ambion). Two flcDNA libraries, from adults and eggs were constructed using a modified biotinylated Cap-trapper approach^{4,5}. Through sequencing of 34,030 clones, 4,534 flcDNA sequences were obtained from adult worms and 12,867 from eggs, with an average sequence length of 955 bp and maximum length of 5,742 bp. In total of 20,478 flcDNAs were used in this study, comprising the above described and 3,077 from public databases^{2,3}. Based on the flcDNA clones, GIS-PET libraries were constructed as described⁶. A total of 57,600 clone sequences from the two GIS-PET libraries were determined and 688,210 PETs were extracted using PET-Tool⁷,

which represented 307,056 non-redundant PET sequences. With criteria of consistent orientation and reasonable distance (<300 kb), 132,211 of these sequences were mapped to non-repetitive regions including 62,627 PETs to a single scaffold and the remaining 69,584 PETs across scaffolds.

Prediction and integration of protein coding genes

The final prediction set was composed from two ExonHunter runs. Firstly, gene predictions were undertaken on the whole genome level. Secondly, additional genes were predicted by ExonHunter in a single gene mode in regions that were spanned by PETs or ESTs but did not contain any predictions in the first run.

The ExonHunter prediction was evaluated by comparison with available experimental datasets. Overall, 10,739 (85%) of 12,657 genes predicted from the genomic data were supported using this approach (Supplementary Fig. 2). The length distribution of predicted genes agreed with the mapping of EST, flcDNA and PET data on the genome (not shown). Moreover, 18,094 (88%) of the 20,478 flcDNAs were found in the predicted genes ($E\text{-value} \leq 1e-20$).

Additionally, sensitivity and specificity of gene prediction were evaluated by Eval⁸ based on *S. japonicum* flcDNAs with near identical alignment to the genome scaffold⁹. Comparison of ExonHunter predictions to the test dataset showed 82% sensitivity and 72% specificity in predicting exact internal exon boundaries, and 91% sensitivity and 86% specificity at the nucleotide level.

Finally, the 12,657 predicted genes were merged with 11,261 putative EST-derived

CDSs (proteins) and 18,761 putative flcDNA CDSs (proteins), yielding an estimate of 13,469 protein-coding genes in *S. japonicum* for further functional analysis. Of these, 9,689 (72%) were derived from predicted genes, 2,509 (19%) from flcDNA, and 1,271 (9%) from ESTs (Supplementary Fig. 3).

Annotation

S. japonicum protein-coding genes were classified into several catalogues step by step (Supplementary Fig. 4), retrieving reliable and functional information at higher priority. Each gene was only classified into one catalogue. Labels were appended to each gene according to its catalogue, describing the reliability level of functional information that the annotation carried. First, reciprocal best hit analysis was used to build ortholog relationships between proteins of *S. japonicum* and KEGG KO groups (BLASTP, E-value $\leq 1e-10$). These proteins were marked as “KO defined protein”. Second, proteins of *S. japonicum* that could not be defined as KO proteins were searched in the Swiss-Prot databases (E-value $\leq 1e-10$, UniProtKB/Swiss-Prot Release 55.4). Those positive proteins are marked as “informative annotated protein”. Third, the proteins were scanned for protein domains by InterPro (<http://www.ebi.ac.uk/interpro/>, Perl-based version 4.3.1, iprscan_DATA_17.0 and iprscan_PTHR_DATA_14.0). Those proteins with InterPro ID assignment are marked as “domain-containing protein”. Fourth, the proteins were searched in other available protein databases of published genomes (E-value $\leq 1e-10$). These proteins are marked as “conserved hypothetical protein”. Fifth, proteins supported by *S. japonicum* EST cluster proteins and flcDNA proteins (E-value $\leq 1e-10$), proteomics data (Summed Mascot score: 30) and *S. mansoni*

EST cluster proteins (E-value $\leq 1e-10$) are marked as “expressed protein”. Finally, proteins without any annotation and supporting information are annotated as “hypothetical protein”.

Of the 13,469 proteins, 2,516 (18.7% of the total) had an ortholog relationship to 1,546 KEGG KO groups and were mapped to reference pathways (Supplementary Table 12). Of the remainder, 4,710 (35.0%) were annotated by reference to known proteins in Swiss-Prot, and 1,898 (14.1%) matched to InterPro protein domains. Finally, among genes for which no informative function could be assigned, 364 (2.7%) were supported by annotated genes in other genomes, while 3,238 (24.0%) were supported by *S. japonicum* EST or proteomic data. A small number of hypothetical *S. japonicum* genes (743, 5.5%) remain unclassifiable to date.

Gene cataloguing

Functional assignment was possible for 6,972 (52%) out of the 13,469 genes when KEGG, UniProt or InterPro supportive information was mapped to Gene Ontology (GO) categories. Comparison of *S. japonicum* (lophotrochozoans) genes in each GO category to those of *Homo sapiens* and other deuterostomes (*Danio rerio* and *Gallus gallus*), and ecdysozoans (*D. melanogaster*, *C. elegans*, and *Brugia malayi*), revealed that about 40-50% of *S. japonicum* genes had orthologs with other species, except for the functional categories of “biological adhesion” and “locomotion”, for which values were only 20-30% (Supplementary Fig. 4). Interestingly, higher vertebrates represented by *H. sapiens* shared more orthologs with schistosomes in most gene categories than did other

taxa.

Genome variation analysis

A locally developed neighborhood quality standard (NQS) based SNP pipeline was established. The quality of the SNPs derived from the WGS was first verified against the finished clonal BAC sequences which theoretically contain no SNPs, and then was tested by a reference approach using the polymorphism detection tool, ssahaSNP (www.sanger.ac.uk/Software/analysis/ssahaSNP). Both approaches yielded consistent results.

Pathway mapping

Our genome-based annotation efforts doubled the number of schistosome genes with a KO assignment, and 25% more pathway units were recognized than previously estimated based on EST data (Supplementary Table 12). Of the 13,469 protein-coding genes in *S. japonicum*, 2,982 hit 1,512 KOs in the reference pathways. Among these genes, 2,005 had clear ortholog relationships to 1,208 (21.4%) KOs, indicating the reliability of the pathway mapping. As a comparison, 2,648 *D. melanogaster* genes were assigned into 1,267 (22.5%) KO in the KEGG released pathways. Although the automatic mapping was comprehensive, we endeavoured to enhance the pathway coverage by merging and then manually checking all results by backward searching of KEGG pathway units in the *S. japonicum* genome scaffolds. We also used those shotgun reads which were not yet placed in the current assembly, and those EST/cDNA not yet considered in the estimate of 13,469 protein-coding genes.

Carbohydrate metabolism.

The glycolysis/gluconeogenesis, tricarboxylic acid (TCA) cycle, pentose phosphate and glycogen metabolism pathways are all complete in *S. japonicum* (Supplementary Figs. 8.1-8.4). A fully functional mitochondrial electron transport system and oxidative phosphorylation pathway (Supplementary Fig. 8.5) is present for energy synthesis. Therefore, *S. japonicum* can utilize carbohydrates including glucose as energy/carbon sources to generate ATP, reducing power and essential metabolic intermediates for amino acid and nucleotide biosynthesis, and can also synthesize glycogen for energy storage and utilization¹⁰. This ability ensures the basic energy supply and the operation of the metabolic machinery required for survival, not only in its definitive and intermediate hosts, but also during its free-living aquatic phases as miracidia and cercariae.

Amino acid metabolism.

S. japonicum is unable to synthesize any of the following 11 amino acids - phenylalanine, tryptophan, methionine, lysine, threonine, leucine, valine, isoleucine, and histidine, arginine and tyrosine. However, all of twenty aminoacyl-tRNA synthetases were identified in the *S. japonicum* genome; these are required for the first stage of protein biosynthesis (Supplementary Figs. 9.1-9.10). The incompleteness of the urea cycle in *S. japonicum* (Supplementary Fig. 9.5) suggests that the excretion of nitrogen metabolites is not urea-based.

Purine and pyrimidine metabolism.

As with other parasitic helminths, *S. japonicum* cannot synthesize purines *de novo*¹¹. Here, we provide definitive evidence that 9 of the 10 genes required to anabolize inosine monophosphate (IMP) from phosphoribosyl pyrophosphate (PRPP) are absent; only adenylosuccinate lyase is present¹² (K01756, EC 4.3.2.2) (Supplementary Fig. 10.1). However, the presence of adenine phosphoribosyl transferase (K000759, [EC:2.4.2.7]), hypoxanthine phosphoribosyl transferase (K000760, [EC:2.4.2.8]), and purine-nucleoside phosphorylase (K03783, [EC:2.4.2.1]), demonstrates that the purine salvage pathway, which converts the purine bases to nucleosides and nucleotides, is intact. In contrast, as anticipated, all the enzymes for pyrimidine *de novo* synthesis from glutamine to uridine monophosphate (UMP) are present in the genome (Supplementary Fig. 10.2). It is noteworthy that similar nucleotide metabolism patterns exist in protozoan (*Trypanosoma*, *Plasmodium* and *Leishmania*) and metazoan (*B. malayi*) parasites, where *de novo* biosynthesis of purines is defective but *de novo* pyrimidine synthesis is intact.

Fatty acid metabolism.

Schistosomes are unable to synthesize fatty acids or steroids *de novo*¹⁰. The genome annotation for *S. japonicum* revealed that pathways for fatty acid biosynthesis and metabolism are incomplete. The only enzyme identified in the fatty acid biosynthesis cycle was 3-oxoacyl-[acyl-carrier-protein] synthase II (K09458, [EC:2.3.1.179]). With regard to the fatty acid β -oxidation cycle, 3-hydroxyacyl CoA dehydrogenase (K00022, [EC:1.1.1.35]) and enoyl-CoA hydratase (K01692, [EC:4.2.1.17]) could potentially contribute to fatty acid modification, but *S. japonicum*

cannot totally degrade fatty acids (Supplementary Figs. 11.1-11.2). It is noteworthy that fatty acid β -oxidation is also incomplete in *Plasmodium*, though the pathway is functional in *Leishmania* and *Trypanosoma*¹³. The ability of schistosomes to generate glycerophospholipids and triglycerides from precursors obtained from their hosts was validated with the fully annotated pathways in the *S. japonicum* genome (Supplementary Fig. 11.3).

Steroid biosynthesis.

Sterols are not synthesized *de novo* in schistosomes, but an active mevalonate pathway is present. The machinery to generate the key intermediate for cholesterol synthesis, farnesyl pyrophosphate, from acetyl-CoA and the synthesis of (S)-3-hydroxy-3-methylglutaryl-CoA is present in *S. japonicum*. Mevalonate, 5-phosphomevalonate and 5-diphosphomevalonate can be generated by three downstream enzymes, hydroxymethylglutaryl-CoA reductase (K00021, [EC:1.1.1.34]), mevalonate kinase (K00869, [EC:2.7.1.36]), and phosphomevalonate kinase (K00938, [EC:2.7.4.2]). Dimethylallyl pyrophosphate and δ -3-isopentenyl pyrophosphate are available for synthesizing geranyl pyrophosphate and then farnesyl pyrophosphate, catalyzed by prenyl transferases (K00787 and K00804, [EC 2.5.1.1, EC 2.5.1.10, EC 2.5.1.29]). However, downstream of farnesyl-pyrophosphate, no cholesterol is generated as the branch *via* squalene is inoperative in parasitic flatworms¹⁴; instead, the metabolic pathway is oriented towards synthesis of Coenzyme Q¹⁵. The presence of sterol *O*-acyltransferase (K00637 [EC:3.1.1.13]), and cholesterol esterase (K01052 [EC:2.3.1.26]) in *S. japonicum* indicates that cholesterol can be converted from

cholesterol ester provided by the host (Supplementary Figs. 11.4).

S. japonicum is dependent on the host for fatty acids and cholesterol acquired from blood and plasma, so we expected to find genes involved in lipid transportation, absorption, and conversion into other metabolites such as phospholipids and triacylglycerols. Indeed, genes coding for apolipoproteins (APOE), low density lipoprotein receptor (LDLR)¹⁶, scavenger receptor, fatty acid binding protein (FABP), ABC transporter family members, ABCA1, ABCG1 and cholesterol esterase were all identified (Supplementary Table 13). Considering the necessity for a balanced supply of different kinds of nucleotides, the identification of the enzymes thymidine kinase (EC 2.7.1.21), hypoxanthine/guanine phosphoribosyl transferase (HGPRT, EC 2.4.2.8), and adenine phosphoribosyl transferase (APRT, EC 2.4.2.7) involved in base and nucleotide interchange among different forms of purine or pyrimidine, as well as recognition of nucleoside salvage-related equilibrative nucleoside transporter¹⁷ (Sjp_0059970, K03323), is significant (Supplementary Figs. 10.1-10.2).

Gene loss and duplication

In *S. japonicum* 3,728 domains were recognized in 8,292 (61.6%) of the total 13,469 proteins coding genes. In *H. sapiens*, *G. gallus*, *D. rerio*, *N. vectensis*, *D. melanogaster*, *A. gambiae* and *C. elegans*, a total of 7,488 domains were found, of which 3,654 were shared with *S. japonicum* and 3,834 were not found in schistosomes (Fig. 2a). In the 3,654 shared domains, 2,993 were shared by all other species studied, 450 were shared by any three taxa of vertebrates, insects, nematode, and sea anemone,

and so on ,133 by two, 78 by one. Of the 3,834 domains absent from *S. japonicum*, 582 were shared by all four taxa, 558 by three, 620 by two, and 2,074 were unique to individual taxa (Supplementary Fig. 6). As an estimation, 1,140 domains were shared by more than three taxa of vertebrates, insects, nematodes, and sea anemone, thus implied as lost domains in *S. japonicum*.

Numbers of genes containing corresponding domains of the analyzed genomes were evaluated by Chi square test (and Fisher exact test if the expected number was less than 5) to detect those gene families that have experienced significant changes. In total, 785 gene families were detected to have significant variation of gene numbers. Among the most expanded gene families in schistosomes, leishmanolysin (major surface protease, MSP, also called gp63, Supplementary Table 8), a member of the metallopeptidase M8 family, has 12 putative family members in *S. japonicum* but there is only one in human, fruit fly, nematode, etc. A BLAST search using *S. japonicum* leishmanolysin proteins in the released genome data for the free-living flatworm *Schmidtea* returned 3 hits (E-value $\leq$ 1e-20, mk4.019481.00.01, mk4.002584.02.01, and mk4.019252.00.01). This family is known to be involved not only in host entry¹⁸⁻²⁰ but is also implicated in various processes during all stages of the life cycle in the protistan parasite *Leishmania*^{18,21,22}. Domain analysis suggested at least 6 putative members in *S. japonicum* possessed a realistic gp63-like structure, N-linked glycosylation sites, a protease family M8 domain, and a zinc-binding active site of the Peptidase M family, thus it is likely that they represent functional proteases. Two members of this family were supported by flcDNA data and three by proteomic data.

Moreover, there are 6 copies of the proteinase inhibitor I14/I15 (hirudin/antistatin, IPR011061) in *S. japonicum*, but only one copy in most of the other species listed in Supplementary Table 9. In leeches, this kind of inhibitor is injected at the bite as an anti-coagulant during feeding^{23, 24}.

References

1. Mullikin, J. C. & Ning, Z. The phusion assembler. *Genome research* **13**, 81-90 (2003).
2. Hu, W. *et al.* Evolutionary and biomedical implications of a *Schistosoma japonicum* complementary DNA resource. *Nat Genet* **35**, 139-147 (2003).
3. Liu, F. *et al.* New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathog* **2**, e29 (2006).
4. Seki, M., Carninci, P., Nishiyama, Y., Hayashizaki, Y. & Shinozaki, K. High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated CAP trapper. *Plant J* **15**, 707-720 (1998).
5. Wei, C. L. *et al.* 5' Long serial analysis of gene expression (LongSAGE) and 3' LongSAGE for transcriptome characterization and genome annotation. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11701-11706 (2004).
6. Ng, P. *et al.* Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nat Methods* **2**, 105-111 (2005).
7. Chiu, K. P. *et al.* PET-Tool: a software suite for comprehensive processing and managing of Paired-End diTag (PET) sequence data. *BMC bioinformatics* **7**, 390 (2006).
8. Keibler, E. & Brent, M. R. Eval: a software package for analysis of genome annotations. *BMC Bioinformatics* **4**, 50 (2003).
9. Brejova, B. *et al.* Finding genes in *Schistosoma japonicum*: annotating novel genomes with help of extrinsic evidence. *Nucleic Acids Res* (2009).
10. Saz, H. J. Comparative energy metabolisms of some parasitic helminths. *J Parasitol* **56**, 634-642 (1970).
11. Berens RL, K. E. a. M. J. *Purine and pyrimidine metabolism, in Biochemistry and Molecular Biology of Parasites (Marr JJ and Müller M eds)*. (Academic Press, New York, 1995).
12. Dovey, H. F., McKerrow, J. H. & Wang, C. C. Purine salvage in *Schistosoma mansoni* schistosomules. *Mol Biochem Parasitol* **11**, 157-167 (1984).
13. Brouwers, J. F., Versluis, C., van Golde, L. M. & Tielens, A. G. 5-Octadecenoic acid: evidence for a novel type of fatty acid modification in schistosomes. *Biochem J* **334** (Pt 2), 315-319 (1998).
14. van Hellemond, J. J. & Tielens, A. G. Adaptations in the lipid metabolism of the protozoan parasite *Trypanosoma brucei*. *FEBS Lett* **580**, 5552-5558 (2006).
15. Kawamukai, M. Biosynthesis, bioproduction and novel roles of ubiquinone. *J Biosci Bioeng* **94**, 511-517 (2002).

16. Fan, J. *et al.* A *Schistosoma japonicum* very low-density lipoprotein-binding protein. *Int J Biochem Cell Biol* **35**, 1436-1451 (2003).
17. Hyde, R. J., Cass, C. E., Young, J. D. & Baldwin, S. A. The ENT family of eukaryote nucleoside and nucleobase transporters: recent advances in the investigation of structure/function relationships and the identification of novel isoforms. *Mol Membr Biol* **18**, 53-63 (2001).
18. McGwire, B. S., Chang, K. P. & Engman, D. M. Migration through the extracellular matrix by the parasitic protozoan *Leishmania* is enhanced by surface metalloprotease gp63. *Infect Immun* **71**, 1008-1010 (2003).
19. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575-1584 (2002).
20. Kulkarni, M. M., Jones, E. A., McMaster, W. R. & McGwire, B. S. Fibronectin binding and proteolytic degradation by *Leishmania* and effects on macrophage activation. *Infect Immun* **76**, 1738-1747 (2008).
21. Cuevas, I. C., Cazzulo, J. J. & Sanchez, D. O. gp63 homologues in *Trypanosoma cruzi*: surface antigens with metalloprotease activity and a possible role in host cell infection. *Infect Immun* **71**, 5739-5749 (2003).
22. Hsiao, C. H., Yao, C., Storlie, P., Donelson, J. E. & Wilson, M. E. The major surface protease (MSP or GP63) in the intracellular amastigote stage of *Leishmania chagasi*. *Mol Biochem Parasitol* **157**, 148-159 (2008).
23. Baskova, I. P. & Zavalova, L. L. Proteinase inhibitors from the medicinal leech *Hirudo medicinalis*. *Biochemistry (Mosc)* **66**, 703-714 (2001).
24. Salzet, M. Anticoagulants and inhibitors of platelet aggregation derived from leeches. *FEBS Lett* **492**, 187-192 (2001).

Supplementary Tables

Supplementary Table 1. Whole genome sequencing reads of *S. japonicum*

Libraries	Clone insert (kb)	Raw reads				Qualified reads ^a				Placed reads ^a			
		Reads No.	Avg. length (bp)	Total base (kb)	Cov. [*] (X)	Paired-read No.	Paired-read ratio (%)	Reads No.	Avg. length (bp)	Total base (kb)	Cov. [*] (X)	Paired-read No.	Paired-read ratio (%)
BAC	80-120	86,893	619	53,759	0.14	86,426	99.5	70,493	740	52,154	0.1	59,202	84.0
Fosmid	36-42	49,225	603	29,678	0.07	44,970	91.4	39,671	620	24,615	0.1	33,444	84.3
Plasmid-L	6-10	1,368,110	540	739,234	1.86	1,304,678	95.4	1,160,389	581	674,461	1.7	1,045,160	90.1
Plasmid-S	1.6-4	2,233,505	677	1,511,124	3.80	2,176,200	97.4	1,980,585	702	1,389,966	3.5	1,862,667	94.0
Total		3,737,733	624	2,333,795	5.87	3,612,274	96.6	3,251,138	659	2,141,119	5.4	3,000,473	92.3

Avg.: average.

^{*}Cov.: coverage, which is estimated based on the assembled genome size of 397,751,405 bp.

[#]Only qualified reads were used in the assembly after removal of vector and any probable contaminating *E. coli* and *Oryctolagus cuniculus* (rabbit) sequences.

[†]Placed reads are total reads in assembled contigs.

Supplementary Table 2. *S. japonicum* genome assembly statistics

Category	Contig		Scaffold				
	all	all	>2 kb	>10 kb	>50 kb	>100 kb	>360 kb
Contig							
Total length (bp)	369,039,322	397,751,405	381,305,335	346,162,913	272,016,489	235,680,740	117,801,217
Total No.	95,265	25,048	13,235	4,663	1,331	795	202
Average size (bp)	3,874	15,880	28,810	74,236	204,370	296,454	583,174
Longest size (bp)	92,534	1,704,383	1,704,383	1,704,383	1,704,383	1,704,383	1,704,383
N50 size (bp)	6,121	173,598	191,326	232,410	315,968	359,541	596,149
N50 No.	17,637	522	476	393	257	203	70
Coverage* (%)	92.8	100	95.9	87.0	68.4	59.3	29.6

* The coverage calculation was based on the assembled genome size of 397,751,405 bp.

Supplementary Table 3. Sequence variation in the *S. japonicum* genome

Region	Length of region (kb)	SNP No.	SNP frequency (No./kb)	Indel No.	Indel frequency (No./kb)
Total	397,751	557,739	1.40	11,291	0.028
Intergenic	275,115	362,286	1.32	7,347	0.027
Intron	108,189	174,233	1.61	3,740	0.035
Donor	/	8	/	6	/
Acceptor	/	10	/	9	/
Exon*	14,447	21,220	1.47	204	0.014
Missense	/	9,992	/	/	/
Samesense	/	11,583	/	/	/
Nonsense	/	166	/	/	/
Extension	/	44	/	/	/

* Only the predicted genes and EST/flcDNA CDSs with reliable alignments in the genome were considered as 'exon regions'. Hence, the ratio of 'Exon length' (14,447kb) in the 'Total genome' (397,751kb) is slightly smaller than the coding ratio in Table 1. The SNP number under "exon" in "missense", "samesense", "nonsense" and "extension" do not add up to the SNP number in the "exon region", because multiple SNP types may occur at one SNP site.

Supplementary Table 4. Distribution of repeats in *S. japonicum*

Repeat Family	Size (bp)	Ratio (%)
Low_complexity	2,011,687	0.51
Microsatellite	2,520,388	0.63
Minisatellite	9,815,590	2.47
Satellite	16,830,062	4.23
SINE		
<i>Sj-alpha1</i>	4,293,863	1.08
<i>Sj-alpha2</i>	3,011,537	0.76
LTR Retrotransposon		
<i>Gulliver</i>	711,318	0.18
18 Novel Sj-Retrotransposons	23,803,907	5.98
Non-LTR Retrotransposon		
<i>Sj-pido</i>	1,540,514	0.39
<i>SjR1</i>	3,358,700	0.84
<i>SjR2</i>	17,602,429	4.43
4 Novel Sj-Retrotransposons	27,743,061	6.97
3 <i>Penelope</i> -like Retrotransposons	3,938,326	0.99
Transposon	856,671	0.22
undefined repeat	41,582,296	10.45
Total repeat	159,620,349	40.13
Genome Size*	397,751,405	100

*The genome size used here was based on the scaffold size of 397 Mbp, and the distribution of repeats was based on this value.

Supplementary Table 5. Statistics for retrotransposons in the *S. japonicum* genome

Retrotransposon	Proposed Group	Total Length (bp)	Genome Portion (%)	Complete copy No.	Partial copy No.
<i>LTR Retrotransposon</i>					
<i>Gulliver</i>	fugitive	4,788	0.18	6	1,223
<i>SjCHGCS1</i>	Boudicca/Saci-3	5,249	1.45	188	5,208
<i>SjCHGCS2</i>	fugitive	4,945	0.49	40	2,062
<i>SjCHGCS3</i>	Saci5	4,451	0.31	28	1,679
<i>SjCHGCS4</i>	Saci5	4,710	0.04	10	130
<i>SjCHGCS5</i>	fugitive	2,148	0.19	79	1,184
<i>SjCHGCS6</i>	fugitive	4,839	0.28	18	2,085
<i>SjCHGCS7</i>	unknown	4,000	0.16	10	1,064
<i>SjCHGCS8</i>	BEL	4,865	0.08	3	401
<i>SjCHGCS9</i>	Saci5	4,779	0.06	13	204
<i>SjCHGCS10</i>	Boudicca/Saci-3	5,025	0.61	126	1,210
<i>SjCHGCS11</i>	Boudicca/Saci-3	6,488	0.21	4	987
<i>SjCHGCS12</i>	fugitive	4,965	0.38	20	2,236
<i>SjCHGCS13</i>	fugitive	5,099	0.16	16	856
<i>SjCHGCS14</i>	fugitive	5,211	0.08	6	294
<i>SjCHGCS15</i>	fugitive	4,711	1.10	15	8,707
<i>SjCHGCS16</i>	BEL	7,232	0.18	4	737
<i>SjCHGCS17</i>	BEL	6,960	0.10	4	389
<i>SjCHGCS18</i>	BEL	5,985	0.10	7	471

Non-LTR Retrotransposon

<i>SjR1</i>	Perere	5,009	0.84	42	3,494
<i>SjR2</i>	SR2	3,921	4.43	400	23,755
<i>Sj-pido</i>	Perere	3,564	0.39	69	1,282
<i>SjCHGCS19</i>	SR2	3,145	4.09	793	17,373
<i>SjCHGCS20</i>	SR2	3,578	2.56	223	14,801
<i>SjCHGCS21</i>	Perere-9	4,275	0.01	1	26
<i>SjCHGCS22</i>	unknown	3,768	0.32	10	2,965

Penelope-like element

<i>Sj-penelope1</i>	Cercyon	3,750	0.36	3	2,717
<i>Sj-penelope2</i>	Cercyon	2,734	0.40	34	2,989
<i>Sj-penelope3</i>	Cercyon	2,512	0.23	11	1,637

Newly defined LTR and Non-LTR retrotransposons of *S. japonicum* were designated *SjCHGCS1-SjCHGCS22*, and the novel *Penelope*-like elements were named as *Sj-penelope1* to *Sj-penelope3*.

Supplementary Table 6. *S. japonicum* ortholog relationship with other species

Species	Protein No.*	1E-10	1E-20 [#]	1E-50	1E-100
<i>D. rerio</i> (zebrafish)	35,321	4,339	3,737	2,107	799
<i>M. musculus</i> (mouse)	29,533	4,320	3,735	2,140	847
<i>H. sapiens</i> (human)	25,077	4,324	3,718	2,150	845
<i>C. familiaris</i> (dog)	19,792	4,231	3,642	2,084	817
<i>G. gallus</i> (chicken)	18,107	3,873	3,311	1,906	730
<i>N. vectensis</i> (sea anemone)	24,773	3,969	3,288	1,808	700
<i>D. melanogaster</i> (fruit fly)	14,144	3,803	3,244	1,829	719
<i>A. aegypti</i> (yellow fever mosquito)	16,789	3,771	3,185	1,807	715
<i>A. gambiae</i> (mosquito)	12,527	3,690	3,128	1,760	690
<i>D. persimilis</i> (fruit fly)	17,325	3,719	3,116	1,689	629
<i>C. intestinalis</i> (sea squirt)	14,002	3,674	3,079	1,688	667
<i>A. mellifera</i> (honey bee)	27,728	3,437	2,832	1,482	616
<i>C. elegans</i> (nematode)	20,077	3,292	2,721	1,454	547
<i>C. briggsae</i> (nematode)	19,507	3,230	2,650	1,417	534
<i>B. malayi</i> (filaria)	11,515	3,193	2,609	1,359	513
<i>B. mori</i> (silkworm)	21,302	3,347	2,582	1,198	379
<i>T. cruzi</i> (Chagas disease agent)	19,607	1,470	1,059	500	172
<i>T. brucei</i> (sleeping sickness agent)	8,711	1,382	1,018	480	171
<i>P. falciparum</i> (malaria parasite)	5,261	1,176	890	405	146

Distinct colors show different groups of selected species.

*Proteins of all reference genomes were downloaded from the KEGG database release 46 (<ftp.genome.jp/pub/kegg/release/archive/kegg/46/>).
#Protein numbers for each species and orthologs calculated under various BLAST E-value thresholds are listed. The list is sorted by an E-value of 1e-20; sorting by other E-values gives rise to minor changes only.

Supplementary Table 7. Summary of gene families in several organisms

Organism name	Gene No.	MCL Cluster No. [*]	Family No. [#]
<i>S. japonicum</i>	13,469	10,266	6,322
<i>A. gambiae</i> (mosquito)	12,527	8,303	5,126
<i>C. elegans</i> (nematode)	20,077	11,287	6,669
<i>D. melanogaster</i> (fruit fly)	14,144	8,871	5,184
<i>D. rerio</i> (zebrafish)	35,321	10,047	5,569
<i>N. vectensis</i> (sea anemone)	24,773	11,987	7,101
<i>G. gallus</i> (chicken)	18,107	9,058	5,424
<i>H. sapiens</i> (human)	25,077	11,181	6,877

^{*}Gene cluster number resulted from MCL algorithm.[#]Gene family number (merged MCL Clusters according to protein domain information).

Supplementary Table 8. Leishmanolysin gene family in *S. japonicum*

Gene ID	Length (aa)	Signal peptide (aa)	N-linked glycosylation (aa position)	Peptidase M8 (IPR001577) (aa)	Peptidase M, zinc-binding site (IPR006025) (aa)	Mascot score*
Sjp_0010240	401	21-22	34/114/157	153-394	223-232	-
Sjp_0045950	511	19-20	29/183	221-307	-	34.34
Sjp_0048370	562	-	86	35-540	193-202	2,294.37
Sjp_0055960	579	20-21	148/245	171-562	-	73.64
Sjp_0067490	510	-	21/55/142	279-778	438-447	2,619.33
Sjp_0071820	590	29-30	43/138/151	182-590	249-258	-
Sjp_0107030	305	-	98/114/120	167-305	-	52.71
Sjp_0108000	586	29-30	27/224	128-362	179-188	-
Sjp_0133220	344	35-36	84/91	34-300	-	-
Sjp_0201970	566	22-23	60/115/173	253-542	222-231	-
Sjp_0212460	526	-	5	225-735	385-394	3,925.01
Sjp_0301130	563	20-21	113/258/450	251-540	-	-

aa: Amino acid residue

*Mascot score' indicates the proteomic support information; the threshold is 30

Domain information was collected by InterPro

Supplementary Table 9. Expanded *S. japonicum* gene families compared with other taxa

Domain Id	Domain Definition*	<i>S. japonicum</i>	<i>H. sapiens</i>	<i>G. gallus</i>	<i>D. rerio</i>	<i>N. vectensis</i>	<i>D. melanogaster</i>	<i>A. gambiae</i>	<i>C. elegans</i>	All p-value#
IPR009053	Prefoldin	76	29	5	11	9	9	5	8	7.96E-105
IPR008940	Protein prenyltransferase	28	3	0	2	3	3	0	2	8.31E-64
IPR010978	tRNA-binding arm	27	3	1	3	1	3	1	2	6.24E-55
IPR010989	t-SNARE	36	17	6	16	9	8	0	7	2.00E-43
IPR001372	Dynein light chain, type 1 and 2	32	3	1	6	2	3	1	6	6.87E-42
IPR012615	Trematode eggshell synthesis	15	0	0	0	0	0	0	0	6.19E-30
IPR001577	Peptidase M8, leishmanolysin	12	1	0	0	3	1	1	1	2.13E-25
IPR011048	Cytochrome cd1-nitrite reductase-like, C-terminal haem d1	17	5	1	1	1	2	1	1	1.19E-18
IPR011001	Saposin-like	20	7	1	2	6	2	2	26	4.63E-18
IPR006015	Universal stress protein (Usp)	7	0	0	0	17	0	0	0	1.93E-10
IPR008991	Translation protein SH3-like	18	9	1	7	7	6	6	6	2.78E-10
IPR003386	Lecithin:cholesterol acyltransferase	6	2	0	0	2	2	1	1	4.35E-06
IPR011061	Proteinase inhibitor I14/I15, hirudin/antistatin	6	1	1	1	0	1	1	2	2.07E-04
IPR006314	Dyp-type peroxidase	3	0	0	0	1	0	0	0	1.07E-03
IPR000905	Peptidase M22, glycoprotease	5	2	0	1	2	2	1	2	9.50E-03
IPR000977	ATP-dependent DNA ligase	5	3	2	2	3	3	0	1	1.76E-02

*Protein family is defined by Interpro protein domains. Protein family size for each species is listed.

#Numbers of genes in corresponding gene families of the analyzed genomes were evaluated by Chi square test (and Fisher exact test if the expected number was less than 5) to statistically detect those gene families with significant change in the gene numbers among the different genomes. The nearer it is to '1', the lower the level of significance. *p*-values illustrate the protein family size variation among all listed species. The nearer it is to '1', the less significance of family size variation is detected.

Supplementary Table 10. *S. japonicum* receptors, ligands and modulators of the Hedgehog, Wnt and Notch pathways

Receptor, ligand and modulator	Predicted gene	cDNA
Notch pathway		
Notch	Sjp_0080880, Sjp_0027000, Sjp_0050520, Sjp_0114750	FSA001-P00029-B09
Serrate (Jag1)	Sjp_0046830	-
TACE (adisinintegrin and metalloproteinase domain 17)	Sjp_0025060	FSA001-P00022-E11
Delta	Sjp_0040580, Sjp_0093680	-
APH-1 (Anterior pharynx defective 1)	Sjp_0114800	SJCHGC06550
Numb	Sjp_0018410	SJCHGC01437
Wnt pathway		
Frizzled	Sjp_0032920, Sjp_0010760, Sjp_0130830	-
FRP	Sjp_0022120, Sjp_0106810	-
LPR5/6	Sjp_0132180, Sjp_0133310	-
Stmb (vang-like)	Sjp_0065870	FSE_057_26_B04
Wnt	Sjp_0048240, Sjp_0058350, Sjp_0003340	FSA_091_13_E02
Porcupine	Sjp_0043730	SJCHGC02316
Hedgehog pathway		
Patched (Ptc)	Sjp_0087770	SJCHGC07066
Hh (Hedgehog)	Sjp_0032690, Sjp_0024280	-
Smoothed	Sjp_0055380	-
Megalin	Sjp_0061930	-

Supplementary Table 11. Growth factors, cytokines, chemokines and their receptors in *S. japonicum*

Gene name	Category	Predicted gene/cDNA/EST
Endothelial cell growth factor 1	Factor	-
Stromal cell derived factor 2 , SDF2	Factor	Sjp_ 0002550
Inhibin beta A chain precursor (Activin beta-A chain)	Factor	Sjp_ 0050850
Activin D precursor	Factor	-
BMP	Factor	Sjp_ 0119760
TGF- β 1	Factor	-
Activin receptor	Receptor	Sjp_ 0000540, FSA_ 064_ 27_ C04
TGF- β receptor	Receptor	Sjp_ 0037610
Epidermal growth factor receptor (EGFR)	Receptor	Sjp_ 0078840, Sjp_ 0107810
Fibroblast growth factor receptor (FGFR)	Receptor	Sjp_ 0080570, Sjp_ 0112660
Fibroblast growth factor receptor like 1 (Ig domain)	Receptor	-
Insulin receptor	Receptor	Sjp_ 0035100
Insulin-like growth factor receptor (IGFR)	Receptor	Sjp_ 0086110, SJCHGC09248

Supplementary Table 12. KEGG pathway mapping of *S. japonicum* genes and ESTs

<i>S. japonicum</i> KEGG pathway mapping process	Number
Pathways in KEGG	237
KO categories in KEGG	10,763
KEGG orthology categories in reference pathway	5,642
Number of pathways hit by a <i>S. japonicum</i> gene	206
Number of <i>S. japonicum</i> genes that hit KO categories (with ortholog relation)	4,352 (2,516)
Number of <i>S. japonicum</i> genes that hit KO categories in reference pathways (with ortholog relation)	2,982 (2,005)
Number of KO categories hit by <i>S. japonicum</i> genes (with ortholog relation)	2,741 (1,546)
Number of KO categories in reference pathway hit by <i>S. japonicum</i> gene (with ortholog relation)	1,512 (1,208)
Number of pathways hit by <i>S. japonicum</i> ESTs	203
Number of <i>S. japonicum</i> ESTs that hit KO categories	2,283
Number of <i>S. japonicum</i> EST that hit KO categories in reference pathway	1,947
Number of KO categories hit by <i>S. japonicum</i> ESTs	1,368
Number of KO categories in reference pathway hit by <i>S. japonicum</i> ESTs	1,206

This table presents the details of KEGG KO and pathway mapping. The yellow colored rows provide a general summary of the KEGG pathway database as background information. The green rows outline the mapping details of the *S. japonicum* integrated gene set (“*S. japonicum* gene”). The blue regions list mapping details of *S. japonicum* EST clusters (“*S. japonicum* EST”). Notably, the number of KOs hit by *S. japonicum* genes increase from 1,368 (hit by previously reported EST clusters) to 2,741 (by integrated gene set in this work) with the same BLAST threshold. And ‘Number of KO categories in reference pathway’ was increased from 1,206 by EST clusters to 1,512 by integrated genes, so that 25% more pathway units were recognized using the integrated gene set. Orthologous relationship were not calculated for ESTs (numbers presented in bracket for intergrated gene set) since the incompleteness of this dataset will affect the BBH results.

Supplementary Table 13. Functional annotation and mapping of genes associated with lipid metabolism

Gene	Gene ID	Function	KO	E-Value*	Domain [#]	Domain ID [#]
LIPE	Sjp_0004410	LIPE; lipase, hormone-sensitive	K07188	1.E-55	Hormone-sensitive lipase, N-terminal	IPR010468
APOA1BP	Sjp_0006580	APOA1BP; apolipoprotein A-I binding protein	—	1.E-66	YjeF-related protein, N-terminal	IPR004443
SCARB2	Sjp_0012630	Scarb2, scavenger receptor class B, member 2, Lysosome membrane protein 2, induces an intracellular signal leading to the translocation of cholesterol from intracellular sites to the plasma membrane	—	7.E-63	CD36 antigen	IPR002159
SCARB2	Sjp_0073560	Scarb2, scavenger receptor class B, member 2, Lysosome membrane protein 2, induces an intracellular signal leading to the translocation of cholesterol from intracellular sites to the plasma membrane	—	4.E-55	CD36 antigen	IPR002159
ABCA1	Sjp_0045330	ATP-binding cassette, sub-family A (ABCI), member 1, induces an intracellular signal leading to the translocation of cholesterol from intracellular sites to the plasma membrane	K05641	0	ABC transporter related	IPR003439
ABCG1	Sjp_0034670	ATP-binding cassette, subfamily G (WHITE), member 1, mediates the transport of cholesterol from cells to high density lipoprotein (HDL).	K05679	1.E-106	ABC-2 type transporter	IPR013525
LRP1	Sjp_0132180	Low density lipoprotein-related protein 1 (alpha-2-macroglobulin receptor), Apolipoprotein E receptor	K04550	2.E-23	Low density lipoprotein-receptor, class A	IPR002172
LIPA	Sjp_0027050	Lipase A, lysosomal acid, cholesterol esterase (Wolman disease)	K01052	1.E-35	AB-hydrolase associated lipase region	IPR006693
LIP1	Sjp_0302430	Lysosomal acid lipase 1 (cholesterol esterase)	K01052	1.E-86	Alpha/beta hydrolase fold-1	IPR000073
FABP7	Sjp_0200770	FABP7 fatty acid-binding protein 7	K08756	4.E-27	Lipocalin-related protein and Bos/Can/Equ allergen	IPR000566
SLC27A1	Sjp_0005620	Solute carrier family 27 (fatty acid transporter), member 1	K08745	1.E-140	AMP-dependent synthetase and ligase	IPR000873

*BLASTP E-Value≤1e-20

#Protein domain ID and description from the Interpro database.

Supplementary Table 14. Neurotransmitter transporters and receptors of *S. japonicum*

Gene name	Predicted gene/cDNA	Similar to*
Neurotransmitter receptor		
Acetylcholine receptor	Sjp_0002750, Sjp_0003730, Sjp_0015560, Sjp_0034800, FSA_088_71_G09, FSE_095_52_D07, FSE_061_88_H11, FSA001-P00002-I07, FSE001-P00007-A19, FSE001-P00030-B04	Rat, chick, human, mouse (E-72)
Dopamine D2-like receptor	Sjp_0062910, Sjp_0039590, Sjp_0039600, FSE_171_37_E05	<i>Drosophila</i> (E-55)
Neuropeptide F receptor	FSE_131_81_A11	<i>Drosophila</i> (E-41)
Glutamate receptor	Sjp_0022600, Sjp_0094830, Sjp_0121320, Sjp_0128710, FSE_139_52_D07, FSA001-P00021-P14, Sjp_0082410	Human, mouse, bee, (E-102)
5-hydroxytryptamine receptor	Sjp_0082170, Sjp_0010170, Sjp_0071540	Planarian, rat (E-73)
Adrenergic receptors	Sjp_0039610, Sjp_0039620, Sjp_0107180, FSA001-P00011-M11	-
γ -aminobutyric-acid receptor	Sjp_0074540, Sjp_0035350, FSE_171_24_H03, FSE001-P00013-O13	-
Ryanodine receptor 44F	FSA001-P00004-D01	<i>Drosophila</i> , lobster, human (E-170)
Octopamine receptor	FSE_225_75_C10, FSA001-P00011-M11	<i>Drosophila</i> (E-70)
Neuropeptide receptor NPR1	Sjp_0083580, Sjp_0096970	Planarian (E-94)
Glycine receptor, alpha 3 subunit	Sjp_0079260	Mouse (E-62)
Glycine receptor, alpha 2 isoform 1	Sjp_0058790	Macaque (E-63)
Myokinin receptor	Sjp_0096480	Mosquito (E-22)
Cardioexcitatory receptor	Sjp_0054780	Snail (E-31)
FMRFamide receptor	Sjp_0042900	Mosquito (E-22)
Myosuppressin receptor	Sjp_0042900, Sjp_0129050	Silkworm (E-17)
Neurotransmitter transporter		
Visceral amine transporter 2 (VAT2)	Sjp_0082950, Sjp_0095450, Sjp_0096200, Sjp_0096200, Sjp_0107460, FSE_161_41_A06, FSE_152_80_H10	-

dopamine transporter	Sjp_0021470, Sjp_0110630, SJCHGC06364	-
Sodium-dependent transporter	noradrenaline Sjp_0011430	-
Serotonin transporter	Sjp_0013570, Sjp_0109350, SJCHGC06385	-
Solute carrier (neurotransmitter transporter), member 1, GABA	family 6 Sjp_0005590, FSE001-P00011-L19	-

*The E value in brackets indicates the BLAST similarity to proteins known from other species.

Supplementary Table 15. Light sensory molecules of *S. japonicum*

Query sequence							BLASTP hit in <i>S. japonicum</i>	
Accession No.	Species*	Name	Abbrev.	GO	GO annotation	Accession	Annotation	BLASTP E value
P19107	DROME	Phosrestin-1 (Phosrestin I) (Arrestin B) (Arrestin-2)	Arr2	GO:0016062	adaptation of rhodopsin mediated signaling	FSE001-P00014-O10	beta-arrestin, putative	5.00E-76
P29829	DROME	Guanine nucleotide-binding protein subunit beta 2	Gbeta76C	GO:0016059	deactivation of rhodopsin mediated signaling	AAW26990	SJCHGC05537 protein	4.00E-89
Q9VL78	DROME	FK506-binding protein 59 (Peptidyl-prolyl cis-trans isomerase) (PPIase) (Rotamase) (dFKBP59)	FKBP59	GO:0050908	detection of light stimulus during visual perception	FSE001-P00035-L15	FK506-binding protein 4/5, K09571	5.00E-48
O60840	HUMAN	Voltage-dependent L-type calcium channel subunit alpha-1F	CAC1F	GO:0050908	detection of light stimulus during visual perception	Sjp_0010120	calcium channel, voltage-dependent, L type, alpha 1D subunit	8.00E-174
Q9VJJ7	DROME	Transient receptor potential-gamma protein (TRPgamma)	trpgamma	GO:0050908	detection of light stimulus during visual perception	Sjp_0091890	transient receptor potential cation channel, subfamily C,	1.00E-119
Q7SXXW4	BRARE	Protein pob	pob	GO:0045494	photoreceptor cell maintenance	FSA001-P00012-K02	Protein pob	3.00E-85
Q96NR8	HUMAN	Retinol dehydrogenase 12 (All-trans and 9-cis retinol dehydrogenase)	RDH12	GO:0045494	photoreceptor cell maintenance	FSE001-P00003-C16	-	3.00E-32
P15372	DROME	Phosrestin-2 (Phosrestin II) (Arrestin A) (Arrestin-1)	Arr1	GO:0045494	photoreceptor cell maintenance	FSE001-P00014-O10	beta-arrestin, putative	5.00E-80
P59837	BOVIN	Retinol dehydrogenase 12 (Double substrate specificity short chain dehydrogenase/reductase 2)	RDH12	GO:0045494	photoreceptor cell maintenance	FSE001-P00027-G09	RalR1, putative Retinol dehydrogenase 11	3.00E-43

Q8BYK4	MOUSE	Retinol dehydrogenase 12	Rdh12	GO:0045494	photoreceptor cell maintenance	FSE001-P00035-C09	RalR1, putative Retinol dehydrogenase 1	3.00E-56
Q8TC12	HUMAN	Retinol dehydrogenase 11	RDH11	GO:0045494	photoreceptor cell maintenance	AAW25854	SJCHGCC03296 protein	2.00E-41
Q9QYF1	MOUSE	Retinol dehydrogenase 11	Rdh11	GO:0045494	photoreceptor cell maintenance	AAW27200	SJCHGCC01149 protein	5.00E-52
O35600	MOUSE	Retinal-specific ATP-binding cassette transporter	Abca4	GO:0045494	photoreceptor cell maintenance	Sjp_0045330	ATP-binding cassette, subfamily A (ABC1), member 1	0
Q6RUQ9	BRARE	Ift88	ift88	GO:0045494	photoreceptor cell maintenance	Sjp_0060900	Intraflagellar transport 88 homolog	0
Q9CWF6	MOUSE	Bardet-Biedl syndrome 2 protein homolog	Bbs2	GO:0045494	photoreceptor cell maintenance	Sjp_0101130	Bardet-Biedl syndrome 2 protein	6.00E-44
P34057	MOUSE	Recoverin (Cancer-associated retinopathy protein)	Rcvrn	GO:0007602	phototransduction	AAW25851	SJCHGCC05658 protein	5.00E-41
P50149	MOUSE	Guanine nucleotide-binding protein G(t), alpha-2 subunit (Transducin alpha-2 chain)	Gnat2	GO:0007602	phototransduction	ABC88001	GTP-binding protein alpha-i subunit	4.00E-116
P91645	DROME	Voltage-dependent calcium channel type A alpha-1 subunit (Cacophony protein)	cac	GO:0007602	phototransduction	Sjp_0010120	calcium channel, voltage-dependent, L type, alpha 1D subunit	1.00E-141
P20612	MOUSE	Guanine nucleotide-binding protein G(t), alpha-1 subunit (Transducin alpha-1 chain)	Gnat1	GO:0007602	phototransduction	Sjp_0037230	guanine nucleotide binding protein (G protein), alpha o	2.00E-109
Q4R1I4	BRARE	Melanopsin (Opsin-4) (amphi-MOP)	OPN4	GO:0007602	phototransduction	Sjp_0082650	polypeptide opsin 4 (melanopsin)	9.00E-59

Q13507	HUMAN	Short transient receptor potential channel 3 (TrpC3) (Htrp-3) (Htrp3)	TRPC3	GO:0007602	phototransduction	Sjp_0091890	transient receptor potential cation channel, subfamily C, opsin 4 (melanopsin)	5.00E-62
P04950	DROME	Opsin Rh3 (Inner R7 photoreceptor cells opsin)	Rh3	GO:0007604	phototransduction, UV	Sjp_0082650	aryl-hydrocarbon receptor-interacting protein 254, putative	5.00E-45
Q924K1	MOUSE	Aryl-hydrocarbon-interacting protein-like 1	Aip1l	GO:0007603	phototransduction, visible light	FSE001-P00024-N19	protein-coupled receptor kinase 3',5'-cyclic-nucleotide phosphodiesterase	3.00E-59
Q1XHL8	BRARE	G protein-coupled receptor kinase 7-1	grk7a	GO:0007603	phototransduction, visible light	Sjp_0030520	protein-coupled receptor kinase 3',5'-cyclic-nucleotide phosphodiesterase	1.00E-98
P35913	HUMAN	Rod cGMP-specific 3',5'-cyclic phosphodiesterase subunit beta precursor (GMP-PDE beta)	PDE6B	GO:0007603	phototransduction, visible light	Sjp_0078360	protein-coupled receptor kinase 3',5'-cyclic-nucleotide phosphodiesterase	2.00E-57
NP_568970	<i>Arabidopsis thaliana</i>	NDPK2 (NUCLEOSIDE DIPHOSPHATE KINASE 2)	NDPK2	GO:0009585	red, far-red light phototransduction	FSE001-P00034-D01	nme2l, nucleoside-diphosphate kinase	9.00E-44
Q55434	SYNY3	Phytochrome-like protein cph2 (Bacteriophytochrome cph2)	PHY2	GO:0009585	red, far-red light phototransduction	AAX25196	SJCHGCC07711 protein	2.00E-23
Q07093	DROME	Head-specific guanylate cyclase (Gycalpa99B)	Gycalpa99B	GO:0009586	rhodopsin mediated phototransduction	Sjp_0082080	guanylate cyclase	7.00E-32
P35359	BRARE	Rhodopsin	rho	GO:0009586	rhodopsin mediated phototransduction	Sjp_0082640	Rhodopsin	2.00E-29
P10523	HUMAN	S-arrestin (Retinal S-antigen) (48 kDa protein) (S-AG) (Rod photoreceptor arrestin)	ARRS	GO:0016056	rhodopsin mediated signaling	FSE001-P00014-O10	beta-arrestin, putative	4.00E-87
Q15835	HUMAN	Rhodopsin kinase precursor (RK) (G protein-coupled receptor kinase 1)	RK	GO:0016056	rhodopsin mediated signaling	Sjp_0030520	G protein-coupled receptor kinase	6.00E-106
P23625	DROME	Guanine nucleotide-binding protein G(q) subunit alpha	Galpa49B	GO:0016056	rhodopsin mediated signaling	Sjp_0040830	guanine nucleotide	1.00E-154

P08100	HUMAN	Rhodopsin (Opsin-2)	OPSD	GO:0016056	rhodopsin mediated signaling	Sjp_0082640	binding protein (G protein), alpha q polypeptide	1.00E-30
P91657	DROME	Opsin Rh5	Rh5	GO:0009588	UV-A, blue light phototransduction	Sjp_0082650	opsin 4 (melanopsin)	1.00E-52
Q9BVG8	HUMAN	Kinesin-like protein KIFC3	KIFC3	GO:0007601	visual perception	FSA_121_43_C06	KIF3B, kinesin family member 3B, (EC 3.6.4.4)	7.00E-61
Q92696	HUMAN	Geranylgeranyl transferase type-2 alpha subunit (Rab GG transferase alpha) (Rab GGTase alpha)	PGTA	GO:0007601	visual perception	FSA001-P00005-N24	-	4.00E-78
O76090	HUMAN	Bestrophin-1 (Vitelliform macular dystrophy protein 2) (TU15B)	BEST1	GO:0007601	visual perception	FSA001-P00007-N05	Bestrophin-3 (Vitelliform macular dystrophy 2-like protein 3)	3.00E-93
O75911	HUMAN	Short-chain dehydrogenase/reductase 3) (retSDR1) (DD83.1)	DHRS3	GO:0007601	visual perception	FSA001-P00033-E07	SDR1, SDR1 protein, (EC 1.1.-.-)	1.00E-37
P26374	HUMAN	Rab proteins geranylgeranyltransferase component A 2 (Rab escort protein 2) (REP-2)	RAE2	GO:0007601	visual perception	FSE001-P00001-I18	Rab proteins geranylgeranyltransferase component A 2	7.00E-36
P36575	HUMAN	Arrestin-C (Cone arrestin) (cArr) (X-arrestin) (C-arrestin)	ARRC	GO:0007601	visual perception	FSE001-P00014-O10	beta-arrestin, putative	5.00E-92
Q8CC27	MOUSE	Voltage-dependent L-type calcium channel subunit beta-2 (CAB2)	Cacnb2	GO:0007601	visual perception	FSE001-P00019-P13	CACNB2, calcium channel, voltage-dependent, beta 2	3.00E-129
Q00169	HUMAN	Phosphatidylinositol transfer protein alpha isoform (PtdInsTP)	PIPNA	GO:0007601	visual perception	FSE001-P00026-F01	subunit, K04863 Phosphatidylinositol transfer	6.00E-90

O15492	HUMAN	(PI-TP-alpha) Regulator of G-protein signaling 16 (RGS16) (RGS-R) (A28-RGS14P)	RGS16	GO:0007601	visual perception	AAW26006	protein alpha isoform SJCHGC09168 protein	1.00E-40
O88531	MOUSE	Palmitoyl-protein thioesterase 1 precursor (PPT-1) (Palmitoyl-protein hydrolase 1) F-box-like/WD repeat protein	Ppt1	GO:0007601	visual perception	AAW26730	SJCHGC04223 protein	4.00E-73
Q8BHI5	MOUSE	TBL1XR1 (Transducin beta-like 1X-related protein 1)	Tbl1xr1	GO:0007601	visual perception	AAW27370	SJCHGC05548 protein	5.00E-140
P11488	HUMAN	Guanine nucleotide-binding protein G(t), alpha-1 subunit (Transducin alpha-1 chain)	GNAT1	GO:0007601	visual perception	ABC88001	GTP-binding protein alpha-i subunit	2.00E-115
O43636	HUMAN	Rod photoreceptor CNG-channel beta subunit	O43636	GO:0007601	visual perception	Sjp_0002290	cyclic nucleotide gated channel, invertebrate	3.00E-96
P29973	HUMAN	cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG-1) (CNG1)	CNGA1	GO:0007601	visual perception	Sjp_0003830	Cyclic nucleotide-gated channel rod photoreceptor subunit alpha (CNG channel 3) (CNG-3) (CNG3)	3.00E-42
P26367	HUMAN	Paired box protein Pax-6 (Oculorhombin) (Aniridia type II protein)	PAX6	GO:0007601	visual perception	Sjp_0004220	paired box protein 6	2.00E-61
Q9DBX2	MOUSE	Phosducin-like protein (PHLP)	Pdcl	GO:0007601	visual perception	Sjp_0007930	phosducin Sodium/potassiu m/calcium	3.00E-42
O60721	HUMAN	Sodium/potassium/calcium exchanger 1 (Na(+)/K(+)/Ca(2+)-exchange protein 1) (Retinal rod Na-Ca+K exchanger)	NCKX1	GO:0007601	visual perception	Sjp_0008110	exchanger 4 precursor (Na(+)/K(+)/Ca(2+)-exchange	1.00E-45

Q9QZ28	MOUSE	Homeobox protein SIX6 (Sine oculis homeobox homolog 6) (Optic homeobox 2) (Six9 protein) Visual system homeobox 1 (Transcription factor VSX1) (Retinal inner nuclear layer homeobox protein) (Homeodomain protein RINX)	Six6	GO:0007601	visual perception	Sjp_0071630	Homeobox protein SIX2	2.00E-55
Q9NZR4	HUMAN		VSX1	GO:0007601	visual perception	Sjp_0072380	visual system homeobox protein 2	1.00E-21
O95343	HUMAN	Homeobox protein SIX3 (Sine oculis homeobox homolog 3)	SIX3	GO:0007601	visual perception	Sjp_0073990	Homeobox protein SIX6 (Sine oculis homeobox homolog 6) (Optic homeobox 2) (Six9 protein) 3',5'-cyclic-nucleotide phosphodiesterase	3.00E-87
P16499	HUMAN	Rod cGMP-specific 3',5'-cyclic phosphodiesterase subunit alpha (GMP-PDE alpha) (PDE V-B1)	PDE6A	GO:0007601	visual perception	Sjp_0078360	glycine receptor, alpha 1	4.00E-57
P48168	MOUSE	Glycine receptor subunit beta precursor (Glycine receptor 58 kDa subunit)	Glr3	GO:0007601	visual perception	Sjp_0079260		1.00E-65
O14718	HUMAN	Visual pigment-like receptor peropsin	OPN3	GO:0007601	visual perception	Sjp_0082640	Rhodopsin	2.00E-31
Q9Y2V3	HUMAN	Retinal homeobox protein Rx (Retina and anterior neural fold homeobox protein)	RX	GO:0007601	visual perception	Sjp_0083190	retina and anterior neural fold homeobox protein	1.00E-27
O00222	HUMAN	Metabotropic glutamate receptor 8 precursor (mGluR8)	MGR8	GO:0007601	visual perception	Sjp_0086150	Metabotropic glutamate receptor 7 precursor	4.00E-37
Q5NCH9	MOUSE	Metabotropic glutamate receptor 6 precursor (mGluR6)	Gmr6	GO:0007601	visual perception	Sjp_0088520	glutamate receptor,	8.00E-73

P53611	HUMAN	Geranylgeranyl transferase type-2 subunit beta (Rab GGTase beta)	PGTB	GO:0007601	visual perception	Sjp_0090410	metabotropic 3	1.00E-100
O00423	HUMAN	Echinoderm microtubule-associated protein-like 1 (EMAP-1) (HuEMAP-1)	EMAL1	GO:0007601	visual perception	Sjp_0091140	Geranylgeranyl transferase type-2 subunit beta (EC 2.5.1.60) Echinoderm microtubule-associated protein-like 1	1.00E-129
P78424	HUMAN	POU domain, class 6, transcription factor 2 (Retina-derived POU domain factor 1) (RPF-1)	PO6F2	GO:0007601	visual perception	Sjp_0096910	POU domain transcription factor, class 6	7.00E-23
Q02846	HUMAN	Retinal guanylyl cyclase 1 precursor (Guanylate cyclase 2D, retinal) (RETGC-1) (Rod outer segment membrane guanylate cyclase) (ROS-GC)	GUC2D	GO:0007601	visual perception	Sjp_0108870	Retinal guanylyl cyclase 1 precursor (EC 4.6.1.2)	2.00E-61

***Species name:**

BOVIN *Bos taurus*
 BRABE *Branchiostoma belcheri*
 BRARE *Danio rerio (Zebrafish) (BrachyDanio rerio)*
 DROME *Drosophila melanogaster*
 HUMAN *Homo sapiens*
 MOUSE *Mus musculus*
 SYNY3 *Synechocystis sp. (strain PCC 6803)*

Supplementary Table 16. Chemical sensory molecules of *S. japonicum*

Query sequence			BLASTP hit in <i>S. japonicum</i>					
Accession No.	Species*	Name	Abbrev.	GO	GO annotation	Accession	Annotation	BLASTP E value
P29274	HUMAN	Adenosine A2a receptor	AA2AR	GO:0007600	sensory perception	Sjp_0039620	Alpha-1A adrenergic receptor (Alpha 1A-adrenoceptor)	4.00E-29
P62874	MOUSE	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta 1 (Transducin beta chain 1)	Gnb1	GO:0050909	sensory perception of taste	AAW26990	SJCHGC05537 protein	2.00E-178
P08510	DROME	Potassium voltage-gated channel protein Shaker	Sh	GO:0050909	sensory perception of taste	Sjp_0069340	potassium voltage-gated channel, Shaker-related subfamily A, member	1.00E-147
P29617	DROME	Protein prospero	pros	GO:0050909	sensory perception of taste	Sjp_0054530	Protein prospero	2.00E-34
Q7RTX1	HUMAN	Taste receptor type 1 member 1 precursor (G-protein coupled receptor 70)	TS1R1	GO:0050917	sensory perception of umami taste	Sjp_0088520	glutamate receptor, metabotropic 3	8.00E-33

*Species name:

DROME *Drosophila melanogaster*
HUMAN *Homo sapiens*
MOUSE *Mus musculus*

Supplementary Table 17. Sound sensory molecules of *S. japonicum*

Query sequence			BLASTP hit in <i>S. japonicum</i>					
Accession No.	Species*	Name	Abbrev.	GO	GO annotation	Accession	Annotation	BLASTP E value
Q9Y5L4	HUMAN	Mitochondrial import inner membrane translocase subunit Tim13	TIM13	GO:0007605	sensory perception of sound	FSA001-P00013-P20	-	1.00E-24
Q9HBG4	HUMAN	Vacuolar proton translocating ATPase 116 kDa subunit a isoform 4 (V-ATPase 116 kDa isoform a4)	VPP4	GO:0007605	sensory perception of sound	FSA001-P00020-K20	Vha100-1, H ⁺ -transporting ATPase	0
Q9Y664	HUMAN	Kaptn (Actin-associated protein 2E4)	KPTN	GO:0007605	sensory perception of sound	FSA001-P00032-M08	Kaptn	2.00E-31
O88342	MOUSE	WD repeat protein 1 (Actin-interacting protein 1) (AIP1)	Wdr1	GO:0007605	sensory perception of sound	FSE_045_67_C09	AIP1, actin-interacting protein 1, putative	1.00E-115
Q9DBF1	MOUSE	Alpha-aminoadipic semialdehyde dehydrogenase (Alpha-AASA dehydrogenase) (Antiquitin-1)	Aldh7a1	GO:0007605	sensory perception of sound	FSE_156_41_A06	-	5.00E-52
Q6RFG0	BRARE	Rab escort protein 1	chm	GO:0050910	detection of mechanical stimulus during sensory perception of sound	FSE001-P00001-I18	Rab proteins geranylgeranyltransferase component A 2	4.00E-43
P20060	MOUSE	Beta-hexosaminidase beta chain precursor (N-acetyl-beta-glucosaminidase) (Beta-N-acetylhexosaminidase) (Hexosaminidase B)	Hexb	GO:0007605	sensory perception of sound	AAW26910	SJCHGC06873 protein	2.00E-121

O60907	HUMAN	F-box-like/WD repeat protein TBL1X (Transducin beta-like 1X protein) (Transducin-beta-like 1, X-linked) (SMAP55)	TBL1X	GO:0007605	sensory perception of sound	AAX27370	SJCHGC05548 protein	2.00E-140
P23760	HUMAN	Paired box protein Pax-3 (HUP2)	PAX3	GO:0007605	sensory perception of sound	Sjp_0004220	paired box protein 6	7.00E-41
Q8R4F0	MOUSE	Mucolipin-3	Mcoln3	GO:0007605	sensory perception of sound	Sjp_0004890	mucolipin, invertebrate	3.00E-38
O35625	MOUSE	Axin-1 (Axis inhibition protein 1) (Fused protein) POU domain, class 3, transcription factor 4 (Brain-specific homeobox/POU domain protein 4) (Brain-4) (Brn-4)	Axin1	GO:0007605	sensory perception of sound	Sjp_0004930	axin 1	4.00E-21
P62515	MOUSE	(Brain-specific homeobox/POU domain protein 4) (Brain-4) (Brn-4)	Pou3f4	GO:0007605	sensory perception of sound	Sjp_0006200	POU domain transcription factor, class 3	1.00E-36
Q01814	HUMAN	Plasma membrane calcium-transporting ATPase 2 (PMCA2)	AT2B2	GO:0007605	sensory perception of sound	Sjp_0009630	Ca2+ transporting ATPase, plasma membrane	0
Q9R0K7	MOUSE	Plasma membrane calcium-transporting ATPase 2 (PMCA2)	Atp2b2	GO:0050910	detection of mechanical stimulus during sensory perception of sound	Sjp_0009630	Ca2+ transporting ATPase, plasma membrane	1.00E-160
Q99246	MOUSE	Voltage-dependent L-type calcium channel subunit alpha-1D	Cacna1d	GO:0007605	sensory perception of sound	Sjp_0010110	calcium channel, voltage-dependent, L type, alpha 1F subunit	4.00E-68
Q6QRQ2	BRARE	L-type calcium channel pore subunit isoform 1.3a	cacna1d	GO:0050910	detection of mechanical stimulus during sensory perception of sound	Sjp_0010120	calcium channel, voltage-dependent, L type, alpha 1D subunit	4.00E-171
Q9VMR4	DROME	Mechanosensory transduction channel NOMPC	nompC	GO:0050910	detection of mechanical stimulus during sensory perception of sound	Sjp_0016540	Ankyrin-2 (Brain ankyrin) (Ankyrin-B) (Ankyrin, nonerythroid)	4.00E-30

P68500	MOUSE	Contactin-5 precursor (Neural recognition molecule NB-2)	Cntn5	GO:0007605	sensory perception of sound	Sjp_0021090	neurofascin	1.00E-52
O60610	HUMAN	Protein diaphanous homolog 1 (Diaphanous-related formin-1) (DRF1)	DIAP1	GO:0007605	sensory perception of sound	Sjp_0022610	diaphanous 1	6.00E-115
P15313	HUMAN	Vacuolar ATP synthase subunit B, kidney isoform (V-ATPase subunit B 1)	VATB1	GO:0007605	sensory perception of sound	Sjp_0031510	V-type H ⁺ -transporting ATPase subunit B	0
Q9JK97	MOUSE	Potassium voltage-gated channel subfamily KQT member 4 (KQT-like 4)	Kcnq4	GO:0007605	sensory perception of sound	Sjp_0035900	potassium voltage-gated channel, KQT-like subfamily, member 2	3.00E-33
P58743	HUMAN	Prestin (Solute carrier family 26 member 5)	S26A5	GO:0007605	sensory perception of sound	Sjp_0037090	Chloride anion exchanger (Protein DRA) (Down-regulated in adenoma) (Solute carrier family 26 member 3)	8.00E-61
Q99502	HUMAN	Eyes absent homolog 1	EYA1	GO:0007605	sensory perception of sound	Sjp_0045910	protein-tyrosine phosphatase	3.00E-64
Q9Y219	HUMAN	Jagged-2 precursor (Jagged2) (HJ2)	JAG2	GO:0007605	sensory perception of sound	Sjp_0046830	jagged	5.00E-103
P37244	MOUSE	Thyroid hormone receptor beta-2	Thrb	GO:0007605	sensory perception of sound	Sjp_0050210	nuclear receptor, subfamily 1, group A, member 2	2.00E-35
P63157	MOUSE	BarH-like 1 homeobox protein (Bar-class homeodomain protein MBH2) (BarH-related homeobox protein 1)	Barhl1	GO:0007605	sensory perception of sound	Sjp_0066280	BarH-like	5.00E-32
P55012	MOUSE	Solute carrier family 12 member 2 (Bumetanide-sensitive sodium-(potassium)-chloride cotransporter 1)	Slc12a2	GO:0050910	detection of mechanical stimulus during sensory perception of sound	Sjp_0074520	Solute carrier family 12 member 9	1.00E-60

Q08460	MOUSE	Calcium-activated potassium channel subunit alpha 1 (Maxi K channel)	Kcnma1	GO:0007605	sensory perception of sound	Sjp_0074820	potassium large conductance calcium-activated channel, subfamily M ₁	9.00E-113
Q9QYE5	MOUSE	Jagged-2 precursor (Jagged2)	Jag2	GO:0007605	sensory perception of sound	Sjp_0080880	Notch	7.00E-81
P70323	MOUSE	T-box transcription factor TBX1 (T-box protein 1) (Testis-specific T-box protein)	Tbx1	GO:0007605	sensory perception of sound	Sjp_0086540	T-box transcription factor TBX1 (T-box protein 1) (Testis-specific T-box protein)	5.00E-66
O95834	HUMAN	Echinoderm microtubule-associated protein-like 2 (EMAP-2) (HuEMAP-2)	EMAL2	GO:0007605	sensory perception of sound	Sjp_0091140	Echinoderm microtubule-associated protein-like 1	5.00E-122
P51787	HUMAN	Potassium voltage-gated channel subfamily KQT member 1 (KQT-like 1)	KCNQ1	GO:0007605	sensory perception of sound	Sjp_0095320	potassium voltage-gated channel, KQT-like subfamily, member 1	8.00E-27
P50586	MOUSE	Tubby protein	Tub	GO:0007605	sensory perception of sound	Sjp_0101840	Tubby protein homolog	4.00E-32
O43511	HUMAN	Pendrin (Sodium-independent chloride/iodide transporter) (Solute carrier family 26 member 4)	S26A4	GO:0007605	sensory perception of sound	Sjp_0114830	Chloride anion exchanger	3.00E-31
Q15319	HUMAN	POU domain, class 4, transcription factor 3 (Brain-specific homeobox/POU domain protein 3C) (Brn-3C)	PO4F3	GO:0007605	sensory perception of sound	Sjp_0130780	POU domain transcription factor, class 4	2.00E-34

***Species name:**BRARE *Danio rerio (Zebrafish) (BrachyDanio rerio)*DROME *Drosophila melanogaster*HUMAN *Homo sapiens*

MOUSE

Mus musculus

Supplementary Table 18. Olfactory sensory molecules of *S. japonicum*

Query sequence		BLASTP hit in <i>S. japonicum</i>					
Accession No.	Species*	Name	Abbrev.	GO	GO annotation	Accession	BLASTP E value
Q8VHH7	MOUSE	Adenylate cyclase type 3 (AC-III) (AC3)	Adcy3	GO:0007608	sensory perception of smell	Sjp_0078110	8.00E-41
Q8C1Z7	MOUSE	Bardet-Biedl syndrome 4 protein homolog	Bbs4	GO:0007608	sensory perception of smell	Sjp_0015830	1.00E-56
Q16280	HUMAN	Cyclic nucleotide-gated olfactory channel (CNG channel 2)	CNGA2	GO:0007608	sensory perception of smell	Sjp_0002290	3.00E-75
Q8CGK7	MOUSE	Guanine nucleotide-binding protein G(olf) subunit alpha (Adenylate cyclase-stimulating G alpha protein, olfactory type)	Gnal	GO:0007608	sensory perception of smell	AAW27527	7.00E-144
P63092	HUMAN	Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	GNAS2	GO:0007608	sensory perception of smell	Sjp_0032600	7.00E-44
Q02280	DROME	Potassium voltage-gated channel protein eag (Ether-a-go-go protein)	eag	GO:0007608	sensory perception of smell	Sjp_0027370	0
P22270	DROME	Tyramine/octopamine receptor precursor (Tyr/Oct-Dro)	TyrR	GO:0007608	sensory perception of smell	FSE_225_75_C10	1.00E-41

*Species name:

- DROME *Drosophila melanogaster*
- HUMAN *Homo sapiens*
- MOUSE *Mus musculus*

Supplementary Table 19. Mechanical/thermoception molecules of *S. japonicum*

Query sequence			BLASTP hit in <i>S. japonicum</i>					
Accession No.	Species*	Name	Abbrev.	GO	GO annotation	Accession	Annotation	BLASTP E value
Q5UM14	BRARE	TRPA2	trpa1b	GO:0050974	detection of mechanical stimulus during sensory perception	Sjp_0003070	transient receptor potential cation channel, subfamily A, member 1	1.00E-56
Q8BLA8	MOUSE	Transient receptor potential cation channel subfamily A member 1	Trpa1	GO:0050955	thermoception	Sjp_0003070	transient receptor potential cation channel, subfamily A, member 1	4.00E-57
Q8R4D5	MOUSE	Transient receptor potential cation channel subfamily M member 8	Trpm8	GO:0050955	thermoception	Sjp_0011270	transient receptor potential cation channel, subfamily M, member 3	9.00E-48

*Species name:

BRARE *Danio rerio* (Zebrafish) (*BrachyDanio rerio*)
MOUSE *Mus musculus*

Supplementary Table 20. Equilibrioception molecules of *S. japonicum*

Query sequence				BLASTP hit in <i>S. japonicum</i>				
Accession No.	Species*	Name	Abbrev.	GO	GO annotation	Accession	Annotation	BLASTP E value
Q6QQE1	BRARE	Cadherin 23	cdh23	GO:0050957	equilibrioception	Sjp_0046380	Protocadherin-11 X-linked precursor (Protocadherin-11) (Protocadherin on the X chromosome)	2.00E-50
P17208	MOUSE	POU domain, class 4, transcription factor 1 (Brain-specific homeobox/POU domain protein 3A) (Brn-3A) (Brn-3.0)	Pou4f1	GO:0051355	proprioception during equilibrioception	Sjp_0130780	POU domain transcription factor, class 4	2.00E-34
P48432	MOUSE	Transcription factor SOX-2	Sox2	GO:0050973	detection of mechanical stimulus during equilibrioception	Sjp_0090170	transcription factor SOX1/2/3/14/21 (SOX group B)	3.00E-37

*Species name:

BRARE *Danio rerio* (Zebrafish) (*BrachyDanio rerio*)

MOUSE *Mus musculus*

Supplementary Table 21. Pain sensory molecules of *S. japonicum*

Query sequence		BLASTP hit in <i>S. japonicum</i>						
Accession No.	Species *	Name	Abbrev.	GO	GO annotation	Accession	Annotation	BLASTP E value
P04354	CHICK	Calbindin (Vitamin D-dependent calcium-binding protein, avian-type)	CALB1	GO:0019233	sensory perception of pain	FSE_109_43_C06	Calbindin-32	4.00E-56
Q9QXT8	MOUSE	Calsenilin (Kv channel-interacting protein 3) (KChIP3)	Kcnip3	GO:0019233	sensory perception of pain	AAW25114	SJCHGC01061 protein	9.00E-54
Q91XM9	MOUSE	Discs large homolog 2 (Postsynaptic density protein PSD-93)	Dlg2	GO:0019233	sensory perception of pain	Sjp_0011120	guanylate kinase	1.00E-37
Q04573	MOUSE	Neuropeptide Y receptor type 1 (NPY1-R)	Npy1r	GO:0019233	sensory perception of pain	Sjp_0096970	Neuropeptide Y receptor, K04209,invertebrate	3.00E-44
Q9QYS2	MOUSE	Metabotropic glutamate receptor 3 precursor (mGluR3)	Grm3	GO:0051930	regulation of sensory perception of pain	Sjp_0088520	glutamate receptor, metabotropic 3	9.00E-82
*Species name:								
CHICK		<i>Gallus gallus</i>						
MOUSE		<i>Mus musculus</i>						

Supplementary Table 22. Hormone receptors of *S. japonicum*

Gene name	Predicted gene/cDNA	Similar to*
Receptor for hypothalamus hormone		
Thyrotropin-releasing hormone receptor	Sjp_0015340, Sjp_0131730	Mouse, human (E-39)
Corticotropin releasing hormone receptor	Smp_125420	Human, Drosophila (E-19)
Prolactin releasing hormone receptor	Sjp_0096970, Smp_118040	Dog (E-49)
Growth hormone releasing hormone receptor	Smp_125420, Sjp_0078680	Mouse (E-14), rat (E-12)
Neurotensin receptor 3	Sjp_0001900	Mouse, sea urchin (E-36)
Arginine vasopressin receptor	Smp_126890, Sjp_0039620	Drosophila (E-26, E-21)
Receptor for pituitary hormone		
Insulin receptor	Sjp_0035100, Sjp_0070030, Sjp_0080940, Sjp_0086110	Echinococcus, insect, fish (E-83)
Follicle stimulating hormone receptor	Sjp_0039620	Rat (E-12)
Luteinizing hormone receptor	Sjp_0039620	Rat (E-13)
Receptor for sex hormone		
Progesterin and adipoQ receptor family member 6	FSA001-P00009-K02, FSE001-P00026-N24, FSE001-P00001-D02	Human and mouse (E-46)
Membrane-associated progesterone receptor component 2	FSE001-P00019-D12	Human (E-35)

Estrogen-related receptor beta like 1	FSE001-P00011-A06	Frog (E-52)
Activin receptor type-1	FSA_064_27_C04	Chicken (E-41)
Receptor for adrenal gland hormone		
Adrenergic receptor	Sjp_0039620, Sjp_0039610	Planarien, mouse, fish
Glucocorticoid receptor	Sjp_0026760, Sjp_0002350	Mouse, frog (E-23)
Receptor for thyroid hormone		
Thyroid hormone receptor	Smp_134490, Smp_174260, Sjp_0050210, Sjp_0027230	Mouse, chicken, zebrafish (E-34 ~ E-36)
Somatostatin receptor	Sjp_0063930, Sjp_0012920	Mosquito (E-13)
Other receptor		
Adiponectin receptor	FSA001-P00026-A07, FSE001-P00016-J16	Fly (E-96)
Leptin receptor	FSA001-P00022-F13, FSA001-P00029-P18, FSA001-P00032-H02, FSE001-P00021-A18	Human (E-25)
Allatostatin receptor	Sjp_0112860	Cockroach (E-45)
Ecdysone receptor	FSE001-P00030-D02	zebrafish (E-10)
Cholecystokinin A receptor	Smp_118040, Sjp_0096970	Chicken (E-30)
Neuromedin U receptor	Smp_011940, Sjp_0096970	Drosophila (E-28), human (E-25)
Secretin receptor	Sjp_0078680	Rat (E-15)

*The E value in brackets indicates the BLAST similarity to known proteins in other species.

Supplementary Table 23. The predicted metallo-proteases of *S. japonicum*

Protease family	Protein name	Annotation	Predicted active site residues	Predicted metal ligands	Hit name	E-value comparing MEROPS	Annotation in MEROPS	Peptidase unit
M01	Sjp_0025870	Thyrotropin-releasing hormone-degrading ectoenzyme, putative	S37E, E126Y	N36H, Q40H, N64E	MER47667	1.6E-16	aminopeptidase A (<i>Xenopus tropicalis</i>) [M01.003]	{peptidase unit: 458-677}
M01	Sjp_0025880	Glutamyl aminopeptidase (EC 3.4.11.7)	E251, >Y	H250, H254, >E	MER53995	1.9E-22	aminopeptidase PILS (<i>Gallus gallus</i>) [M01.018]	{peptidase unit: 249-467} from GB:CAH65246
M01	Sjp_0041390	puromycin-sensitive aminopeptidase, K08776, putative	E197, Y282	H196, H200, E219	MER02746	3E-54	cytosol alanyl aminopeptidase (<i>Homo sapiens</i>) [M01.010]	{peptidase unit: 268-501} from Sw:P55786
M01	Sjp_0025860	glutamyl aminopeptidase, K01261, putative	E339, Y424	H338, H342, E361	MER21283	3.7E-66	family M1 unassigned peptidases (<i>Anopheles gambiae</i>) [M01.UPW]	{peptidase unit: 316-544} from GB:EAA44681
M01	Sjp_0022850	leukotriene-A4 hydrolase, K01254, [PATH ko00590], putative	E295, X382Y	H294, H298, E317	MER53983	3.4E-73	leukotriene A4 hydrolase (<i>Gallus gallus</i>) [M01.004]	{peptidase unit: 210-451} from GB:CAG32097
M01	Sjp_0022540	transcription initiation factor TFIID subunit D2, K03128, [PATH ko03022], putative	A271E, >Y	L270H, H274, F293E	MER26493	8.5E-37	Tata binding protein associated factor (<i>Homo sapiens</i>) [M01.972]	{peptidase unit: 268-436} from Sw:Q6P1X5
M03A	Sjp_0131320	mitochondrial intermediate peptidase, K01410, putative	E61	H60, H64, E89	MER22450	5E-36	mitochondrial intermediate peptidase (<i>Anopheles gambiae</i>) [M03.006]	{peptidase unit: 31-676} from GB:EAA04703
M03A	SJCHGC08280	SJCHGC08280, similar to mitochondrial intermediate peptidase (<i>Mus musculus</i>)	>E	>H, >H, >E	MER49611	9.30E-19	mitochondrial intermediate peptidase (<i>Brachydanio rerio</i>) [M03.006]	{peptidase unit: 37-419} from GB:XP_685496
M03A	Sjp_0015820	thimet oligopeptidase, K01392, putative	<E, <E, <E, E419	<H, <H, <E, <H, <H, T397E, H418, H422, E447	MER69077	4.5E-75	thimet oligopeptidase (<i>Monodelphis domestica</i>) [M03.001]	{peptidase unit: 3-675}
M03A	SJCHGC06674	SJCHGC06674, similar to Thimet oligopeptidase 1 (<i>Homo sapiens</i>)	R219E, >E, >E, >E	F54H, A56H, V145E, >H, >H, >E, >H, >E	MER69077	1.10E-63	thimet oligopeptidase (<i>Monodelphis domestica</i>) [M03.001]	{peptidase unit: 3-675}

M08	Sjp_0071820	leishmanolysin (EC 3.4.24.36), K01404, putative	E253, M370	H252, H256, H359	MER22457	2.7E-81	invadolysin (<i>Anopheles gambiae</i>) [M08.002]	{peptidase unit: 31-616} from GB:EAA07618
M08	Sjp_0048370	Leishmanolysin-like peptidase, putative	E197, M314	H196, H200, H303	MER22457	1E-74	invadolysin (<i>Anopheles gambiae</i>) [M08.002]	{peptidase unit: 31-616} from GB:EAA07618
M08	Sjp_0011250	Leishmanolysin-like peptidase, putative	<E, M1	<H, <H, <H	MER22457	2.4E-41	invadolysin (<i>Anopheles gambiae</i>) [M08.002]	{peptidase unit: 31-616} from GB:EAA07618
M08	Sjp_0107030	Leishmanolysin-like peptidase, putative	E166, M283	H165, H169, H272	MER22457	7.8E-35	invadolysin (<i>Anopheles gambiae</i>) [M08.002]	{peptidase unit: 31-616} from GB:EAA07618
M08	SJCHGC07540	SJCHGC07540, similar to XP_393414-similar to CG3953-PA (<i>Apis mellifera</i>)	E10, >M	H9, H13, >H	MER22457	7.00E-14	invadolysin (<i>Anopheles gambiae</i>) [M08.002]	{peptidase unit: 31-616} from GB:EAA07618
M08	Sjp_0067490	Leishmanolysin-like peptidase, putative	E442, M557	H441, H445, H546	MER11819	5.9E-75	invadolysin (<i>Drosophila melanogaster</i>) [M08.002]	{peptidase unit: 89-683} from GB:AAN71624
M08	SJCHGC03485	SJCHGC03485, weak similar to XP_344026-similar to leishmanolysin-like (metallopeptidase M8 family)	<E, M34	<H, <H, H23	MER11819	8.40E-13	invadolysin (<i>Drosophila melanogaster</i>) [M08.002]	{peptidase unit: 89-683} from GB:AAN71624
M08	SJCHGC02921	SJCHGC02921, weak similar to CG3953-PA (<i>Drosophila melanogaster</i>)	<E, <M	<H, <H, <H	MER11819	2.90E-11	invadolysin (<i>Drosophila melanogaster</i>) [M08.002]	{peptidase unit: 89-683} from GB:AAN71624
M08	SJCHGC02836	SJCHGC02836, weak similar to CG3953-PA (<i>Drosophila melanogaster</i>)	<E, <M	<H, <H, <H	MER11819	7.60E-10	invadolysin (<i>Drosophila melanogaster</i>) [M08.002]	{peptidase unit: 89-683} from GB:AAN71624
M08	Sjp_0108000	leishmanolysin, K01404, putative	E183, M303	H182, H186, H292	MER14492	5.2E-76	leishmanolysin-2 (<i>Homo sapiens</i>) [M08.003]	{peptidase unit: 51-634} from GB:CAC42883
M08	Sjp_0010240	Leishmanolysin-like peptidase, putative	E227, M342	H226, H230, H331	MER14492	1.4E-59	leishmanolysin-2 (<i>Homo sapiens</i>) [M08.003]	{peptidase unit: 51-634} from GB:CAC42883
M08	Sjp_0067500	hypothetical protein	E389, M518	H388, H392, H507	MER14492	3.9E-43	leishmanolysin-2 (<i>Homo sapiens</i>) [M08.003]	{peptidase unit: 51-634} from GB:CAC42883
M08	Sjp_0133220	Leishmanolysin precursor (EC 3.4.24.36)	<E, M72	<H, <H, H61	MER14492	5.5E-41	leishmanolysin-2 (<i>Homo sapiens</i>) [M08.003]	{peptidase unit: 51-634} from GB:CAC42883
M08	Sjp_0045950	Leishmanolysin-like peptidase, putative	<E, M254	<H, <H, H246	MER14492	5.2E-34	leishmanolysin-2 (<i>Homo sapiens</i>) [M08.003]	{peptidase unit: 51-634} from GB:CAC42883

M10A	Sjp_0075540	matrix metalloproteinase-16 (membrane-inserted), K07996, putative	E55	H54, H58, H64	MER01091	2.2E-25	envelysin (<i>Paracentrotus lividus</i>) [M10.010] LALP peptidase { <i>Loxosceles intermedia</i> }} (<i>Loxosceles intermedia</i>) [M12.333]	{peptidase unit: 163-347} from Sw:P22757
M12A	Sjp_0029490	SPAN protein precursor, putative	E23	H22, H26, H32	MER91011	2.1E-14		{peptidase unit: 58-246} from GB:ABK20019
M12B	Sjp_0025060	a disintegrin and metalloproteinase domain I7,K06059,[PATH ko04330,PATH ko05120], putative	E377	H376, H380, H386	MER03094	4.4E-25	ADAM17 peptidase (<i>Homo sapiens</i>) [M12.217]	{peptidase unit: 205-474} from Sw:P78536
M12B	SICHGC06709	SICHGC06709, similar to ATS3_HUMAN ADAMTS-3 precursor	>E	>H, >H, >H	MER16700	2.90E-17	ADAMTS14 peptidase (<i>Homo sapiens</i>) [M12.024]	{peptidase unit: 254-460} from Sw:Q8WXS8
M12B	Sjp_0011910	ADAM metalloproteinase with thrombospondin type 1 motif, 9,K08624, putative	<E	<H, <H, <H	MER54012	1.9E-36	ADAMTS7 peptidase (<i>Gallus gallus</i>) [M12.231]	{peptidase unit: 373-622} from GB:XP_413938
M12B	Sjp_0092270	ADAM metalloproteinase with thrombospondin type 1 motif, 2,K08618, putative	E54	H53, H57, H63	MER54012	2.9E-24	ADAMTS7 peptidase (<i>Gallus gallus</i>) [M12.231]	{peptidase unit: 373-622} from GB:XP_413938
M12B	Sjp_0042280	a disintegrin and metalloproteinase domain 33,K08616, putative	E168	H167, H171, H177	MER38665	1.5E-22	subfamily M12B unassigned peptidases (<i>Anopheles gambiae</i>) [M12.UPB]	{peptidase unit: 90-296} from GB:EAA45314
M13	Sjp_0102100	Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), putative	>E, >D	>H, >H, >E	MER01056	1.3E-28	endothelin-converting enzyme 1 (<i>Bos taurus</i>) [M13.002]	{peptidase unit: 78-754} from Sw:P42891
M13	Sjp_0102090	endothelin-converting enzyme,K01415, putative	E181, D244	H180, H184, E240	MER01991	3.9E-59	endothelin-converting enzyme 2 (<i>Bos taurus</i>) [M13.003]	{peptidase unit: 111-787} from Sw:Q10711
M13	Sjp_0045990	Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), putative	<E, D62	<H, <H, E58	MER19839	1.8E-19	family M13 unassigned peptidases (<i>Anopheles gambiae</i>) [M13.UPW]	{peptidase unit: 1-685} from GB:EAA14296
M13	Sjp_0018670	Endothelin-converting enzyme-like 1 (EC 3.4.24.-) (Xce protein), putative	<E, <D	<H, <H, <E	MER19839	1.3E-14	family M13 unassigned peptidases (<i>Anopheles gambiae</i>) [M13.UPW]	{peptidase unit: 1-685} from GB:EAA14296

M13	Sjp_0059900	Neprilysin-1, putative	<E, D336	<H, <H, E332	MER11159	3E-15	family M13 unassigned peptidases (<i>Drosophila melanogaster</i>) [M13.UPW]	{peptidase unit: 168-849} from GB:NP_511056
M13	Sjp_0134850	Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), putative	<E, E62D	<H, <H, A58E	MER27380	6.5E-16	family M13 unassigned peptidases (<i>Locusta migratoria</i>) [M13.UPW]	{peptidase unit: 42-727} from GB:AAN73018
M13	Sjp_0055990	Neprilysin (EC 3.4.24.11)	<E, <D	<H, <H, <E	MER01050	1.2E-15	neprilysin (<i>Homo sapiens</i>) [M13.001]	{peptidase unit: 52-750} from Sw:P08473
M13	Sjp_0079080	Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), putative	E84, A128D	H83, H87, D124E	MER26946	8.1E-16	similar to Kell blood group protein { <i>Rattus norvegicus</i> } (<i>Rattus norvegicus</i>) [M13.950]	{peptidase unit: 1-254} from GB:XP_216130
M14A	Sjp_0005020	carboxypeptidase A2_K01298, putative	R244, E391	<H, <E, H315	MER19052	1.2E-45	subfamily M14A unassigned peptidases (<i>Anopheles gambiae</i>) [M14.UPA]	{peptidase unit: 113-416} from GB:EAA06398
M14B	Sjp_0094120	Cytosolic carboxypeptidase-like protein 5, putative	R407	H356, E359, >H	MER33174	6.6E-55	Mernase-AA216 hypothetical peptidase (<i>Homo sapiens</i>) [M14.025]	{peptidase unit: 149-350} from GB:CAD38704
M14B	SJCHGC06984	SJCHGC06984, similar to carboxypeptidase EAplysia californica (California sea hare)	<R, E17	<H, <E, <H	MER04959	6.60E-29	metallocarboxypeptidase D peptidase unit 2 (<i>Anas platyrhynchos</i>) [M14.016]	{peptidase unit: 502-906}
M14B	Sjp_0013530	carboxypeptidase D_K07752, putative	R130, E236	H69, E72, H178	MER15121	9.2E-68	subfamily M14B unassigned peptidases (<i>Arabidopsis thaliana</i>) [M14.UPB]	{peptidase unit: 55-444} from GB:CAD87770
M14B	Sjp_0088200	Carboxypeptidase D precursor (EC 3.4.17.22) (Metalloproteinase D) (Protein silver), putative	<R, >E	<H, <E, H18	MER03495	1.4E-21	subfamily M14B unassigned peptidases (<i>Caenorhabditis elegans</i>) [M14.UPB]	{peptidase unit: 68-478} from GB:AAB37821
M14X	Sjp_0004350	carboxypeptidase A (EC 3.4.17.1), K01290, putative	Y186E	H137, E140, >H	MER84362	5.4E-39	family M14 non-peptidase homologues (<i>Tetrahymena thermophila</i>) [M14.UNW]	{peptidase unit: 466-650} from GB:EAR88956
M16A	Sjp_0009910	insulysin_K01408.[PATH ko05010], putative	E71, E142	H68, H72, E149	MER53603	4.6E-60	insulysin (<i>Canis familiaris</i>) [M16.002]	{peptidase unit: 44-255}

M16A	Sjp_0009920	insulysin,K01408,[PATH ko05010], putative	A441E, K512E	P438H, V442H, L519E	MER78753	6E-44	insulysin unit 3 { { <i>Homo sapiens</i> } } (<i>Homo sapiens</i>) [M16.984]	{peptidase unit: 724-934} from Sw:P14735
M16A	Sjp_0047620	nardilysin,K01411, putative	E133, E204	H130, H134, E211	MER01220	4.5E-48	nardilysin (<i>Rattus norvegicus</i>) [M16.005]	{peptidase unit: 182-391} from Sw:P47245
M16A	SJCHGC09278	SJCHGC09278, similar to NRD1 protein (<i>Homo sapiens</i>)	E104, E175	H101, H105, E182	MER01220	7.00E-48	nardilysin (<i>Rattus norvegicus</i>) [M16.005]	{peptidase unit: 182-391} from Sw:P47245
M16B	Sjp_0027410	mitochondrial processing peptidase,K01412, putative	E51, E121	H48, H52, E128	MER62893	9.9E-90	mitochondrial processing peptidase beta-subunit (<i>Bos taurus</i>) [M16.003]	{peptidase unit: 63-293} from Sw:Q3SZ71
M16B	SJCHGC02536	SJCHGC02536, similar to peptidase (mitochondrial processing) beta	E51, E121	H48, H52, E128	MER62893	9.90E-90	mitochondrial processing peptidase beta-subunit (<i>Bos taurus</i>) [M16.003]	{peptidase unit: 63-293} from Sw:Q3SZ71
M16B	SJCHGC02537	SJCHGC02537, similar to RIKEN cDNA 3110004O18 (<i>Mus musculus</i>)	E87, >E	H84, H88, >E	MER62893	3.90E-38	mitochondrial processing peptidase beta-subunit (<i>Bos taurus</i>) [M16.003]	{peptidase unit: 63-293} from Sw:Q3SZ71
M16B	Sjp_0021970	mitochondrial processing peptidase,K01412, putative	E91, A162E	H88, K92H, E169	MER27090	2.7E-60	mitochondrial processing peptidase non-peptidase alpha subunit (<i>Anopheles gambiae</i>) [M16.971]	{peptidase unit: 96-331} from GB:EAA12324
M16B	Sjp_0030820	Cytochrome b-c1 complex subunit 2, mitochondrial precursor, putative	R81E, >E	H78, R82H, >E	MER03544	5.8E-09	ubiquinol-cytochrome c reductase core protein II (<i>Homo sapiens</i>) [M16.974]	{peptidase unit: 42-266} from Sw:P22695
M16C	Sjp_0082360	Presequence protease, mitochondrial precursor (EC 3.4.24.-)	<E, <E	<H, <H, <E	MER26956	2.5E-15	eupitirylisin (<i>Rattus norvegicus</i>) [M16.009]	{peptidase unit: 153-646} from GB:XP_225517
M16C	SJCHGC03612	SJCHGC03612, similar to MGC69133 protein (<i>Xenopus laevis</i>)	E75, E148	H72, H76, >E	MER26956	1.40E-38	eupitirylisin (<i>Rattus norvegicus</i>) [M16.009]	{peptidase unit: 153-646} from GB:XP_225517
M17	Sjp_0004310	Putative aminopeptidase W07G4.4, putative	K296, R371	K284, D289, D307, D367, E369	MER03499	2.7E-115	family M17 unassigned peptidases (<i>Schistosoma mansoni</i>) [M17.UPW]	{peptidase unit: 6-520} from GB:AAB41442

M17	Sjp_0025490	Putative aminopeptidase W07G4.4 (EC 3.4.11.-), putative	>K, >R	>K, >D, >D, >D, >E	MER03499	3.1E-114	family M17 unassigned peptidases (<i>Schistosoma mansonii</i>) [M17.UPW]	{peptidase unit: 6-520} from GB:AAB41442
M17	Sjp_0118990	Putative aminopeptidase W07G4.4 (EC 3.4.11.-), putative	<K, R71	<K, <D, D7, D67, E69	MER03499	1.1E-42	family M17 unassigned peptidases (<i>Schistosoma mansonii</i>) [M17.UPW]	{peptidase unit: 6-520} from GB:AAB41442
M17	SIHGC01390	his-tagged cytosolic leucine aminopeptidase [<i>Schistosoma japonicum</i>]	K279, R354	K267, D272, D290, D350, E352	MER03499	4.60E-202	family M17 unassigned peptidases (<i>Schistosoma mansonii</i>) [M17.UPW]	{peptidase unit: 6-520} from GB:AAB41442
M18	Sjp_0060580	aspartyl aminopeptidase, K01267, putative	D1037, E1257	H1035, D1223, E1258, D1323, H1417	MER86152	1.2E-102	aspartyl aminopeptidase (<i>Aspergillus terreus</i>) [M18.002]	{peptidase unit: 8-498} from GB:EAU38396
M20A	Sjp_0113330	aminoacylase, K01436, [PATH ko00220], putative	<D, <E	<H, <D, <E, <D/E, H169	MER38710	4.9E-36	aminoacylase (<i>Anopheles gambiae</i>) [M20.973]	{peptidase unit: 1-423} from GB:EA A06832
M22	SIHGC02749	SIHGC02749, similar to I313300A acylphosphatase isozyme Chl <i>Gallus gallus</i> (chicken)		>H, >H	MER74280	1.60E-06	family M22 non-peptidase homologues (<i>Aeromonas hydrophila</i>) [M22.UNW]	{peptidase unit: 19-788} from GB:ABK39492
M22	Sjp_0109990	O-sialoglycoprotein endopeptidase, K01409, putative		<H, <H	MER22512	2.5E-23	Kae1 putative peptidase (<i>Anopheles gambiae</i>) [M22.003]	{peptidase unit: 5-327} from GB:EA A04730
M22	Sjp_0083860	Probable O-sialoglycoprotein endopeptidase (EC 3.4.24.57) (mOsgep), putative		<H, <H	MER22512	8E-23	Kae1 putative peptidase (<i>Anopheles gambiae</i>) [M22.003]	{peptidase unit: 5-327} from GB:EA A04730
M22	Sjp_0083850	O-sialoglycoprotein endopeptidase, K01409, putative		H28, H32	MER01577	2E-39	Kae1 putative peptidase (<i>Homo sapiens</i>) [M22.003]	{peptidase unit: 2-314} from Sw:Q9NPF4
M22	Sjp_0083840	Probable O-sialoglycoprotein endopeptidase (EC 3.4.24.57)		<H, <H	MER05592	3.7E-29	Kae1 putative peptidase (<i>Rattus norvegicus</i>) [M22.003]	{peptidase unit: 2-314} from Sw:Q9WVS2
M22	Sjp_0050120	Probable O-sialoglycoprotein endopeptidase (EC 3.4.24.57), putative		>H, >H	MER79234	4.2E-29	Kae1 putative peptidase (<i>Xenopus tropicalis</i>) [M22.003]	{peptidase unit: 2-314} from GB:AAI21531
M22	Sjp_0112020	O-sialoglycoprotein endopeptidase, K01409, putative		H79, H83	MER13498	2.7E-25	Mername-AA018 peptidase (<i>Homo sapiens</i>) [M22.004]	{peptidase unit: 37-402} from GB:CAC14666

M22	SJCHGC07257	SJCHGC07257, similar to AAH58172-Unknown (protein for MGC:67870) (<i>Mus musculus</i>)	H74, H78	MER13498	1.20E-25	Membrane-AA018 peptidase (<i>Homo sapiens</i>) [M22.004]	{peptidase unit: 37-402} from GB:CAI4666
M23B	Sjp_0112430	2-oxoglutarate dehydrogenase E2 component	T91H	MER92387	2E-13	subfamily M23B non-peptidase homologues (<i>Mycobacterium gilvum</i>) [M23.UNB]	{peptidase unit: 100-369} from GB:ABP45412
M23B	SJCHGC04170	SJCHGC04170, similar to CAG10633-unnamed protein product (<i>Tetraodon nigroviridis</i>)	T124H	MER92387	1.90E-14	subfamily M23B non-peptidase homologues (<i>Mycobacterium gilvum</i>) [M23.UNB]	{peptidase unit: 100-369} from GB:ABP45412
M23B	Sjp_0035730	myosin heavy chain, K10352, putative	-1348H	MER68428	1.8E-12	subfamily M23B non-peptidase homologues (<i>Naotomonas pharaonis</i>) [M23.UNB]	{peptidase unit: 206-455} from GB:YP_326268
M23B	Sjp_0054830	myosin heavy chain, K10352, putative	N851H	MER68428	2.1E-10	subfamily M23B non-peptidase homologues (<i>Naotomonas pharaonis</i>) [M23.UNB]	{peptidase unit: 206-455} from GB:YP_326268
M24A	Sjp_0097930	methionyl aminopeptidase, K01265, putative	H64	MER21301	3.3E-69	methionyl aminopeptidase 1 (<i>Anopheles gambiae</i>) [M24.001]	{peptidase unit: 132-393} from GB:EAA08766
M24A	SJCHGC08091	SJCHGC08091, similar to ENSANGP000000021308 (<i>Anopheles gambiae</i>)	H183	MER21301	2.30E-40	methionyl aminopeptidase 1 (<i>Anopheles gambiae</i>) [M24.001]	{peptidase unit: 132-393} from GB:EAA08766
M24A	Sjp_0084000	methionyl aminopeptidase, K01265, putative	H115	MER50562	1.5E-41	methionyl aminopeptidase 1 (<i>Leishmania major</i>) [M24.001]	{peptidase unit: 128-384} from GB:CAJ07139
M24A	Sjp_0009900	methionyl aminopeptidase, K01265, putative	H228	MER01728	2.2E-157	methionyl aminopeptidase 2 (<i>Homo sapiens</i>) [M24.002]	{peptidase unit: 156-478} from Sw:P50579
M24B	Sjp_0092540	X-Pro aminopeptidase, K01262, putative	H204, H307, H318	MER13463	9.1E-61	aminopeptidase P homologue (<i>Homo sapiens</i>) [M24.026]	{peptidase unit: 242-501} from Sw:Q9NQH7

M24B	SICHGC00876	SJCHGC00876, similar to NP_071381-hypothetical protein LOC63929 (<i>Homo sapiens</i>)	H308, H411, H422	D325, D336, H415, E442, E468	MER13463	5.30E-61	aminopeptidase P homologue (<i>Homo sapiens</i>) [M24.026]	{peptidase unit: 242-501} from Sw:Q9NQH7
M24B	Sjp_0095970	X-Pro aminopeptidase,K01262, putative	H568, H658, H671	D588, D599, H662, >E, >E	MER27092	6.6E-40	aminopeptidase P1 (<i>Anopheles gambiae</i>) [M24.009]	{peptidase unit: 330-589} from GB:CAA04929
M24B	Sjp_0027920	X-Pro dipeptidase,K01271, putative	H938, H1049, H1060	D959, D970, H1053, E1096, E1136	MER10979	1.8E-114	Xaa-Pro dipeptidase {eukaryote} (Suberites domuncula) [M24.007]	{peptidase unit: 191-486} from GB:CAA75231
M24X	Sjp_0128810	methionyl aminopeptidase (EC 3.4.11.18), K01265, putative	H136	N159D, I170D, H238, D277E, R369E	MER64643	2.9E-106	proliferation-associated protein 2G4, 38kDa {{ <i>Rattus norvegicus</i> }} (<i>Rattus norvegicus</i>) [M24.978]	{peptidase unit: 8-348} from GB:AAH79095
M28A	Sjp_0051220	glutaminyl-peptide cyclotransferase,K00683, putative	D128, E177	H126, D140, E178, N219D/E, >H	MER15094	6.6E-36	family M28 homologue similar to glutaminyl-peptide cyclotransferase {{ <i>Rattus norvegicus</i> }} (<i>Mus musculus</i>) [M28.979]	{peptidase unit: 60-362} from Sw:Q9CYK2
M28B	Sjp_0085790	glutamate carboxypeptidase II,K01301, putative	D344, E389	H342, D352, E390, D418, H512	MER05238	1.7E-48	glutamate carboxypeptidase III (<i>Homo sapiens</i>) [M28.012]	{peptidase unit: 281-566} from Sw:Q9Y3Q0
M28X	Sjp_0060780	Endoplasmic reticulum metalloproteinase 1, putative	D186, E230	H184, D196, E231, E257, H335	MER53354	4.8E-62	Fxa peptidase {{ <i>Rattus norvegicus</i> }} (<i>Pan troglodytes</i>) [M28.018]	{peptidase unit: 189-444} from GB:XP_520478
M28X	Sjp_0008170	Nicalin-1 precursor (Nicastrin-like protein 1), putative	D233, G285E	Y231H, A247D, K286E, >D/E, >H	MER53665	1E-22	nicalin (<i>Canis familiaris</i>) [M28.978]	{peptidase unit: 532-726} from GB:XP_542180
M38	Sjp_0026210	dihydroorotase,K01465,[PATH ko00240], putative	D64	<H, <H, <K, <H, <H	MER05767	4.1E-28	dihydro-orotase (<i>Homo sapiens</i>) [M38.972]	{peptidase unit: 1438-1742} from Sw:P27708
M38	Sjp_0009470	Dihydropyrimidinase-related protein 3 (DRP-3)	>D	D67H, H69, S159K, Q214H, R258H	MER30143	2.3E-37	dihydropyrimidinase related protein-1 (<i>Homo sapiens</i>) [M38.974]	{peptidase unit: 17-411} from Sw:Q14194
M38	Sjp_0056150	dihydropyrimidinase-like 2,K07528,[PATH ko04360], putative	G338D	H80, C82H, R171K, N220H, Q260H	MER30151	2.7E-117	dihydropyrimidinase related protein-3 (<i>Homo sapiens</i>) [M38.976]	{peptidase unit: 17-411} from Sw:Q14195

M38	SIHGGC09277	SIHGGC09277, similar to MGC68668 protein (<i>Xenopus laevis</i>)	G324D	H66, C68H, R157K, N206H, Q246H	MER30151	2.70E-117	dihydropyrimidinase related protein-3 (<i>Homo sapiens</i>) [M38.976]	{peptidase unit: 17-411} from Sw:Q14195
M38	Sjp_0064410	N-acetylglucosamine-6-phosphate deacetylase, K01443, [PATH ko00530], putative	>D	Q80H, N82H, K192, H221, H242	MER33184	2E-75	hypothetical protein like 5730457F11RIK (<i>Homo sapiens</i>) [M38.979]	{peptidase unit: 17-415} from GB:AAH18734
M38	SIHGGC02614	SIHGGC02614, similar to Hypothetical protein CBG06850 (<i>Caenorhabditis briggsae</i>)	>D	<H, <H, K47, H76, >H	MER33184	7.80E-16	hypothetical protein like 5730457F11RIK (<i>Homo sapiens</i>) [M38.979]	{peptidase unit: 17-415} from GB:AAH18734
M41	Sjp_0090130	AFG3 family protein, K08956, putative	E259	H258, H262, >D	MER05496	9.3E-28	Afg3-like protein 2 (<i>Homo sapiens</i>) [M41.007]	{peptidase unit: 522-790} from Sw:Q9Y4W6
M41	SIHGGC02177	>gb AAO32953.1 putative AFG3-like protein 2 [<i>Schistosoma japonicum</i>]	<E	<H, <H, D25	MER05496	7.70E-39	Afg3-like protein 2 (<i>Homo sapiens</i>) [M41.007]	{peptidase unit: 522-790} from Sw:Q9Y4W6
M41	SIHGGC02179	SIHGGC02179, similar to putative AFG3-like protein 2 (<i>Schistosoma japonicum</i>)	<E	<H, <H, <D	MER05496	3.40E-14	Afg3-like protein 2 (<i>Homo sapiens</i>) [M41.007]	{peptidase unit: 522-790} from Sw:Q9Y4W6
M41	Sjp_0093270	nucleoside-triphosphatase (EC 3.6.1.15), K01516, putative	E426	H425, H429, D503	MER21311	6E-60	i-AAA peptidase (<i>Anopheles gambiae</i>) [M41.004]	{peptidase unit: 551-785} from GB:EAA13918
M41	Sjp_0095400	spastic paraplegia 7, K09552, putative	E369	H368, R372H, Y451D	MER04454	3.1E-44	paraplegin (<i>Homo sapiens</i>) [M41.006]	{peptidase unit: 522-795} from Sw:Q9UQ90
M48A	Sjp_0042800	STE24 endopeptidase, K06013, putative	E462	H461, H465, E542	MER02646	2.3E-76	farnesylated-protein converting enzyme 1 (<i>Homo sapiens</i>) [M48.003]	{peptidase unit: 164-475} from Sw:O75844
M49	Sjp_0090260	dipeptidyl-peptidase III, K01277, putative	E438	H437, H442, E495	MER27463	2E-183	dipeptidyl-peptidase III (<i>Xenopus laevis</i>) [M49.001]	{peptidase unit: 15-744} from GB:AAH43905
M50A	Sjp_0070220	S2P endopeptidase, K07765, putative	E141	H140, H144, >D	MER56624	2E-56	S2P peptidase (<i>Strongylocentrotus purpuratus</i>) [M50.001]	{peptidase unit: 30-474} from GB:XP_787875
M67A	Sjp_0117270	26S proteasome non-ATPase regulatory subunit 7, putative	R280E	>H, >H, >D	MER30134	1.2E-24	26S proteasome non-ATPase regulatory subunit 7 (<i>Homo sapiens</i>) [M67.973]	{peptidase unit: 7-125} from Sw:P51665
M67A	Sjp_0109020	COP9 signalosome complex subunit 6	T66E	T123H, G125H, D132	MER30137	2.6E-25	COP9 signalosome subunit 6 (<i>Homo sapiens</i>) [M67.972]	{peptidase unit: 41-143} from Sw:Q7L5N1

M67A	SICHGC04182	SICHGC04182, similar to COP9 signalosome subunit 6	T48E	>H, >H, >D	MER30137	4.20E-14	COP9 signalosome subunit 6 (<i>Homo sapiens</i>) [M67.972]	{peptidase unit: 41-143} from Sw:Q7L5N1
M67A	Sjp_0052560	COP9 signalosome complex subunit 5, K09613, putative	E1306	H1368, H1370, D1381	MER22069	1.4E-108	Jab1/MPN domain metalloenzyme (<i>Arabidopsis thaliana</i>) [M67.002]	{peptidase unit: 26-336} from Sw:Q8LAZ7
M67A	Sjp_0019080	26S proteasome non-ATPase regulatory subunit 14, putative	E55	H116, H118, D129	MER20371	3.6E-134	Poh1 peptidase (<i>Anopheles gambiae</i>) [M67.001]	{peptidase unit: 15-325} from GB:EAA10169
M67A	SICHGC06520	Pad1 homolog [<i>Schistosoma japonicum</i>]	E55	H116, H118, D129	MER20371	7.10E-135	Poh1 peptidase (<i>Anopheles gambiae</i>) [M67.001]	{peptidase unit: 15-325} from GB:EAA10169
M67C	Sjp_0090150	ubiquitin thiolesterase, K01072, putative	E216	H271, H273, D284	MER30146	4.8E-49	AMSH deubiquitinating peptidase (<i>Homo sapiens</i>) [M67.006]	{peptidase unit: 251-405} from Sw:O95630

Supplementary Table 24. The predicted cysteine proteases of *S. japonicum*

Protease family	Protein name	Annotation	Predicted active site residues	Hit name	E-value comparing MEROPS	Annotation in MEROPS	Peptidase unit
C01A	Sjp_0124210	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen Sm31), putative	<Q, <C, I59H, N79	MER54623	9.3E-29	cathepsin B (Paralichthys olivaceus) [C01.060]	{peptidase unit: 70-330} from GB:ABM47001
C01A	Sjp_0089450	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen Sj31), putative	Q112, C118, H288, N308	MER00692	1.2E-151	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157
C01A	Sjp_0077120	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen Sj31), putative	Q112, L118C, H286, N306	MER00692	9.6E-139	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157
C01A	Sjp_0131900	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen Sj31), putative	Q112, S118C, H288, N308	MER00692	2.4E-133	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157
C01A	Sjp_0077140	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen Sj31), putative	Q79, X85C, H255, N275	MER00692	3.3E-122	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157
C01A	Sjp_0112990	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen Sj31), putative	Q77, C83, H254, N276	MER00692	7.3E-103	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157

C01A	Sjp_0029530	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen Sm31), putative	Q92, X98C, Q268H, N288	MER00692	1.9E-88	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157
C01A	Sjp_0089420	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen S31), putative	Q112, S118C, >H, >N/Q	MER00692	1.1E-82	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157
C01A	Sjp_0077110	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen S31), putative	<Q, <C, H93, N113	MER00692	9.3E-46	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157
C01A	Sjp_0068040	Cathepsin B-like cysteine proteinase precursor (EC 3.4.22.-) (Antigen S31), putative	E111Q, D117C, >H, >N/Q	MER00692	1E-42	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157
C01A	SJCHGC02852	similar to cathepsin B in <i>Schistosoma japonicum</i>	Q114, C120, H291, N313	MER00692	9.50E-103	cathepsin B-like peptidase {platyhelminth} (<i>Schistosoma japonicum</i>) [C01.062]	{peptidase unit: 81-341} from Sw:P43157
C01A	Sjp_0107700	Cathepsin F, K01373, putative	Q103, C109, >H, >N/Q	MER65230	2.2E-51	cathepsin F (<i>Tetraodon nigroviridis</i>) [C01.018]	{peptidase unit: 1-203} from GB:CAF92720
C01A	Sjp_0014320	Cathepsin K precursor (EC 3.4.22.38), putative	<Q, S8C, >H, >N/Q	MER05622	1.1E-16	cathepsin K (<i>Macaca mulatta</i>) [C01.036]	{peptidase unit: 108-329} from Sw:P61277
C01A	Sjp_0132650	Cathepsin L precursor (EC 3.4.22.15)	Q168, C174, >H, >N/Q	MER29278	1.2E-58	cathepsin L-like endopeptidase {insect} (<i>Hydra vulgaris</i>) [C01.092]	{peptidase unit: 107-324} from GB:AAO65603
C01A	Sjp_0116670	Cathepsin L (EC 3.4.22.15)	<Q, <C, H24, N44	MER36273	1.2E-28	cathepsin L-like endopeptidase {insect} (<i>Rhipicephalus haemaphysaloides</i>) [C01.092]	{peptidase unit: 115-332} from GB:AAQ16117

C01A	Sjp_0044890	Cathepsin L (EC 3.4.22.15)	<Q, <C, H19, N39	MER36273	2.7E-28	cathepsin L-like endopeptidase {insect} (<i>Rhipicephalus</i> <i>haemaphysaloides</i>) [C01.092]	{peptidase unit: 115-332} from GB:AAQ16117
C01A	Sjp_0027020	Cathepsin L-like proteinase precursor (EC 3.4.22.-), putative	Q178, G184C, H321, N343	MER02418	1E-57	cathepsin L-like peptidase {Fasciola} sp.} (<i>Fasciola</i> <i>gigantica</i>) [C01.033]	{peptidase unit: 107-323} from GB:BAA23743
C01A	Sjp_0044950	Cathepsin L-like proteinase precursor (EC 3.4.22.-), putative	<Q, S14C, Q148H, N169	MER02418	5.5E-48	cathepsin L-like peptidase {Fasciola} sp.} (<i>Fasciola</i> <i>gigantica</i>) [C01.033]	{peptidase unit: 107-323} from GB:BAA23743
C01A	Sjp_0044970	Cathepsin L-like proteinase precursor (EC 3.4.22.-), putative	<Q, <C, Q30H, N50	MER12037	1.2E-20	cathepsin L-like peptidase {Fasciola} sp.} (<i>Fasciola</i> <i>gigantica</i>) [C01.033]	{peptidase unit: 107-323} from GB:AAF44677
C01A	SJCHGC02868	>gb AAP05886.1 similar to AJ279092 cathepsin L1 in <i>Fasciola hepatica</i> [<i>Schistosoma japonicum</i>]	Q139, S145C, Q282H, N303	MER02418	7.50E-54	cathepsin L-like peptidase {Fasciola} sp.} (<i>Fasciola</i> <i>gigantica</i>) [C01.033]	{peptidase unit: 107-323} from GB:BAA23743
C01A	Sjp_0044980	Cathepsin L-like proteinase precursor (EC 3.4.22.-), putative	Q77, S83C, >H, >N/Q	MER00635	2.3E-21	cathepsin L-like peptidase {Fasciola} sp.} (<i>Fasciola hepatica</i>) [C01.033]	{peptidase unit: 108-323} from Sw:Q24940
C01A	Sjp_0043440	Cathepsin C, K01275, putative	X243Q, C249, H400, N422	MER00705	1.6E-85	dipeptidyl-peptidase I (<i>Rattus</i> <i>norvegicus</i>) [C01.070]	{peptidase unit: 229-459} from Sw:P80067
C01A	Sjp_0043460	Cathepsin C precursor (EC 3.4.22.-), putative	<Q, <C, H183, N205 Q249, C255, H406, N428	MER00705	1.2E-47	dipeptidyl-peptidase I (<i>Rattus</i> <i>norvegicus</i>) [C01.070]	{peptidase unit: 229-459} from Sw:P80067
C01A	SJCHGC01753	Cathepsin C [<i>Schistosoma japonicum</i>]		MER00705	7.70E-75	dipeptidyl-peptidase I (<i>Rattus</i> <i>norvegicus</i>) [C01.070]	{peptidase unit: 229-459} from Sw:P80067

C01A	Sjp_0007940	Cathepsin B	Q117, C123, H292, N312	MER19717	5.7E-141	SmCB2 peptidase {{Schistosoma sp.} (<i>Schistosoma mansoni</i>) [C01.061]	{peptidase unit: 85-344} from GB:CAC85211
C01A	Sjp_0107130	Peptidase C1A, papain C-terminal, domain-containing	<Q, <C, <H, N15	MER19717	9.8E-15	SmCB2 peptidase {{Schistosoma sp.} (<i>Schistosoma mansoni</i>) [C01.061]	{peptidase unit: 85-344} from GB:CAC85211
C01A	SJCHGC02034	SJCHGC02034, similar to cathepsin B endopeptidase (<i>Schistosoma japonicum</i>)	Q49, -52C, >H, >N/Q	MER19717	5.20E-13	SmCB2 peptidase {{Schistosoma sp.} (<i>Schistosoma mansoni</i>) [C01.061]	{peptidase unit: 85-344} from GB:CAC85211
C01A	Sjp_0104290	Cathepsin L, K01365, [PATH ko04612], putative	Q134, C140, H277, N297	MER02333	1.4E-95	SmCL2-like peptidase (<i>Schistosoma mansoni</i>) [C01.044]	{peptidase unit: 102-316} from GB:CAA83538
C01A	Sjp_0120460	Cathepsin S precursor (EC 3.4.22.27), putative	Q139, >C, >H, >N/Q	MER62697	6.1E-11	subfamily C1A unassigned peptidases (<i>Tetrahymena thermophila</i>) [C01.UPA]	{peptidase unit: 3-337} from GB:EAR95553
C01A	Sjp_0044990	Cathepsin S precursor (EC 3.4.22.27), putative	>Q, >C, >H, >N/Q	MER62697	7.3E-10	subfamily C1A unassigned peptidases (<i>Tetrahymena thermophila</i>) [C01.UPA]	{peptidase unit: 3-337} from GB:EAR95553
C01A	Sjp_0044960	Cathepsin S precursor (EC 3.4.22.27), putative	E130Q, >C, >H, >N/Q	MER62697	2.7E-08	subfamily C1A unassigned peptidases (<i>Tetrahymena thermophila</i>) [C01.UPA]	{peptidase unit: 3-337} from GB:EAR95553
C02	Sjp_0012570	Calpain-11, K08580, putative	Q208, C214, H374, N398	MER13513	5.6E-139	calpain {{Schistosoma} sp.} (<i>Schistosoma japonicum</i>) [C02.023]	{peptidase unit: 124-380} from GB:BAA74718
C02	Sjp_0098670	Calpain (EC 3.4.22.-)	<Q, <C, <H, N48	MER13513	5.8E-37	calpain {{Schistosoma} sp.} (<i>Schistosoma japonicum</i>) [C02.023]	{peptidase unit: 124-380} from GB:BAA74718
C02	SJCHGC06439	SJCHGC06439, similar to calpain (<i>Schistosoma japonicum</i>)	Q148, C154, >H, >N	MER13513	2.10E-43	calpain {{Schistosoma} sp.} (<i>Schistosoma japonicum</i>) [C02.023]	{peptidase unit: 124-380} from GB:BAA74718

C02	Sjp_0097500	Calpain (EC 3.4.22.-)	Q50, C56, >H, >N	MER03200	4.7E-62	calpain {{Schistosoma} sp.} (<i>Schistosoma mansoni</i>) [C02.023]	{peptidase unit: 124-380} from Sw:P27730
C02	Sjp_0044810	Calpain B (EC 3.4.22.-)	Q38, C44, H201, N229	MER49593	2.8E-102	calpain B {{ <i>Gecarcinus lateralis</i> }} (<i>Gecarcinus lateralis</i>) [C02.024]	{peptidase unit: 118-375} from GB:AAT77811
C02	Sjp_0028660	Calpain, invertebrate,K08585, putative	Q24, C30, H191, N217	MER49593	1.5E-90	calpain B {{ <i>Gecarcinus lateralis</i> }} (<i>Gecarcinus lateralis</i>) [C02.024]	{peptidase unit: 118-375} from GB:AAT77811
C02	SJCHGC01809	similar to calpain	Q170, C176, H333, N361	MER49593	1.10E-105	calpain B {{ <i>Gecarcinus lateralis</i> }} (<i>Gecarcinus lateralis</i>) [C02.024]	{peptidase unit: 118-375} from GB:AAT77811
C02	Sjp_0029310	Calpain-9 (EC 3.4.22.-)	Q189, C195, Q361H, X385N	MER00770	2.1E-52	calpain-1 (<i>Homo sapiens</i>) [C02.001]	{peptidase unit: 85-337} from Sw:P07384
C02	Sjp_0035290	Calpain-7,K08576, putative	<Q, <C, H447, N467	MER05201	2.5E-45	calpain-7 (<i>Mus musculus</i>) [C02.008]	{peptidase unit: 261-523} from Sw:Q9R1S8
C02	Sjp_0052870	Protein C6orf103, putative	<Q, <C, >H, >N	MER03201	2E-39	hypothetical protein flj40251 (<i>Homo sapiens</i>) [C02.972]	{peptidase unit: 100-349} from Sw:Q8N7X0
C02	Sjp_0011610	Calpain-5 (EC 3.4.22.-) (nCL-3), putative	R75Q, F81C, >H, >N	MER54124	4E-29	Mername-AA248 (<i>Gallus gallus</i>) [C02.026]	{peptidase unit: 51-326}
C02	SJCHGC03597	SJCHGC03597, similar to CAG00593-unnamed protein product (<i>Tetraodon nigroviridis</i>)	R75Q, F81C, >H, >N	MER54124	2.90E-18	Mername-AA248 (<i>Gallus gallus</i>) [C02.026]	{peptidase unit: 51-326}
C12	SJCHGC01421	SJCHGC01421, similar to JC7117 ubiquitin carboxy-terminal hydrolase-6 (EC 3.1.-.-)	Q79, C85, H159, D174	MER12102	3.70E-51	family C12 unassigned peptidases (<i>Gallus gallus</i>) [C12.UPW]	{peptidase unit: 4-230} from GB:NP_990156
C12	Sjp_0020870	Ubiquitin carboxyl-terminal hydrolase BAP1 (EC 3.4.19.12)	>Q, >C, >H, >D	MER19047	2.3E-21	ubiquitinyl hydrolase-BAP1 (<i>Anopheles gambiae</i>) [C12.004]	{peptidase unit: 4-235} from

C12	Sjp_0070430	ubiquitin carboxyl-terminal hydrolase BAP1, K08588, putative	Q95, C101, >H, >D	MER03989	2.9E-19	ubiquitinyl hydrolase-BAP1 (<i>Homo sapiens</i>) [C12.004]	GB:EAA00168 {peptidase unit: 3-237} from Sw:Q92560
C12	Sjp_0032670	ubiquitin carboxyl-terminal hydrolase L3, K05609, putative	Q75, C81, H155, >D	MER00836	3E-38	ubiquitinyl hydrolase-L3 (<i>Homo sapiens</i>) [C12.003]	{peptidase unit: 4-230} from Sw:P15374
C12	Sjp_0060830	ubiquitin carboxyl-terminal hydrolase L5, K05610, putative	Q79, C85, H160, D175	MER05539	1.2E-64	ubiquitinyl hydrolase-UCH37 (<i>Homo sapiens</i>) [C12.005]	{peptidase unit: 6-227} from Sw:Q9Y5K5
C13	Sjp_0074300	Putative GPI-anchor transamidase precursor (EC 3.-.-) (GPI transamidase), putative	<H, C41	MER04448	1.3E-32	glycosylphosphatidylinositol:protein transamidase (<i>Drosophila melanogaster</i>) [C13.005]	{peptidase unit: 42-307} from Sw:Q8T4E1
C13	Sjp_0119400	Hemoglobinase precursor (EC 3.4.22.34) (Antigen Sj32), putative	<H, <C	MER02548	0.0000034	legumain (<i>Caenorhabditis elegans</i>) [C13.004]	{peptidase unit: 39-301} from GB:CAB01126
C13	Sjp_0001230	Legumain, K01369, [PATH ko04612], putative	H145, C186	MER54549	8.2E-83	legumain (Ixodes ricinus) [C13.004]	{peptidase unit: 36-292} from GB:AAS94231
C13	Sjp_0119250	Hemoglobinase precursor (EC 3.4.22.34) (Antigen Sj32), putative	H41, C82	MER54549	7.9E-24	legumain (Ixodes ricinus) [C13.004]	{peptidase unit: 36-292} from GB:AAS94231
C13	SJCHGC01905	>sp P42665 HGLB_SCHJA Hemoglobinase precursor (Antigen Sj32)	H145, C186	MER54549	3.60E-83	legumain (Ixodes ricinus) [C13.004]	{peptidase unit: 36-292} from GB:AAS94231
C14	Sjp_0009160	Caspase 3, apoptosis-related cysteine protease,, putative	H135, C177	MER20107	1.7E-57	caspase-3 (<i>Takifugu rubripes</i>) [C14.003]	{peptidase unit: 15-280} from GB:AAM48291
C14	SJCHGC00800	SJCHGC00800, similar to caspase 3-like (<i>Takifugu rubripes</i>)	H122, C164	MER20107	2.30E-57	caspase-3 (<i>Takifugu rubripes</i>) [C14.003]	{peptidase unit: 15-280} from GB:AAM48291

C14	Sjp_0039420	Caspase 7, apoptosis-related cysteine protease, K04397, [PATH ko04210, PATH ko05010, PATH ko05050], putative	H150, C193	MER62533	1.2E-62	caspase-7 (<i>Xenopus tropicalis</i>) [C14.004]	{peptidase unit: 38-319} from GB:CAJ82745
C14	Sjp_0071380	Caspase 2, apoptosis-related cysteine protease, K02186, putative	H214, C257	MER00854	9.9E-37	CED-3 peptidase (<i>Caenorhabditis elegans</i>) [C14.002]	{peptidase unit: 203-498} from Sw:P42573
C15	Sjp_0059190	pyroglutamyl-l-peptidase, K01304, putative	E95, C158, H177	MER56614	8.9E-23	pyroglutamyl-peptidase I {chordate} (<i>Strongylocentrotus purpuratus</i>) [C15.010]	{peptidase unit: 2-196} from GB:XP_786904
C19	Sjp_0061650	ubiquitin thiolesterase, K01072, putative	N50, C55, >H, >N/D	MER24829	6.2E-24	family C19 non-peptidase homologues (<i>Drosophila melanogaster</i>) [C19.UNW]	{peptidase unit: 268-1126} from GB:AA09142
C19	SJCHGC02219	SJCHGC02219, similar to ENSANGP00000001529 (<i>Anopheles gambiae</i>)	N39, C44, >H, >D/N	MER24829	1.90E-19	family C19 non-peptidase homologues (<i>Drosophila melanogaster</i>) [C19.UNW]	{peptidase unit: 268-1126} from GB:AA09142
C19	Sjp_0131530	U4/U6.U5 tri-snRNP-associated protein 2, putative	N187, D192C, >H, >D/N	MER32988	1.4E-63	family C19 non-peptidase homologues (<i>Magnaporthe grisea</i>) [C19.UNW]	{peptidase unit: 104-563}
C19	SJCHGC05849	SJCHGC05849, similar to BAC32203-unnamed protein product (<i>Mus musculus</i>)	N187, D192C, >H, >D/N	MER32988	2.60E-58	family C19 non-peptidase homologues (<i>Magnaporthe grisea</i>) [C19.UNW]	{peptidase unit: 104-563}
C19	Sjp_0023350	Ubiquitin carboxyl-terminal hydrolase 27, putative	<N, <C, H22, D38	MER56237	1.5E-15	family C19 unassigned peptidases (<i>Apis mellifera</i>) [C19.UPW]	{peptidase unit: 132-495}
C19	Sjp_0037330	ubiquitin thiolesterase, K01072, putative	<N, <C, H469, >D/N	MER90166	2.7E-19	family C19 unassigned peptidases (<i>Aspergillus nidulans</i>) [C19.UPW]	{peptidase unit: 350-784} from GB:EAA62367
C19	Sjp_0035300	Probable ubiquitin carboxyl-terminal hydrolase R10E11.3 (EC 3.1.2.15)	<N, <C, H910, D925	MER38083	6.5E-10	family C19 unassigned peptidases (<i>Drosophila melanogaster</i>) [C19.UPW]	{peptidase unit: 320-752} from GB:AAQ22510

C19	Sjp_0070290	ubiquitin thiolesterase, K01072, putative	N266, C271, >H, >D/N	MER56568	3.5E-77	family C19 unassigned peptidases (<i>Strongylocentrotus purpuratus</i>) [C19.UPW]	{peptidase unit: 385-1036} from GB:XP_781672
C19	Sjp_0106640	ubiquitin thiolesterase, K01072, putative	N42, C47, >H, >D/N	MER49607	3.2E-50	ubiquitin-specific peptidase 12 (<i>Brachydanio rerio</i>) [C19.020]	{peptidase unit: 22-377}
C19	SJCHGC02682	SJCHGC02682, similar to AAH74641-Unknown (protein for MGC:69343) (<i>Xenopus tropicalis</i>)	N107, C112, >H, >D/N	MER02667	3.50E-21	ubiquitin-specific peptidase 14 (<i>Homo sapiens</i>) [C19.015]	{peptidase unit: 73-494} from Sw:P54578
C19	Sjp_0082040	ubiquitin thiolesterase, K01072, putative	<N, <C, >H, >D/N	MER04834	2.9E-28	ubiquitin-specific peptidase 2 (<i>Homo sapiens</i>) [C19.013]	{peptidase unit: 4-362} from Sw:O75604
C19	Sjp_0082030	Ubiquitin carboxyl-terminal hydrolase 2, putative	<N, <N, <C, <C, <N, <C, H44, N61	MER69124	4.4E-19	ubiquitin-specific peptidase 2 (<i>Monodelphis domestica</i>) [C19.013]	{peptidase unit: 207-600}
C19	Sjp_0023310	ubiquitin thiolesterase, K01072, putative	N135, C140, >H, >D/N	MER12130	1.2E-31	ubiquitin-specific peptidase 22 (<i>Homo sapiens</i>) [C19.035]	{peptidase unit: 138-525} from Sw:Q9UPT9
C19	Sjp_0079300	Ubiquitin carboxyl-terminal hydrolase 24, putative	<N, <C, >H, >D/N	MER05706	5.1E-56	ubiquitin-specific peptidase 24 (<i>Homo sapiens</i>) [C19.047]	{peptidase unit: 901-1303} from Sw:Q9UPU5
C19	Sjp_0054130	ubiquitin thiolesterase, K01072, putative	<N, <C, H161, N176	MER14033	6.8E-41	ubiquitin-specific peptidase 36 (<i>Homo sapiens</i>) [C19.042]	{peptidase unit: 84-434} from Sw:Q9P275
C19	Sjp_0104920	ubiquitin thiolesterase, K01072, putative	<N, <C, >H, >D/N	MER36093	6E-26	ubiquitin-specific peptidase 44 (<i>Brachydanio rerio</i>) [C19.057]	{peptidase unit: 257-695} from GB:AAH48060
C19	SJCHGC05977	SJCHGC05977, similar to RIKEN cDNA 2410018108 (<i>Mus musculus</i>)	N42, C47, H322, D337	MER14629	2.30E-129	ubiquitin-specific peptidase 46 (<i>Homo sapiens</i>) [C19.052]	{peptidase unit: 1-366} from Sw:P62068
C19	Sjp_0047570	ubiquitin thiolesterase, K01072, putative	<N, <C, G273H, N282	MER49598	0.0000068	ubiquitin-specific peptidase 5 (<i>Brachydanio rerio</i>) [C19.001]	{peptidase unit: 286-831} from GB:NP_999920

C19	Sjp_0047560	ubiquitin thiolesterase,K01072, putative	N263, C268, >H, >D/N, >H, >N/D	MER69159	2.4E-45	ubiquitin-specific peptidase 5 (<i>Monodelphis domestica</i>) [C19.001]	{peptidase unit: 285-859}
C19	Sjp_0092320	Ubiquitin carboxyl-terminal hydrolase 8 (EC 3.1.2.15), putative	<N, <C, >H, >D/N	MER01884	5.4E-25	ubiquitin-specific peptidase 8 (<i>Homo sapiens</i>) [C19.011]	{peptidase unit: 717-1118} from Sw:P40818
C19	Sjp_0130570	Ubiquitin carboxyl-terminal hydrolase 8 (EC 3.1.2.15)	<N, <C, >H, >D/N	MER01884	1.5E-16	ubiquitin-specific peptidase 8 (<i>Homo sapiens</i>) [C19.011]	{peptidase unit: 717-1118} from Sw:P40818
C19	Sjp_0032170	ubiquitin thiolesterase,K01072, putative	<N, <C, >H, >D/N	MER14628	1.7E-58	Usp9y g.p. {{ <i>Mus musculus</i> }} (<i>Mus musculus</i>) [C19.051]	{peptidase unit: 1521-1969} from GB:NP_683745
C26	Sjp_0103900	CTP synthase,K01937,[PATH ko00240], putative	C532, >H	MER89297	3.6E-25	family C26 unassigned peptidases (<i>Desulfotomaculum reducens</i>) [C26.UPW]	{peptidase unit: 347-529} from GB:ABO51683
C26	Sjp_0049430	GMP synthase (glutamine-hydrolysing),K01951,[PATH ko00230,PATH ko00251], putative	C80, H166	MER43387	1.5E-58	guanine 5'-monophosphate synthetase (<i>Homo sapiens</i>) [C26.950]	{peptidase unit: 43-244} from Sw:P49915
C44	Sjp_0087440	Putative glutamate synthase [NADPH] (EC 1.4.1.13) (NADPH-GOGAT), putative	<C	MER67018	1.8E-44	family C44 unassigned peptidases (<i>Rhodopirellula baltica</i>) [C44.UPW]	{peptidase unit: 25-403} from GB:NP_866822
C44	Sjp_0034580	glucosamine--fructose-6-phosphate aminotransferase	<C	MER52013	8.4E-68	glucosamine-fructose-6-phosphate aminotransferase (<i>Emeritella nidulans</i>) [C44.971]	{peptidase unit: 2-347} from GB:AAW49003
C44	SJCHGC07417	SJCHGC07417, similar to CAF90172-unnamed protein product (<i>Tetraodon nigroviridis</i>)	<C	MER12158	5.10E-13	glutamine:fructose-6-phosphate amidotransferase (<i>Homo sapiens</i>) [C44.972]	{peptidase unit: 2-207} from Sw:O94808
C46	Sjp_0024280	hedgehog,K06224,[PATH ko04340], putative	<C, <S/T, <H	MER02538	3.8E-09	Indian hedgehog protein (<i>Homo sapiens</i>) [C46.003]	{peptidase unit: 203-403} from

C48	Sjp_0112700	sentrin-specific protease 8,K08597, putative	H102, D119, Q156, C162	MER11488	2.3E-21	family C48 unassigned peptidases (<i>Drosophila melanogaster</i>) [C48.UPW]	Sw:Q14623
C48	SJCHGC05036	>gb AAP06338.1 similar to AE003822 CG8493 gene product in <i>Drosophila melanogaster</i> [<i>Schistosoma japonicum</i>]	H102, D119, Q156, C162	MER11488	4.50E-22	family C48 unassigned peptidases (<i>Drosophila melanogaster</i>) [C48.UPW]	{peptidase unit: 23-177} from GB:AAM68667
C48	Sjp_0033730	sentrin-specific protease 1,K08592, putative	H456, D473, Q520, C526, H236, D331, L373Q, E379C	MER11012	7.5E-38	SENPI peptidase (<i>Homo sapiens</i>) [C48.002]	{peptidase unit: 456-614} from Sw:Q9P0U3
C48	Sjp_0074570	sentrin-specific protease 7,K08596, putative	D331, L373Q, E379C	MER14095	5.6E-14	SENPI peptidase (<i>Homo sapiens</i>) [C48.009]	{peptidase unit: 760-951} from Sw:Q9BQF6
C50	Sjp_0089510	separase, K02365,[PATH ko04110,PATH ko04111], putative	H1188, C1214	MER64756	6.8E-32	separase (<i>Xenopus laevis</i>) [C50.001]	{peptidase unit: 1372-1876} from GB:AAH70872
C54	Sjp_0081190	autophagy-related protein 4,K08342,[PATH ko04140], putative	Y56, C77, >D, >H	MER13561	5.5E-48	autophagin-1 (<i>Homo sapiens</i>) [C54.003]	{peptidase unit: 28-342} from Sw:Q9Y4P1
C54	SJCHGC05841	SJCHGC05841, similar to AAH76463-Unknown (protein for IMAGE:7037575) (<i>Danio rerio</i>)	Y56, C77, >D, >H	MER13561	2.80E-49	autophagin-1 (<i>Homo sapiens</i>) [C54.003]	{peptidase unit: 28-342} from Sw:Q9Y4P1
C54	SJCHGC04295	SJCHGC04295, weak similar to APG4 autophagy 4 homolog C	Y102, >C, >D, >H	MER14316	2.90E-05	autophagin-3 (<i>Homo sapiens</i>) [C54.004]	{peptidase unit: 43-407} from Sw:Q96DT6

C56	Sjp_0008070	protein DJ-1, K05687, putative	S77E, C104, A105H	MER11449	1.2E-30	DJ-1 putative peptidase (<i>Drosophila melanogaster</i>) [C56.002]	{peptidase unit: 47-162} from GB:BAB84672
C56	SJCHGC05973	SJCHGC05973, similar to CG1349-PA (<i>Drosophila melanogaster</i>)	S77E, C104, A105H	MER11449	1.20E-30	DJ-1 putative peptidase (<i>Drosophila melanogaster</i>) [C56.002]	{peptidase unit: 47-162} from GB:BAB84672
C64	Sjp_0026540	Zinc finger Ran-binding domain-containing protein 1 (Protein TRABID), putative	C550, H689	MER29052	1.6E-61	TRABID protein (<i>Homo sapiens</i>) [C64.004]	{peptidase unit: 415-653} from Sw:Q9UGI0
C65	SJCHGC02577	>gb AAP06318.1 similar to AE003628 CG4968 gene product in <i>Drosophila melanogaster</i> [<i>Schistosoma japonicum</i>]	D46, C49, >H	MER56292	1.60E-45	otubain-1 (<i>Canis familiaris</i>) [C65.001]	{peptidase unit: 1-241}
C78	Sjp_0082310	Probable Ufm1-specific protease 2, putative	C46, H173	MER41174	3.8E-36	UfSP1 peptidase (<i>Mus musculus</i>) [C78.001]	{peptidase unit: 1-217} from GB:BAB28174
C78	SJCHGC03148	SJCHGC03148, similar to expressed protein (<i>Arabidopsis thaliana</i>)	C46, >H	MER41174	2.70E-13	UfSP1 peptidase (<i>Mus musculus</i>) [C78.001]	{peptidase unit: 1-217} from GB:BAB28174
C78	Sjp_0087200	expressed protein, putative	C274, >H	MER60310	6.7E-26	UfSP2 peptidase (<i>Mus musculus</i>) [C78.002]	{peptidase unit: 253-454} from GB:BA C37214
C78	SJCHGC07577	SJCHGC07577, similar to ENSANGP00000015223 (<i>Anopheles gambiae</i>)	C77, >H	MER60310	1.40E-25	UfSP2 peptidase (<i>Mus musculus</i>) [C78.002]	{peptidase unit: 253-454} from GB:BA C37214

Supplementary Table 25. The predicted serine proteases of *S. japonicum*

Protease family	Protein name	Annotation	Predicted active site residues	Hit name	E-value comparing MEROPS	Annotation in MEROPS	Peptidase unit
S01A	Sjp_0028090	Cercarial protease precursor	H67, D125, X217S	MER03620	2E-73	cercarial elastase {{Schistosoma}} (Schistosoma mansoni) [S01.144]	{peptidase unit: 1-264} from Sw:P12546
S01A	Sjp_0012180	protease, serine, 7 (enterokinase), K01316, putative	H292, D341, S437	MER54210	2.1E-45	enteropeptidase (Gallus gallus) [S01.156]	{peptidase unit: 805-1039} from GB:XP_425539
S01A	Sjp_0047680	Transmembrane serine protease 8 precursor (EC 3.4.21.-)	F68H, D118, S223	MER04442	3.4E-39	intestinal serine peptidase {rodent} (Mus musculus) [S01.042]	{peptidase unit: 37-279} from Sw:Q9QYZ9
S01A	Sjp_0023390	Transmembrane protease, serine 6 (EC 3.4.21.-)	H113, D163, S259	MER05278	8E-45	matriptase-2 (Homo sapiens) [S01.308]	{peptidase unit: 568-802} from Sw:Q8IU80
S01A	Sjp_0100980	Trypsin precursor (EC 3.4.21.59), putative	H70, D135, >S	MER11236	8.1E-20	subfamily S1A unassigned peptidases (Drosophila melanogaster) [S01.UPA]	{peptidase unit: 472-727} from GB:AAF49916
S01A	Sjp_0101080	Enteropeptidase (EC 3.4.21.9)	<H, <D, S187	MER11472	7.3E-15	subfamily S1A unassigned peptidases (Drosophila melanogaster) [S01.UPA]	{peptidase unit: 37-294} from GB:AAF57983
S01A	SJCHGC03379	enterokinase-like	<H, D59, S161	MER11500	1.80E-25	subfamily S1A unassigned peptidases (Drosophila melanogaster) [S01.UPA]	{peptidase unit: 7-250} from GB:AAF59008
S01A	SJCHGC01895	>gb AAP05908.1 similar to AB062685 ribosome-associated protein P40 in Bombyx mori [Schistosoma japonicum]	H248, D313, >S	MER11236	9.00E-20	subfamily S1A unassigned peptidases (Drosophila melanogaster) [S01.UPA]	{peptidase unit: 472-727} from GB:AAF49916
S08A	Sjp_0047820	membrane-bound transcription factor protease, site 1, K08653, putative	<D, H229, >N, >S	MER22444	1.3E-27	site-1 peptidase (Anopheles gambiae) [S08.063]	{peptidase unit: 331-627} from GB:EAA00127
S08A	Sjp_0037290	tripeptidyl-peptidase II, K01280, putative	<D, H243, X361N, >S	MER11484	1.3E-18	tripeptidyl-peptidase II (Drosophila melanogaster) [S08.090]	{peptidase unit: 92-609} from Sw:Q9V6K1
S08B	Sjp_0076700	Furin-like protease 1, isoforms 1/1-X/2 precursor, putative	<D, H35, >N, >S	MER00373	3.6E-47	furin-1 {insect} (Drosophila melanogaster) [S08.048]	{peptidase unit: 348-640} from Sw:P30430

S08B	Sjp_0107540	Furin-like protease 1, isoform 1-CRR precursor (EC 3.4.21.75)	D64, >H, >N, >S	MER00373	7.2E-13	furin-1 {insect} (<i>Drosophila melanogaster</i>) [S08.048]	{peptidase unit: 348-640} from Sw:P30430
S08B	Sjp_0092390	Proprotein convertase subtilisin/kexin type 6 precursor, putative	D141, H181, >N, >S	MER01603	7.6E-56	KPC-1 peptidase {{ <i>Caenorhabditis elegans</i> }} (<i>Caenorhabditis elegans</i>) [S08.047]	{peptidase unit: 199-489} from GB:CAB04086
S08B	Sjp_0037780	Endoprotease bli-4 precursor, putative	<D, <H, N47, S120	MER01603	2.2E-42	KPC-1 peptidase {{ <i>Caenorhabditis elegans</i> }} (<i>Caenorhabditis elegans</i>) [S08.047]	{peptidase unit: 199-489} from GB:CAB04086
S08B	Sjp_0018070	Furin-1 precursor (EC 3.4.21.75)	<D, <H, N160, S236	MER00376	8E-45	proprotein convertase 1 (<i>Homo sapiens</i>) [S08.072]	{peptidase unit: 143-441} from Sw:P29120
S08B	Sjp_0062970	proprotein convertase subtilisin/kexin type 2 (EC 3.4.21.94), K01360, putative	D206, H247, X349N, S424	MER28254	2.9E-140	subfamily S8B non-peptidase homologues (<i>Halocynthia roretzi</i>) [S08.UNB]	{peptidase unit: 141-463} from GB:BAC05491
S08B	SJCHGC02912	SJCHGC02912, similar to proprotein convertase PC2 precursor (<i>Rana ridibunda</i>)	D207, >H, >N, >S	MER28254	2.90E-25	subfamily S8B non-peptidase homologues (<i>Halocynthia roretzi</i>) [S08.UNB]	{peptidase unit: 141-463} from GB:BAC05491
S09A	Sjp_0016520	Prolyl endopeptidase (EC 3.4.21.26)	E93S, S181D, W222H	MER20513	6.2E-47	prolyl oligopeptidase (<i>Anopheles gambiae</i>) [S09.001]	{peptidase unit: 517-798} from GB:EAA14977
S09A	Sjp_0080730	prolyl oligopeptidase, K01322, putative	S550, D637, H676	MER28369	3.7E-104	prolyl oligopeptidase (<i>Xenopus laevis</i>) [S09.001]	{peptidase unit: 471-753} from GB:AAH98959
S09B	Sjp_0049200	dipeptidyl-peptidase 9, K08656, putative	S770, D848, H880	MER04923	7.4E-60	dipeptidyl-peptidase 9 (<i>Homo sapiens</i>) [S09.019]	{peptidase unit: 258-506} from Sw:Q86T12
S09B	Sjp_0033000	dipeptidyl-peptidase 4, K01278, putative	S399, D475, H507	MER00399	5.5E-39	fibroblast activation protein alpha subunit (<i>Homo sapiens</i>) [S09.007]	{peptidase unit: 498-758} from Sw:Q12884
S09B	SJCHGC07346	SJCHGC07346, similar to BAC34034-unnamed protein product (<i>Mus musculus</i>)	S4, D82, H114	MER11413	1.70E-34	S9 homologue {invertebrate} (<i>Drosophila melanogaster</i>) [S09.016]	{peptidase unit: 862-1100} from GB:AAS65214
S09C	Sjp_0065680	acylaminoacyl-peptidase, K01303, putative	S509, D602, H636	MER13962	1.5E-48	acylaminoacyl-peptidase (<i>Mus musculus</i>) [S09.004]	{peptidase unit: 477-728} from Sw:Q8R146

S09X	Sjp_0015690	cholinesterase,K01050, putative	S244, >D, >H	MER33188	2.8E-42	acetylcholinesterase (<i>Homo sapiens</i>) [S09.979]	{peptidase unit: 114-499} from Sw:P22303
S09X	SJCHGC07322	SJCHGC07322, similar to acetylcholinesterase precursor (<i>Loligo opalescens</i>)	>S, >D, >H	MER37861	4.00E-14	BC026374 protein {{ <i>Mus musculus</i> }} (<i>Mus musculus</i>) [S09.962]	{peptidase unit: 1-489} from GB:AAH26374
S09X	SJCHGC07451	SJCHGC07451, similar to carboxylesterase (Schizaphis graminum)	>S, >D, >H	MER37861	6.50E-11	BC026374 protein {{ <i>Mus musculus</i> }} (<i>Mus musculus</i>) [S09.962]	{peptidase unit: 1-489} from GB:AAH26374
S09X	Sjp_0004400	Hormone-sensitive lipase (EC 3.1.1.79) (HSL), putative	<S, D116, H146	MER43146	2E-10	family S9 unassigned peptidases (<i>Bacillus subtilis</i>) [S09.UPW]	{peptidase unit: 88-352} from GB:AAC12774
S09X	Sjp_0028510	lysophospholipase II,K06130,[PATH ko00564], putative	S57, D113, H146	MER89387	6.7E-21	family S9 unassigned peptidases (<i>Francisella tularensis</i>) [S09.UPW]	{peptidase unit: 88-199} from GB:ABO47484
S09X	SJCHGC02460	>gb AAD52700.1 lysophospholipase [<i>Schistosoma japonicum</i>]	S116, D172, H205	MER89387	2.60E-20	family S9 unassigned peptidases (<i>Francisella tularensis</i>) [S09.UPW]	{peptidase unit: 88-199} from GB:ABO47484
S09X	Sjp_0045440	acetylcholinesterase,K01049,[PA TH ko00564], putative	S248, >D, >H	MER90438	2.5E-35	family S9 unassigned peptidases (<i>Geobacillus thermodenitrificans</i>) [S09.UPW]	{peptidase unit: 70-430} from GB:ABO68347
S09X	Sjp_0004410	hormone-sensitive lipase,K07188,[PATH ko04910], putative	S396, >D, >H	MER33274	1.1E-44	hormone-sensitive lipase (<i>Homo sapiens</i>) [S09.993]	{peptidase unit: 621-873} from Sw:Q05469
S09X	Sjp_0026950	Abhydrolase domain-containing protein FAM108A precursor, putative	<S, D76, H105	MER17353	9.8E-35	Mernname-AA194 putative peptidase (<i>Homo sapiens</i>) [S09.052]	{peptidase unit: 82-305} from GB:BAB15709
S09X	Sjp_0116590	Abhydrolase domain-containing protein FAM108C1, putative	S172, >D, >H	MER30290	1.1E-46	Mernname-AA197 putative peptidase (<i>Mus musculus</i>) [S09.055]	{peptidase unit: 62-285} from GB:NP_666208
S09X	Sjp_0036280	Acetylcholinesterase precursor (EC 3.1.1.7) (AChE), putative	S319, >D, >H	MER33235	1.6E-27	neurologin 4, X-linked (<i>Homo sapiens</i>) [S09.988]	{peptidase unit: 137-510} from Sw:Q8N0W4
S09X	Sjp_0019620	carboxylesterase (EC 3.1.1.1), K01044, putative	G235S, >D, >H	MER33236	3.9E-37	neurologin 4, Y-linked (<i>Homo sapiens</i>) [S09.989]	{peptidase unit: 137-510} from Sw:Q8NFFZ3
S09X	Sjp_0008850	Neurologin-4, X-linked precursor (Neurologin X)	>S, >D, >H	MER33236	1.1E-14	neurologin 4, Y-linked (<i>Homo sapiens</i>) [S09.989]	{peptidase unit: 137-510} from Sw:Q8NFFZ3

S09X	SJCHGC06633	SJCHGC06633 Bat5-prov protein (<i>Xenopus laevis</i>)	<S, D65, >H	MER37840	2.30E-25	protein bat5 (<i>Homo sapiens</i>) [S09.065]	{peptidase unit: 264-545} from Sw:O95870
S10	Sjp_0024270	Lysosomal protective protein precursor, putative	S183, D421, >H	MER45439	1.6E-83	serine carboxypeptidase A (<i>BrachyDanio rerio</i>) [S10.002]	{peptidase unit: 28-471} from GB:AAH56531
S10	SJCHGC09294	SJCHGC09294, similar to protective protein for beta-galactosidase	S87, D325, >H	MER45439	4.10E-82	serine carboxypeptidase A (<i>BrachyDanio rerio</i>) [S10.002]	{peptidase unit: 28-471} from GB:AAH56531
S10	Sjp_0105430	carboxypeptidase C,K01287, putative	>S, >D, >H	MER53737	1.6E-27	serine carboxypeptidase A (<i>Canis familiaris</i>) [S10.002]	{peptidase unit: 58-499}
S10	Sjp_0023750	Lysosomal protective protein precursor (EC 3.4.16.5) (Cathepsin A)	<S, D221, H282	MER00430	6E-43	serine carboxypeptidase A (<i>Homo sapiens</i>) [S10.002]	{peptidase unit: 39-480} from Sw:P10619
S12	Sjp_0006630	Serine beta-lactamase-like protein LACTB, putative	S103, K106, >Y	MER05923	3.2E-26	LACT-1 peptidase (<i>Mus musculus</i>) [S12.004]	{peptidase unit: 100-551} from Sw:Q9EP89
S12	SJCHGC01499	SJCHGC01499, similar to NP_502657-lactamase beta (40546) (<i>Caenorhabditis elegans</i>)	S103, K106, >Y	MER05923	1.10E-28	LACT-1 peptidase (<i>Mus musculus</i>) [S12.004]	{peptidase unit: 100-551} from Sw:Q9EP89
S14	Sjp_0046620	ATP-dependent Clp protease, protease subunit,K01358, putative	X139S, H164, D213	MER02211	1.2E-66	peptidase Clp {type 3} (<i>Homo sapiens</i>) [S14.003]	{peptidase unit: 43-249} from Sw:Q16740
S16	Sjp_0044150	Lon-like ATP-dependent protease,K08675, putative	S972, K1015	MER00495	1.6E-47	PIM1 peptidase (<i>Homo sapiens</i>) [S16.002]	{peptidase unit: 705-959} from Sw:P36776
S16	SJCHGC03265	SJCHGC03265, similar to Lon protease-like protein (<i>Homo sapiens</i>)	S34, K77	MER00495	9.40E-48	PIM1 peptidase (<i>Homo sapiens</i>) [S16.002]	{peptidase unit: 705-959} from Sw:P36776
S26A	SJCHGC08565	SJCHGC08565, weak similar to CAC39221-dJ1137017.1	<S, K25	MER13949	3.20E-14	mitochondrial signal peptidase {metazoa} (<i>Homo sapiens</i>) [S26.013]	{peptidase unit: 27-144} from GB:CAC39221
S26B	Sjp_0058200	signal peptidase I,K03100,[PATH ko03060], putative	S58, H98	MER63473	7.2E-48	signalase {eukaryote} 21 kDa component (<i>Bombyx mori</i>) [S26.010]	{peptidase unit: 43-179} from GB:ABD36356
S26B	SJCHGC06703	SJCHGC06703, similar to NP_150596-similar to signal peptidase complex (18kD) (<i>Homo sapiens</i>)	S58, H98	MER63473	6.00E-60	signalase {eukaryote} 21 kDa component (<i>Bombyx mori</i>) [S26.010]	{peptidase unit: 43-179} from GB:ABD36356

S28	SJCHGC02147	SJCHGC02147, similar to dipeptidylpeptidase 7	S156, D415, H440	MER06058	3.20E-118	dipeptidyl-peptidase II (<i>Rattus norvegicus</i>) [S28.002]	{peptidase unit: 39-494} from Sw:Q9EPB1
S28	Sjp_0085250	lysosomal Pro-X carboxypeptidase, K01285, putative	S164, D415, >H	MER00446	2.1E-98	lysosomal Pro-Xaa carboxypeptidase (<i>Homo sapiens</i>) [S28.001]	{peptidase unit: 46-496} from Sw:P42785
S28	SJCHGC06818	>gb AAP05884.1 similar to prolylcarboxypeptidase (angiotensinase C) in <i>Homo sapiens</i> [<i>Schistosoma japonicum</i>]	S164, >D, >H	MER00446	2.80E-64	lysosomal Pro-Xaa carboxypeptidase (<i>Homo sapiens</i>) [S28.001]	{peptidase unit: 46-496} from Sw:P42785
S28	SJCHGC06819	SJCHGC06819, similar to ENSANGP00000018571 (<i>Anopheles gambiae</i>)	<S, D248, H274	MER00446	5.40E-63	lysosomal Pro-Xaa carboxypeptidase (<i>Homo sapiens</i>) [S28.001]	{peptidase unit: 46-496} from Sw:P42785
S33	Sjp_0084990	Abhydrolase domain-containing protein 4 (EC 3.1.1.-)	S88, >D, >H	MER11509	3.2E-50	family S33 unassigned peptidases (<i>Drosophila melanogaster</i>) [S33.UPW]	{peptidase unit: 77-391} from GB:AAO41479
S33	Sjp_0085000	Abhydrolase domain-containing protein 4 (EC 3.1.1.-)	<S, D47, H73	MER11509	4.6E-24	family S33 unassigned peptidases (<i>Drosophila melanogaster</i>) [S33.UPW]	{peptidase unit: 77-391} from GB:AAO41479
S33	SJCHGC04331	SJCHGC04331, similar to CG1882-PA (<i>Drosophila melanogaster</i>)	<S, D25, H51	MER11509	1.20E-17	family S33 unassigned peptidases (<i>Drosophila melanogaster</i>) [S33.UPW]	{peptidase unit: 77-391} from GB:AAO41479
S33	SJCHGC00650	SJCHGC00650, similar to misexpression suppressor of KSR 2 in <i>Drosophila melanogaster</i> (<i>Schistosoma japonicum</i>)	G140S, D277, G308H	MER42914	6.10E-45	NDRG3 protein (<i>Homo sapiens</i>) [S33.987]	{peptidase unit: 15-301} from Sw:Q9UGV2
S33	Sjp_0062170	esterase / lipase, K01066, [PATH ko00650, PATH ko00960, PATH ko00623], putative	S138, >D, >H	MER37853	1.2E-51	protein phosphatase methyltransferase 1 (<i>Homo sapiens</i>) [S33.984]	{peptidase unit: 53-371} from Sw:Q9Y570
S33	SJCHGC03406	SJCHGC03406, similar to BAC36554-unnamed protein product (<i>Mus musculus</i>)	S138, >D, >H	MER37853	5.90E-39	protein phosphatase methyltransferase 1 (<i>Homo sapiens</i>) [S33.984]	{peptidase unit: 53-371} from Sw:Q9Y570
S33	Sjp_0104720	Putative esterase/lipase H10193 (EC 3.1.-.-), putative	S95, S196D, H257	MER31610	2.1E-45	Williams-Beuren syndrome critical region protein 21 epoxide hydrolase (<i>Homo sapiens</i>) [S33.976]	{peptidase unit: 15-308} from GB:AAO7472

S33	SJCHGC09171	SJCHGC09171, similar to XP_341105-similar to Williams-Beuren syndrome critical region protein 21 (<i>Rattus norvegicus</i>)	S95, S196D, >H	MER31610	1.50E-37	Williams-Beuren syndrome critical region protein 21 epoxide hydrolase (<i>Homo sapiens</i>) [S33.976]	{peptidase unit: 15-308} from GB:AAS07472
S54	Sjp_0086430	Peptidase S54, rhomboid, domain-containing	>S, >H	MER04528	1.7E-37	RHBDF1 protein (<i>Homo sapiens</i>) [S54.952]	{peptidase unit: 403-855} from GB:BAB15318
S54	Sjp_0070660	rhomboid-related protein 1/2/3, K02857, putative	>S	MER15452	3.2E-15	Rhomboid-2 { {Drosophila}-type} (<i>Drosophila melanogaster</i>) [S54.010]	{peptidase unit: 167-431} from GB:AAO39503
S59	Sjp_0085990	Nuclear pore complex protein Nup98-Nup96 precursor	H974, S976	MER45454	3.7E-52	nucleoporin 145 (<i>Mus musculus</i>) [S59.001]	{peptidase unit: 710-937} from GB:AAH57608
S60	Sjp_0018270	melanoma-associated antigen p97, K06569, putative	-103K, E293S	MER33291	1.3E-57	melanotransferrin domain 1 (<i>Homo sapiens</i>) [S60.973]	{peptidase unit: 20-361} from Sw:P08582

Supplementary Table 26. The predicted threonine proteases of *S. japonicum*

Protease family	Protein name	Annotation	Predicted active site residues	Hit name	E-value comparing MEROPS	Annotation in MEROPS	Peptidase unit
T01A	Sjp_0080900	20S proteasome subunit beta 1,K02738,[PATH ko03050], putative	T19	MER25535	1.2E-64	proteasome catalytic subunit 1 (<i>Anopheles gambiae</i>) [T01.010]	{peptidase unit: 27-216} from GB:EAA14913
T01A	SJCHGC05732	>gb AAP06465.1 similar to XM_027825 proteasome (prosome, macropain) subunit, beta type 6 in <i>Homo sapiens</i> [<i>Schistosoma japonicum</i>]	T19	MER25535	3.60E-64	proteasome catalytic subunit 1 (<i>Anopheles gambiae</i>) [T01.010]	{peptidase unit: 27-216} from GB:EAA14913
T01A	Sjp_0002420	20S proteasome subunit beta 2,K02739,[PATH ko03050], putative	T39	MER20025	6.6E-72	proteasome catalytic subunit 2 (<i>Anopheles gambiae</i>) [T01.011]	{peptidase unit: 47-248} from GB:EAA13087
T01A	Sjp_0062270	20S proteasome subunit beta 5,K02737,[PATH ko03050], putative	T63	MER03132	3.9E-91	proteasome catalytic subunit 3 (Botryllus schlosseri) [T01.012]	{peptidase unit: 57-257} from GB:CAA66313
T01A	SJCHGC05978	SJCHGC05978, similar to AAM18885-unknown (<i>Branchiostoma floridae</i>)	T63	MER03132	3.90E-91	proteasome catalytic subunit 3 (Botryllus schlosseri) [T01.012]	{peptidase unit: 57-257} from GB:CAA66313
T01A	Sjp_0004200	20S proteasome subunit alpha 6,K02725,[PATH ko03050], putative	S33T	MER56264	3.1E-78	proteasome subunit alpha 1 (<i>Pan troglodytes</i>) [T01.976]	{peptidase unit: 20-235} from GB:XP_508298
T01A	SJCHGC09091	Psm2	P33T	MER04582	4.10E-80	proteasome subunit alpha 2 (Carassius auratus) [T01.972]	{peptidase unit: 32-225} from Sw:O73672
T01A	Sjp_0102270	20S proteasome subunit alpha 7,K02727,[PATH ko03050], putative	<T	MER00553	3.1E-63	proteasome subunit alpha 3 (<i>Homo sapiens</i>) [T01.977]	{peptidase unit: 35-238} from Sw:P25788

T01A	Sjp_0060160	20S proteasome subunit alpha 3, K02728, [PATH ko03050], putative		MER00554	3.1E-80	proteasome subunit alpha 4 (<i>Homo sapiens</i>) [T01.973]	{peptidase unit: 33-239} from Sw:P25789
T01A	SJCHGC05022	SJCHGC05022, similar to AAH76196-Unknown (protein for MGC:92716) (<i>Danio rerio</i>)	<T	MER29424	2.80E-68	proteasome subunit alpha 6 (<i>Bos taurus</i>) [T01.971]	{peptidase unit: 38-242} from Sw:Q2YDE4
T01A	Sjp_0024000	20S proteasome subunit alpha 1, K02730, [PATH ko03050], putative		MER00557	3.5E-43	proteasome subunit alpha 6 (<i>Homo sapiens</i>) [T01.971]	{peptidase unit: 24-140} from Sw:P60900
T01A	Sjp_0064130	20S proteasome subunit alpha 8, K06690, putative		MER04372	7E-80	proteasome subunit alpha 7 (<i>Homo sapiens</i>) [T01.974]	{peptidase unit: 24-232} from Sw:O14818
T01A	SJCHGC04094	SJCHGC04094, similar to XP_393583-similar to ENSANGP0000007022 (<i>Apis mellifera</i>)		MER04372	3.90E-40	proteasome subunit alpha 7 (<i>Homo sapiens</i>) [T01.974]	{peptidase unit: 24-232} from Sw:O14818
T01A	Sjp_0090370	20S proteasome subunit beta 6, K02732, [PATH ko03050], putative	<T	MER00551	1.9E-52	proteasome subunit beta 1 (<i>Homo sapiens</i>) [T01.986]	{peptidase unit: 39-230} from Sw:P20618
T01A	Sjp_0075650	20S proteasome subunit beta 4, K02734, [PATH ko03050], putative	N10T	MER02676	1.4E-48	proteasome subunit beta 2 (<i>Homo sapiens</i>) [T01.984]	{peptidase unit: 5-147} from Sw:P49721
T01A	Sjp_0091190	expressed protein, putative	<T	MER80407	8.4E-15	proteasome subunit beta 3 (<i>Aedes aegypti</i>) [T01.983]	{peptidase unit: 9-205} from GB:EA42270
T01A	SJCHGC02212	SJCHGC02212, similar to AAP06451-similar to NM_011971 proteasome (prosome, macropain) subunit, beta type 3 in <i>Mus musculus</i> (<i>Schistosoma japonicum</i>)	<T	MER00527	1.20E-62	proteasome subunit beta 3 (<i>Bos taurus</i>) [T01.983]	{peptidase unit: 9-205} from Sw:P33672
T01A	Sjp_0068310	20S proteasome subunit beta 7, K02736, [PATH ko03050], putative	T31	MER00559	4.5E-58	proteasome subunit beta 4 (<i>Xenopus laevis</i>) [T01.987]	{peptidase unit: 32-233} from Sw:P28024
T01A	SJCHGC06662	SJCHGC06662, similar to proteasome subunit N3 (<i>Oncorhynchus mykiss</i>)	T31	MER00559	2.30E-57	proteasome subunit beta 4 (<i>Xenopus laevis</i>) [T01.987]	{peptidase unit: 32-233} from Sw:P28024

T01A	Sjp_0020330	Rho-associated protein kinase 1 (EC 2.7.11.1)	S35T	MER35527	4.9E-53	subfamily T1A non-peptidase homologues (<i>Schistosoma japonicum</i>) [T01.UNA]	{peptidase unit: 35-241} from GB:AP06025
T02	Sjp_0064670	N4-(beta-N-acetylglucosaminyl)-L-asparaginase,K01444,[P ATH ko00511], putative	T229	MER06143	3.3E-33	glycosylasparaginase precursor (<i>Arabidopsis thaliana</i>) [T02.001]	{peptidase unit: 241-376} from Sw:Q56W64
T02	Sjp_0074330	taspase, threonine aspartase, 1,K08657, putative	T207	MER61559	2.8E-30	isoaspartyl dipeptidase {threonine type} (Burkholderia xenovorans) [T02.002]	{peptidase unit: 7-316} from GB:YP_560152
T02	SJCHGC02349	SJCHGC02349, similar to CAG09520-unnamed protein product (<i>Tetradon nigroviridis</i>)	T198	MER61559	2.10E-30	isoaspartyl dipeptidase {threonine type} (Burkholderia xenovorans) [T02.002]	{peptidase unit: 7-316} from GB:YP_560152
T02	SJCHGC09117	SJCHGC09117, similar to XP_394866-similar to N (4)-(Beta-N-acetylglucosaminyl)-L-asparaginase	T217	MER85244	6.60E-36	isoaspartyl dipeptidase {threonine type} (Roseiflexus sp. RS-1) [T02.002]	{peptidase unit: 5-319} from GB:EA.T25092
T03	Sjp_0004210	gamma-glutamyltransferase	>S/T	MER16970	4.5E-21	gamma-glutamyltransferase-like protein 3 (<i>Homo sapiens</i>) [T03.017]	{peptidase unit: 118-664} from Sw:Q9UJ14
T03	SJCHGC03544	SJCHGC03544, similar to JC7331 gamma-glutamyltransferase (EC 2.3.2.2) homolog - rat	>S/T	MER16970	9.20E-22	gamma-glutamyltransferase-like protein 3 (<i>Homo sapiens</i>) [T03.017]	{peptidase unit: 118-664} from Sw:Q9UJ14

Supplementary Table 27. The predicted aspartic proteases of *S. japonicum*

Protease family	Protein name	Annotation	Predicted active site residues	Hit name	E-value comparing MEROPS	Annotation in MEROPS	Peptidase unit
A01A	Sjp_0006800	Cathepsin D	D84, Y129, D270	MER52779	5.2E-131	cathepsin D (Opisthorchis viverrini) [A01.009]	{peptidase unit: 56-386} from GB:AAZ39883
A01A	Sjp_0049310	Lysosomal aspartic protease precursor (EC 3.4.23.-), putative	D478, Y523, D663	MER52779	5.8E-73	cathepsin D (Opisthorchis viverrini) [A01.009]	{peptidase unit: 56-386} from GB:AAZ39883
A01A	Sjp_0073710	Cathepsin D precursor (EC 3.4.23.5), putative	D23, Y68, >D	MER52779	8.7E-38	cathepsin D (Opisthorchis viverrini) [A01.009]	{peptidase unit: 56-386} from GB:AAZ39883
A01A	Sjp_0073700	Lysosomal aspartic protease precursor (EC 3.4.23.-), putative	<D/H, <F/S/Y, D15	MER52779	6.5E-27	cathepsin D (Opisthorchis viverrini) [A01.009]	{peptidase unit: 56-386} from GB:AAZ39883
A01A	Sjp_0073720	Lysosomal aspartic protease precursor (EC 3.4.23.-), putative	<D/H, <F/S/Y, D12	MER52779	1.1E-25	cathepsin D (Opisthorchis viverrini) [A01.009]	{peptidase unit: 56-386} from GB:AAZ39883
A01A	Sjp_0051260	beta-site APP-cleaving enzyme 2 (memapsin 1) (EC 3.4.23.45), K07747, putative	<D/H, <F/S/Y, >D	MER05870	0.00000003	memapsin-2 (<i>Homo sapiens</i>) [A01.004]	{peptidase unit: 52-428} from Sw:P56817
A01A	Sjp_0007750	Lysosomal aspartic protease precursor (EC 3.4.23.-), putative	D96, Y141, >D	MER62900	2.7E-62	subfamily A1A unassigned peptidases (<i>Schistosoma mansoni</i>) [A01.009]	{peptidase unit: 67-392} from GB:ABB77194
A01A	Sjp_0085930	Cathepsin D precursor (EC 3.4.23.5), putative	<D/H, <F/S/Y, D24	MER62900	4.7E-62	subfamily A1A unassigned peptidases (<i>Schistosoma mansoni</i>) [A01.009]	{peptidase unit: 67-392} from GB:ABB77194
A01A	SJCHGC02875	SJCHGC02875, similar to aspartic proteinase (<i>Schistosoma mansoni</i>)	D96, Y141, D268	MER62900	5.50E-137	subfamily A1A unassigned peptidases (<i>Schistosoma mansoni</i>) [A01.009]	{peptidase unit: 67-392} from GB:ABB77194

A02X	Sjp_0008720	Retrovirus-related Pol polyprotein from transposon opus, Includes: Protease, putative	D47	MER16568	2.1E-24	family A2 unassigned peptidases (Clonorchis sinensis) [A02.UPW]	{peptidase unit: 278-375} from GB:AAK07487
A02X	Sjp_0002790	Transposon Ty3-G Gag-Pol polyprotein, putative	D165	MER16568	3.9E-21	family A2 unassigned peptidases (Clonorchis sinensis) [A02.UPW]	{peptidase unit: 278-375} from GB:AAK07487
A22A	Sjp_0071020	presenilin-1, K04505, [PATH ko04310, PATH ko04330, PATH ko05010], putative	D233, D405	MER05227	1.4E-101	Psn peptidase {{ <i>Drosophila melanogaster</i> }} (<i>Drosophila melanogaster</i>) [A22.014]	{peptidase unit: 35-541} from Sw:O02194
A22B	SJCHGC09361	>gb AAP05959.1 similar to histocompatibility 13, (NM_078720 shanti in <i>Drosophila melanogaster</i> [<i>Schistosoma japonicum</i>]	D219, D265	MER85383	9.40E-84	impas 1 peptidase (Aedes aegypti) [A22.003]	{peptidase unit: 11-397} from GB:EAT35878

Supplementary Table 28. Elastase-encoding sequences of schistosomes

Name	DEFINITION	Protein accession	gi (protein)	length (aa)	DNA accession	Species
SjCE-2b	cercarial protease precursor (Cercarial elastase).	Sjp_0028090	-	263	Sjp_0028090	<i>S. japonicum</i>
SmCE-1a	cercarial protease precursor (Cercarial elastase).	P12546.1	GI:116116	264	gi 21217526 gb AF510337.1	<i>S. mansoni</i>
SmCE-1a*	cercarial elastase HP1 [<i>S. mansoni</i>].	CAA94312.1	GI:1240058	213	embl Z70296.1 gi 1240057	<i>S. mansoni</i>
SmCE-1b	elastase	AAC46967.1	GI:1103829	274	U31768.1 gi 1103828	<i>S. mansoni</i>
SmCE-1c	elastase	AAC46968.1	GI:1103831	257	U31769.1 gi 1103830	<i>S. mansoni</i>
SmCE-2a	elastase 2a [<i>S. mansoni</i>].	AAM43941.1	GI:21217531	263	gi 21217530 gb AF510339.1	<i>S. mansoni</i>
SmCE-2b	elastase 2b [<i>S. mansoni</i>].	AAM43942.1	GI:21217533	263	gi 21217532 gb AF510340.1	<i>S. mansoni</i>
SdCE-1a	elastase 1a [<i>Schistosomatum douthitti</i>].	AAM43945.1	GI:21217539	144	gi 21217538 gb AF510343.1	<i>Schistosomatum douthitti</i>
SdCE-1b	elastase 1b (CE-1b) gene	AAM43946.1	GI:21217541	46	gi 21217540 gb AF510344.1	<i>Schistosomatum douthitti</i>
ShCE-1a	elastase 1a [<i>S. haematobium</i>].	AAM43943.1	GI:21217535	263	gi 21217534 gb AF510341.1	<i>S. haematobium</i>
ShCE-1b	elastase 1b [<i>S. haematobium</i>].	AAM43944.1	GI:21217537	139	gi 21217536 gb AF510342.1	<i>S. haematobium</i>

*This is a partial sequence of SmCE-1a.

Supplementary Table 29. Immune defense molecules in *S. japonicum*

Name	Immune system related gene	<i>S. japonicum</i> predicted gene	<i>S. japonicum</i> cDNA
Toll-interleukin 1 receptor domain (TIR)	TIR motif-containing protein -1 homolog	Sjp_0057800	FSE001-P00007-E23 FSE001-P00016-P10 FSE001-P00035-L07 SJCHGC00800
PCD	CASP3	Sjp_0009160	
	CASP2	Sjp_0071380	
	CASP7	Sjp_0039420	FSE001-P00005-K03 FSE001-P00013-A03 FSE001-P00019-I15 FSE001-P00032-C07 FSE001-P00035-L20 FSE_202_23_G03, FSE_202_23_G03 FSA_130_21_E03 FSE_224_46_F06 SJCHGC03476
Leucin-rich repeat (LRR)	Leucine-rich repeat-containing protein	Sjp_0017510 Sjp_0120280 Sjp_0002890	
Lectins	Lectin	Sjp_0096850	
Complement	macroglobulin/complement	Sjp_0062670 Sjp_0012630	SJCHGC09210 SJCHGC02849
Scavenger receptor	Scavenger receptor class B, member 2	Sjp_0063070 Sjp_0073560 Sjp_0121730	
Ig superfamily (Igsf)	Immunoglobulin domain-containing protein	Sjp_0095500 Sjp_0113060	

Supplementary Table 30. TIR downstream signaling of *S. japonicum*

Gene name	<i>S. japonicum</i> predicted gene	<i>S. japonicum</i> cDNA
TIR	Sjp_0057800	
ASK-1	Sjp_0069170	FSA001-P00006-B07
MKK6	Sjp_0084080	
	Sjp_0034990	
	Sjp_0112530	
MAPK p38	Sjp_0065080	FSE_037_38_F05
	Sjp_0065090	FSE_037_38_F05
	Sjp_0108760	

Supplementary Table 31: Enzymes involved with glycan and lipid synthesis in *S. japonicum*

Catogory	Gene name	<i>S. japonicum</i> predicted gene	<i>S. japonicum</i> cDNA
N-glycan biosynthesis			
	alpha-1,6-mannosyl-glycoprotein	Sjp_0012740	-
	Dolichol kinase	Sjp_0060110	-
	dolichyl-phosphate beta-glucosyltransferase	Sjp_0059720	SJCHGC01491, FSE001-P00021-H05, FSE001-P00004-I01
	Alpha-1,3(6)-mannosylglycoprotein-beta-1,6-N-acetylglucosaminyl transferase	Sjp_0092620	-
	Geranylgeranyl transferase type II beta subunit	Sjp_0090410, Sjp_0056380	-
	Dolichyl diphosphatase	Sjp_0021010	SJCHGC01705, FSE001-P00023-G24, FSE001-P00002-P06, FSE001-P00002-P06
	Dolichol-phosphate mannosyltransferase	Sjp_0062550	SJCHGC06465, FSE001-P00017-J05
	UDP-N-acetylglucosamine-dolichylphosphate	Sjp_0056980	-
	Beta-1,4-N-acetylglucosaminyltransferase	Sjp_0019650	SJCHGC01011, FSE001-P00035-G05, FSE001-P00003-H10, FSA001-P00026-J17
	Beta-1,-4-mannosyltransferase	Sjp_0077160	FSE001-P00025-I02, FSE001-P00025-I02, FSE001-P00009-H22, FSE001-P00009-H22
	Alpha-1,3-mannosyltransferase	Sjp_0026500	FSE001-P00016-I23, FSE001-P00016-I23, SJCHGC09545, SJCHGC05150
	Alpha-1,2-mannosyltransferase	Sjp_0106360	FSA001-P00011-J01
	Oligosaccharidyl-lipid f-lippase family	Sjp_0082000	-
	Alpha-1,6-mannosyltransferase	Sjp_0092220	FSE001-P00004-F18, FSE001-P00004-F18, FSA001-P00022-H10, FSA001-P00022-H10
	Alpha-1,3-glucosyltransferase	Sjp_0089670, Sjp_0037030	SJCHGC04950, FSE001-P00001-F13, FSE001-P00029-G11, FSE001-P00029-G11
	Alpha-1,2-glucosyltransferase	Sjp_0018130	FSE001-P00004-F14, FSE001-P00004-F14, FSE001-P00017-G05

Dolichyl-diphosphooligosaccharide--protein glycosyltransferase	Sjp_0110220, Sjp_0001310, Sjp_0066590, Sjp_0055830, Sjp_0040630	FSE001-P00006-N15, FSE001-P00005-J21, FSE001-P00003-G02, FSE001-P00001-J14, FSA001-P00030-I12, FSA001-P00004-B20, FSE001-P00034-D07, FSA001-P00006-F01, FSA001-P00010-P12, FSA001-P00020-P02, FSA001-P00032-O03, FSA001-P00033-O20, FSE001-P00012-I03, FSE001-P00026-P02, FSE001-P00034-O12, SJCHGC02763, FSA001-P00019-P21, FSA001-P00019-P21, FSA001-P00019-P21, FSA001-P00005-F22, FSA001-P00005-F22
Mannosyl-oligosaccharide glucosidase	Sjp_0060870, Sjp_0060890	-
Alpha 1,3-glucosidase	Sjp_0043610	SJCHGC00659
Beta-N-acetylglucosaminylglycopeptide beta-1,4-galactosyltransferase	Sjp_0056950, Sjp_0109090, Sjp_0060240	-
Glycoprotein 2-beta-D-xylosyltransferase	Sjp_0055390	FSE001-P00021-E12, FSE001-P00020-M12
Alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase	Sjp_0048170	-
Glycoprotein 6-alpha-L-fucosyltransferase	Sjp_0094930, Sjp_0071340, Sjp_0071320	FSE_162_37_E05, FSE_074_65_A09
UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase	Sjp_0000680	SJCHGC01813
Mannosyl-oligosaccharide alpha-1,2-mannosidase	Sjp_0029600, Sjp_0036240	SJCHGC03680
O-glycan biosynthesis		
Polypeptide GalNAc transferase	Sjp_0015020	FSE_222_31_G04, FSE_100_47_G06 FSE_210_78_F10, SJCHGC06745, SJCHGC05975, SJCHGC02258, FSE_100_41_A06, FSE_087_21_E03, SJCHGC05177
Beta-1,3-galactosyltransferase brn	Sjp_0133490, Sjp_0130670, Sjp_0049800, Sjp_0039270	-
Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyl transferase	Sjp_0030180	-
Galactosylceramide biosynthesis		
Alpha-galactosidas	Sjp_0023240, Sjp_0023260	SJCHGC05460
Glycosphingolipid synthesis-lactoseries		

Beta-1,3-N-acetylglucosaminyltransferase	Sjp_0039280, Sjp_0039290, Sjp_0049790	SJCHGC03667, FSA_062_40_H05
Beta-1,3-galactosyltransferase	Sjp_0024550, Sjp_0102690, Sjp_0024550	SJCHGC05177
Galactoside 3(4)-fucosyltransferase	Sjp_0044100	-
Beta-1,4-galactosyltransferase	Sjp_0079160, Sjp_0079160	FSE001-P00013-P08, SJCHGC01813
Ceramide-glucosyltransferase	Sjp_0093880, Sjp_0054080	-
Beta-galactosidase	Sjp_0072010	-
Glycosphingolipid synthesis-neo-lactoseries		
Beta-1,3-N-acetylglucosaminyltransferase	Sjp_0039280, Sjp_0039290, Sjp_0049790	SJCHGC03667
4-galactosyl-N-acetylglucosaminide 3-alpha-L-fucosyltransferase	Sjp_0036210	SJCHGC09057
Galactoside alpha-1,3-fucosyltransferase 4	Sjp_0012840	-
Galactoside 3(4)-fucosyltransferase	Sjp_0044100	-
Beta-1,4-galactosyltransferase	Sjp_0056950, Sjp_0079160	-
Ceramide glucosyltransferase	Sjp_0093880, Sjp_0054080	-
Glycosphingolipid synthesis-globoseries		
Beta-1,3-N-acetylglactosaminyltransferase	Sjp_0039280, Sjp_0039290, Sjp_0049790	SJCHGC03667, FSA_062_40_H05
Alpha-N-acetylglactosaminidase	Sjp_0023220, Sjp_0107620, Sjp_0023240, Sjp_0023250, Sjp_0023260	FSA001-P00032-J19, FSA001-P00023-B20, FSA001-P00015-J21
Alpha-galactosidase	Sjp_0023250, Sjp_0023220, Sjp_0107620, Sjp_0023240	SJCHGC05460
Beta-1,4-galactosyltransferase	Sjp_0056950, Sjp_0079160	FSE001-P00013-P08, SJCHGC01813
4-galactosyl-N-acetylglucosaminide 3-alpha-L-fucosyltransferase	Sjp_0036210	FSE_213_44_D06, FSE001-P00034-C08, SJCHGC04427, SJCHGC09057
Ceramide glucosyltransferase	Sjp_0093880, Sjp_0054080	-
Glycosphingolipid synthesis-ganglioseries		
Beta-N-acetylhexosaminidase	Sjp_0015430, Sjp_0019530, Sjp_0020650	-

Beta-galactosidase	Sjp_0072010	-	FSE001-P00023-M01, FSE_162_33_A05, FSE_102_66_B09, FSE_053_73_A10, FSA001-P00032-O02, FSA001-P00033-A20, FSA001-P00033-A20
MFS transporter	Sjp_0095020, Sjp_0075690, Sjp_0019920	-	FSE001-P00013-P08, SJCHGC01813
Beta-1,4-galactosyltransferase	Sjp_0056950, Sjp_0079160	-	
Ceramide glucosyltransferase	Sjp_0093880, Sjp_0054080	-	
Chondroitin/dermatan sulfate synthesis			
Galactosylgalactosylxylosylprotein3-beta-glucuronosyltransferase	Sjp_0062810	SJCHGC09394	
Xylosylprotein 4-bet-a-galactosyltransferase	Sjp_0062870	-	
Protein xylos-yltransferase	Sjp_0001670, Sjp_0132550	FSE001-P00021-E12, FSE001-P00020-M12, FSE001-P00023-K21, FSE001-P00023-K21	
Heparin sulfate synthesis			
Alpha-1,4-N-acetylglucosaminyltransferase EXTL	Sjp_0025120	-	
Glucuronyl-/N-acetylglucosaminyl transferase EXT	Sjp_0005240	-	
Galactosylgalacto-sylxylosylprotein3-beta-glucuronosyltransferase	Sjp_0062810	SJCHGC09394	
Xylosylprotein 4-bet-a-galactosyltransferase	Sjp_0062870	-	
Protein xylosyltransferase	Sjp_0001670, Sjp_0132550	FSE001-P00023-K21, FSE001-P00023-K21	
Heparan sulfate 2-O-sulfotransf-erase HS2ST1	Sjp_0060410	FSE001-P00001-D08	
Heparan sulfate N-deacetylase/N-sulfotransferase NDST1	Sjp_0082020	-	
Heparan sulfate 6-O-sulfotransferase HS6ST1	Sjp_0094660	FSE001-P00018-N11	
Keratin sulfate synthesis			
Beta-N-acetylglucosaminylglycopeptide beta-1,4-galactosyltransferase	Sjp_0109090, Sjp_0060240, Sjp_0056950	-	
Glycoprotein 6-alpha-L-fucosyltransferase	Sjp_0071340, Sjp_0026160, Sjp_0071320	-	
Glycolipid metabolism			

aldehyde dehydrogenase	Sjp_0011180	SJCHGC06511, SJCHGC06572, FSE_156_41_A06, FSE001-P00005-E21, SJCHGC05995, FSE001-P00028-I20, FSE001-P00028-I20 FSE001-P00021-I22, FSE001-P00013-G06, FSE001-P00011-A02, FSA001-P00024-J07, FSA001-P00018-G16, FSE001-P00024-A03, FSE001-P00019-E15, FSE001-P00012-K11, FSE001-P00010-J09, FSE001-P00010-C16, FSE001-P00006-I01, FSE001-P00006-D06 SJCHGC07000
aldehyde reductase	Sjp_0099430	
glycerol kinase	Sjp_0060360, Sjp_0034060, Sjp_0054690	
alpha-galactosidase	Sjp_0023250, Sjp_0023220, Sjp_0107620, Sjp_0023240	SJCHGC05460 FSE001-P00033-M07, FSE001-P00033-F20, FSE001-P00032-L18, FSE001-P00026-O15, FSE001-P00021-I19, FSE001-P00018-O16, FSE001-P00015-F16, FSE001-P00011-L20, FSE001-P00008-F01, FSE001-P00008-E11, FSE001-P00008-B10, FSE001-P00007-I23, FSE001-P00006-N15, FSE001-P00005-J21, FSE001-P00003-G02, FSE001-P00001-J14, FSA001-P00030-I12, FSA001-P00004-B20 SJCHGC06417 FSE001-P00025-G23, FSE001-P00018-H11 - FSA001-P00031-C05, FSA001-P00017-M14, FSA001-P00013-F20 SJCHGC05833 - -
oligosaccharyl transferase	Sjp_0046190	
triacylglycerol lipase	Sjp_0060000, Sjp_0006640	
Diacylglycerol O-acyltransferase	Sjp_0108620	
beta-galactosidase	Sjp_0072010	
1-acyl-sn-glycerol-3-phosphate acyltransferase	Sjp_0113900	
phosphatidate phosphatase	Sjp_0028050	
diacylglycerol kinase	Sjp_0102700, Sjp_0125070, Sjp_0044900, Sjp_0041190, Sjp_0006700	

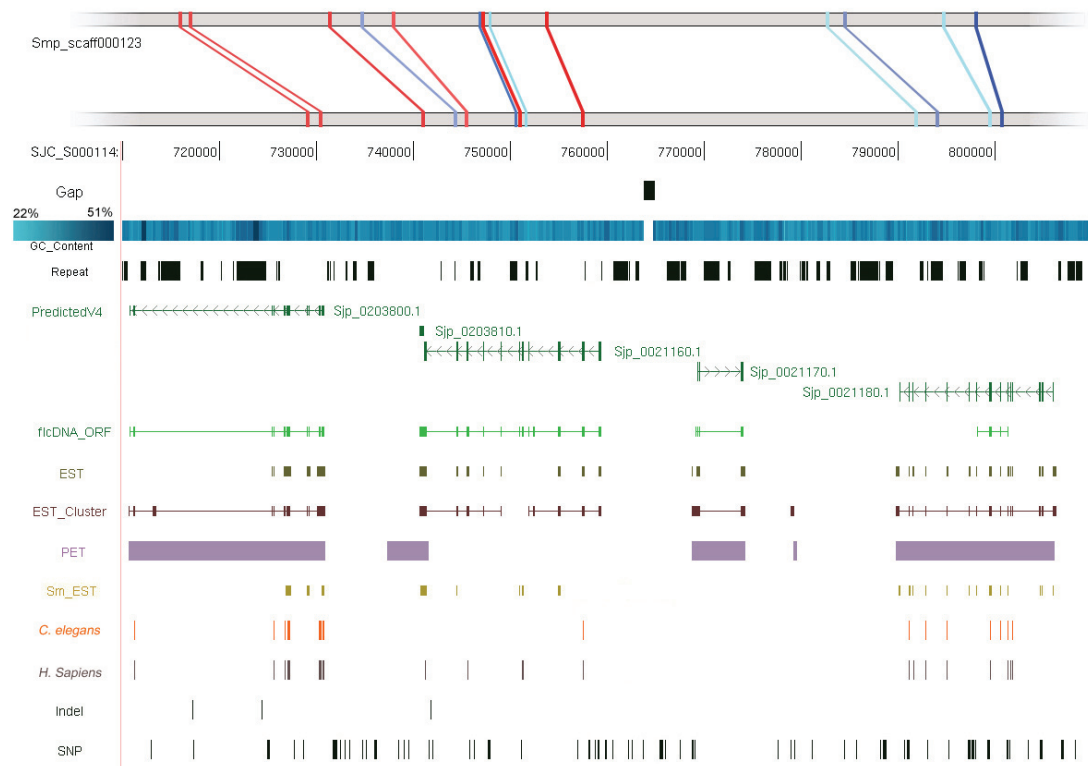
Glycophospholipid metabolism

glycerol-3-phosphate dehydrogenase	Sjp_0117170, Sjp_0104060, Sjp_0003530, Sjp_0017810	FSA001-P00025-M24, FSE001-P00001-L12
Dynein heavy chain at 36C	Sjp_0108580	-
SH3-domain GRB2-like 1	Sjp_0035450	FSE001-P00031-E08, SJCHGC00605, SJCHGC01831
Phospholipase D	Sjp_0079820, Sjp_0075670, Sjp_0081710	-
CDP-diacylglycerol—inositol-3-phosphatidyltransferase	Sjp_0062420	-
phosphatidylserine synthase	Sjp_0027930, Sjp_0027940, Sjp_0022460, Sjp_0061470	FSE_142_22_F03, SJCHGC04840, SJCHGC04840,
phosphatidate phosphatase	Sjp_0028050	SJCHGC05833
diacylglycerol kinase	Sjp_0102700, Sjp_0125070, Sjp_0044900, Sjp_0041190, Sjp_0006700	-
Phosphatidylserine decarboxylase	Sjp_0005350	FSE001-P00034-A09, FSE001-P00032-P10, FSE001-P00027-J12, FSE001-P00012-M17, FSE001-P00004-E03, FSA001-P00011-E07
phosphatidylserine synthase 2	Sjp_0022450	SJCHGC04840, SJCHGC04840
lecithin-cholesterol acyltransferase	Sjp_0079120	SJCHGC01661
conductance calcium-activated potassium channel	Sjp_0101750	-
choline phosphotransferase 1	Sjp_0054500	SJCHGC00742, FSA001-P00003-I11
lysophospholipase II	Sjp_0028510	FSE001-P00017-B01, FSE001-P00017-B01, SJCHGC02460, FSE001-P00026-L01, FSE001-P00025-G19, FSE001-P00015-O14, FSA001-P00030-B04, FSA001-P00009-A20
choline-phosphate cytidylyltransferase	Sjp_0016680	-
ethanolamine-phosphate cytidylyltransferase	Sjp_0030910	SJCHGC06128
lysophospholipase A2	Sjp_0082230, Sjp_0079060, Sjp_0029100, Sjp_0029090	-
ethanolamine-phosphotransferase	Sjp_0052680	-
choline kinase beta	Sjp_0090310	-

choline acetyltransferase	Sjp_0035650	-	FSA_072_84_D11, FSE_163_55_G07
Acetylcholinesterase	Sjp_0036280, Sjp_0045440		FSA001-P00031-C05, FSA001-P00017-M14, FSA001-P00013-F20
1-acyl-sn-glycerol-3-phosphate acyltransferase	Sjp_0113900		
Lipid synthesis and/or catabolism			
choline phosphotransferase	Sjp_0054500		FSA001-P00003-I11, SJCHGC00742
1-alkyl-2-acetylglycero-phosphocholine esterase	Sjp_0084940, Sjp_0084580	-	
conductance calcium-activated potassium channel	Sjp_0101750	-	
1-acyl-sn-glycerol-3-phosphate acyltransferase	Sjp_0113900		FSA001-P00031-C05, FSA001-P00017-M14, FSA001-P00013-F20
phosphatidate phosphatase	Sjp_0028050		SJCHGC05833
ethanolamine-phosphotransferase	Sjp_0052680	-	
ectonucleotide pyrophosphatase/phosphodiesterase 6	Sjp_0027190		FSE001-P00029-A02, SJCHGC02021
Phospholipase D	Sjp_0079820, Sjp_0075670, Sjp_0081710	-	
Sphingolipid synthesis and degradation			
serine palmitoyltransferase	Sjp_0035000, Sjp_0072130, Sjp_0029360		FSE_046_45_E06, FSA001-P00033-M17
3-dehydrosphinganine reductase	Sjp_0074380	-	FSE_207_25_A04, FSE_207_25_A04, FSE_176_89_A12, FSE_176_89_A12, FSE001-P00005-J16
sphingosine kinase	Sjp_0102700		
sphingosine-1-phosphate lyase	Sjp_0056830	-	FSA001-P00032-B03, FSA001-P00032-B03
Ceramidase	Sjp_0114380		SJCHGC01813, FSE001-P00013-P08
beta-1,4-galactosyltransferase	Sjp_0079160, Sjp_0079160	-	
beta-galactosidase	Sjp_0072010		
fatty acid desaturase	Sjp_0061450		FSA001-P00006-E05
ceramide glucosyltransferase	Sjp_0093880, Sjp_0054080	-	
alpha-galactosidase	Sjp_0023250, Sjp_0023220, Sjp_0107620, Sjp_0023240		SJCHGC05460
ceramide cholinephosphotransferase	Sjp_0099270, Sjp_0099280	-	

sphingomyelin phosphodiesterase ceramide kinase phosphatidate phosphatase	Sjp_0064150	SJCHGC06222
	Sjp_0023990	-
	Sjp_0028050	SJCHGC05833
Arachidonic acid synthesis		
carbonyl reductase	Sjp_0128240, Sjp_0075840	FSE001-P00030-J17, FSE001-P00027-L04, FSE001-P00027-B01, FSE001-P00024-C09, FSE001-P00013-F03, FSE001-P00012-N23, FSE001-P00012-E21, FSE001-P00011-C17, FSE001-P00009-B11, FSE001-P00006-I14, FSE001-P00004-D20, FSE001-P00002-N23, FSA001-P00028-N06, FSA001-P00006-G15, FSE001-P00024-I18, FSE001-P00024-I18 FSA001-P00009-E09
	Sjp_0022850	-
	Sjp_0004210	FSE001-P00022-D06, FSE001-P00022-D06, FSE001-P00021-N20, FSE001-P00009-P22, FSE001-P00009-P22, FSA001-P00004-I22, FSA001-P00004-I22
leukotriene-A4 hydrolase gamma-glutamyltranspeptidase	Sjp_0110610	-
	Sjp_0115980	-
	Sjp_0101750	-
arachidonate 15-lipoxygenase conductance calcium-activated potassium channel arachidonate 12-lipoxygenase (R-type)	Smp_156150, Smp_139410, Smp_161450, Smp_161450, Smp_133290	SJCHGC09449
	Smp_034020	-
	Sjp_0022850, Smp_007550	-
Prostaglandin synthesis		
phospholipase A2 arachidonate 5-lipoxygenase leukotriene-A4 hydrolase		

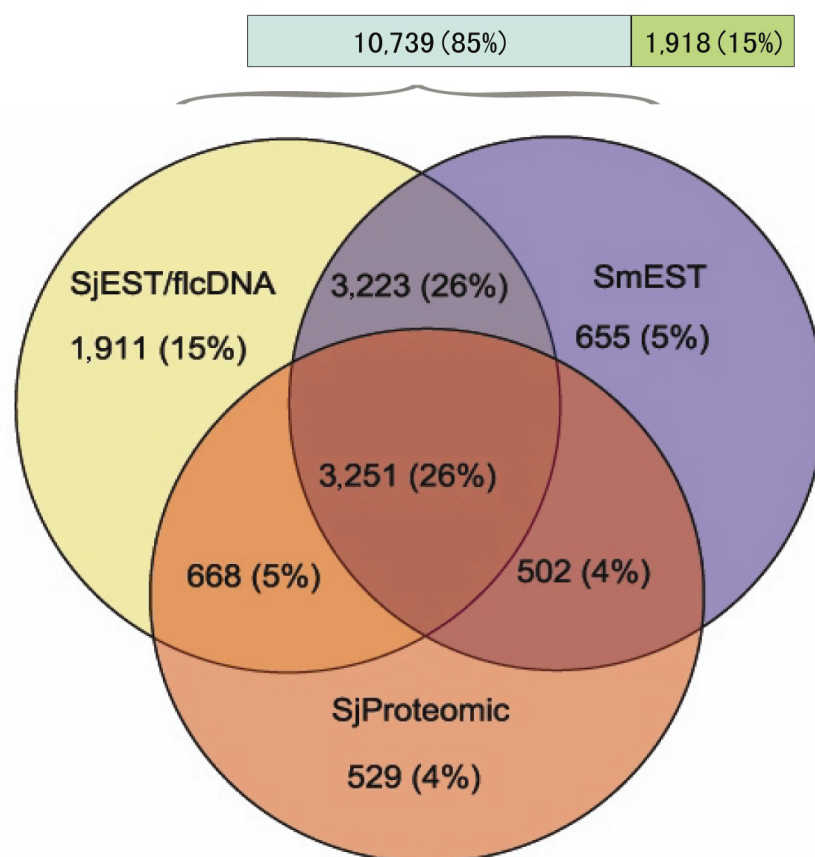
Supplementary Figures



Supplementary Figure 1. Schematic representation of a portion of the *S. japonicum* genome illustrated with various feature tracks

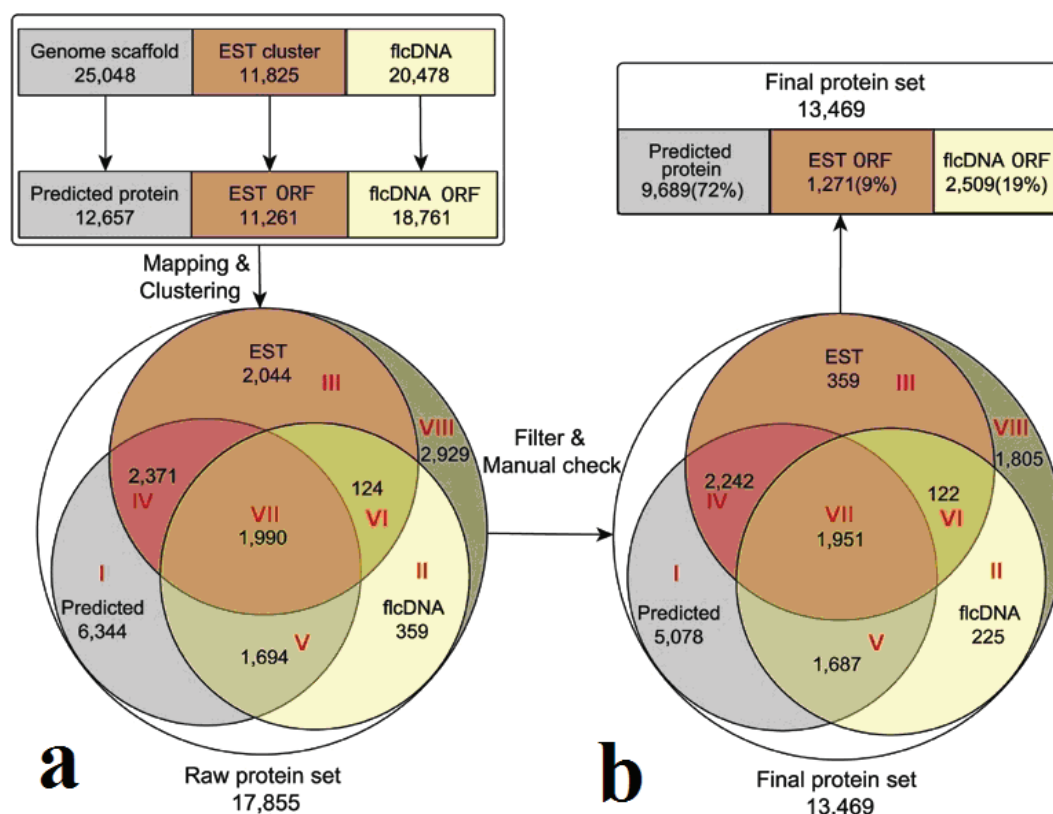
The feature tracks are described from top to bottom. Smp_scaff000123 is the ID of an assembled genomic scaffold sequence of *S. mansoni* and SJC_S000114 is the ID of a scaffold of *S. japonicum* genomic sequence assembly. The corresponding relationship between these two scaffolds is indicated by connecting lines (BLAST E-value $< 1e-80$, red; $< 1e-50$, blue; and $< 1e-20$, cyan), which represent the coding sequences of the two counterpart regions. The black bar in the 'Gap' track indicates an *S. japonicum* scaffold gap in this particular position. In the 'GC_Content' track, the darker color stands for the greater GC percentage, ranging from 22% to 51%. The details for repeats are described in the text. The PredictedV4 track presents the fourth version of *S. japonicum* predicted genes within the scaffold labelled as Sjp_xxxxxxx.1. Alignments of the putative *S. japonicum* CDSs derived from flcDNAs ('flcDNA_CDS'), ESTs ('EST') and EST

clusters ('EST_Cluster' track). The *S. japonicum* PET track theoretically marks the beginning and end position of a transcript. The annotation supporting tracks include the alignments with the *S. mansoni* ESTs (Sm_EST) and the 46th GenBank/EMBL/DDBJ protein dataset release for *C. elegans* and *H. sapiens*. The information for *S. japonicum* indels and SNPs is shown as the last two tracks at the bottom of the figure.



Supplementary Figure 2. Support for predicted genes by expression datasets

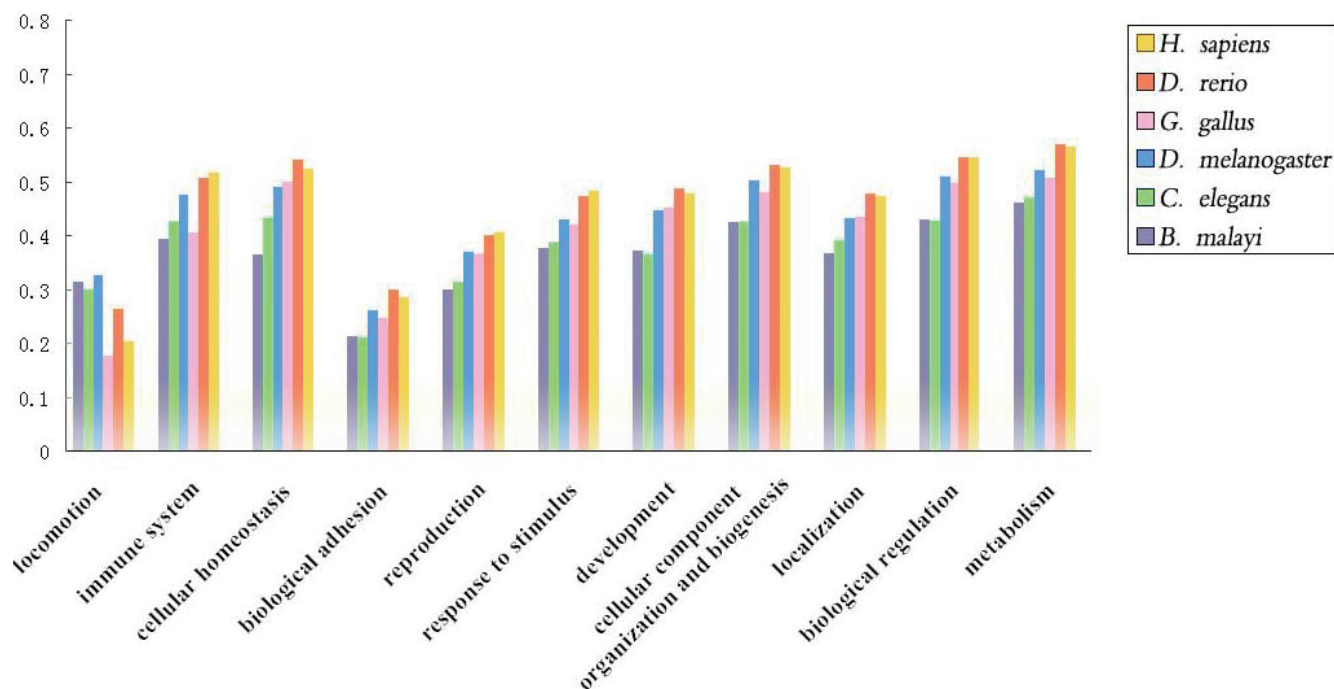
A total of 12,657 *S. japonicum* protein-coding genes were predicted using the gene prediction program, ExonHunter. ExonHunter combines *ab initio* gene prediction with supporting evidence from *S. japonicum* and *S. mansoni* ESTs, the Swiss-Prot protein database, *S. japonicum* PETs, and Pfam. Shown is a Venn diagram of the 10,739 (85%) of the 12,657 predicted genes supported by *S. japonicum* EST and flcDNA (yellow pie), *S. mansoni* EST (blue pie) and *S. japonicum* proteomics (red pie) data. Among the 12,657 protein-coding genes, 5,727 (45.2%) were supported by *S. japonicum* flcDNAs, 8,318 (66.7%) were supported by *S. japonicum* ESTs; in total, 9,053 (71.5%) are supported by flcDNAs or ESTs. In addition, 7,631 (60.2%) were supported by *S. mansoni* ESTs, and 4,950 (39.1%) were supported by *S. japonicum* proteomics data.



Supplementary Figure 3. Integration of *S. japonicum* protein dataset

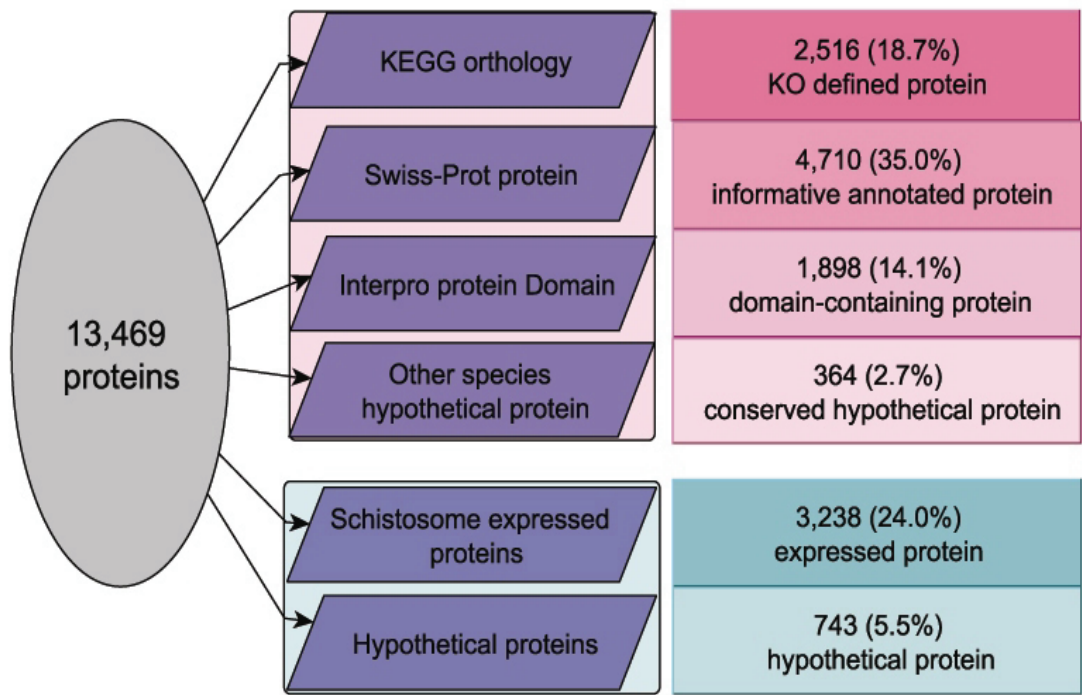
This figure shows the overall view of various *S. japonicum* gene datasets and the process leading to the integrated *S. japonicum* protein dataset, used for functional analysis in this study. Three protein-coding gene datasets, 12,657 predicted genes (Predicted CDSs), 11,261 putative EST ORFs (EST CDSs) and 18,761 putative flcDNA ORFs (flcDNA CDSs) mapped onto the *S. japonicum* genome were grouped according to their positions and overlaps on the draft genome. The mapping and grouping results are showed in Venn diagrams. Venn diagram **a**, 1,990 CDS groups were shared by predicted, flcDNA and EST (VII part), another 2,371 groups were shared by predicted and EST CDSs (IV part), 1,694 groups were shared by predicted and flcDNA CDSs (V part), and 124 groups were shared by flcDNA and EST CDSs (VI part). Parts I, II and III include CDSs that have only one source of data. Venn diagram **b** shows the status of the groups described above after auto filtering and manual checking for coding region conflicts, gene fragmentation and mergence problems. For example, a large proportion of putative CDSs that exclusively came from the EST dataset were discarded (part III changed from 2,044 in Venn diagrams **a** to 359 in Venn diagrams **b**) because of fragmentation problems of EST derived data. The most reasonable CDS, in most cases

the longest one if there are no conflicts, was chosen to represent each group. The cut-off of alignment of EST CDSs and flcDNA CDSs to the genome scaffolds was set at the level of 70% overall identity, thus eliminating 2,929 potential EST/flcDNA CDSs that were not covered by the genome draft (a strict threshold as compared to that used for the genome coverage calculation was used to bolster the reliability). These non-covered CDSs were clustered via sequence similarity (VIII part). Note that in the final 13,469 integrated CDS dataset, 1,805 EST/flcDNA CDSs (part VIII, Venn diagram **b**) that did not have good alignments on the genome draft were included for better coverage and more reliable protein sequence. Some predicted genes (part I, Venn diagram **a**) seem to be fragments of these CDSs, however, most predicted gene structures were preserved in the final gene set.



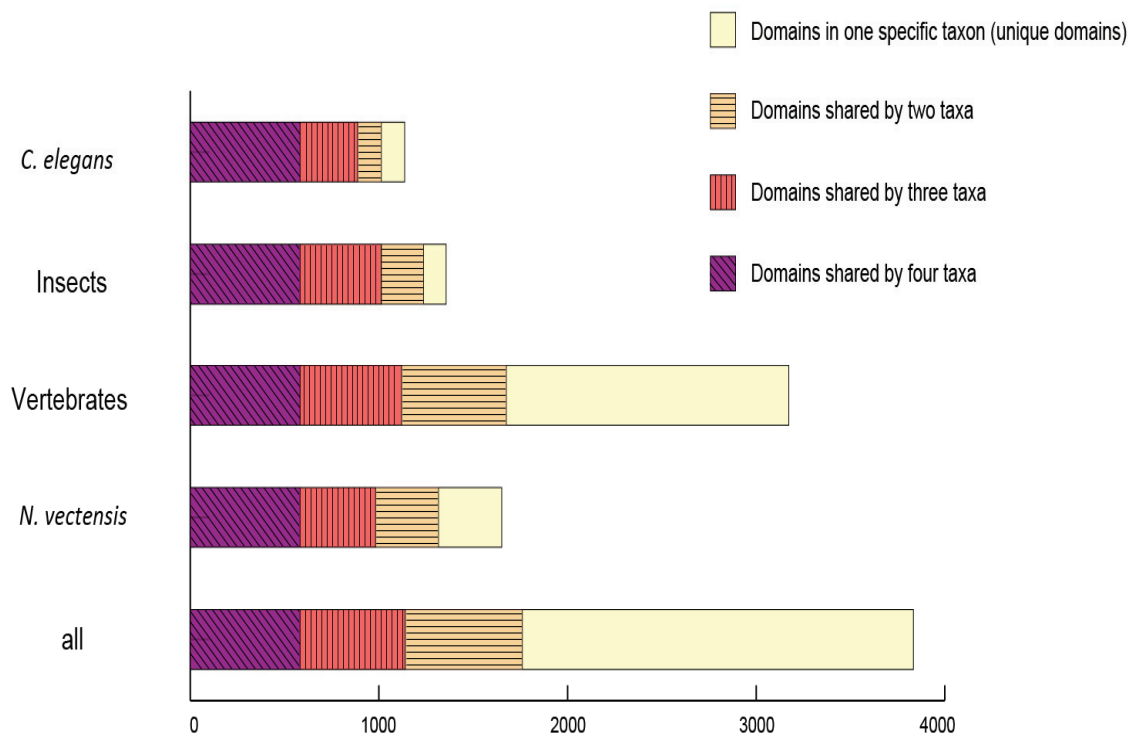
Supplementary Figure 4. Functional categorization of *S. japonicum* genes and comparative analysis with other representative organisms

Subsets of the *S. japonicum* proteins that fall into individual functional categories are grouped and compared with their orthologs in *H. sapiens*, *D. rerio*, *G. gallus*, *D. melanogaster*, *C. elegans*, and the filarial nematode, *B. malayi*. The percentage of *S. japonicum* proteins in a particular subset that has an ortholog relationship to the respective reference genome is shown. This reflects the measure of protein sequence conservation within this particular functional category between *S. japonicum* and each studied genome. Y axis, 0.1 = 10%.



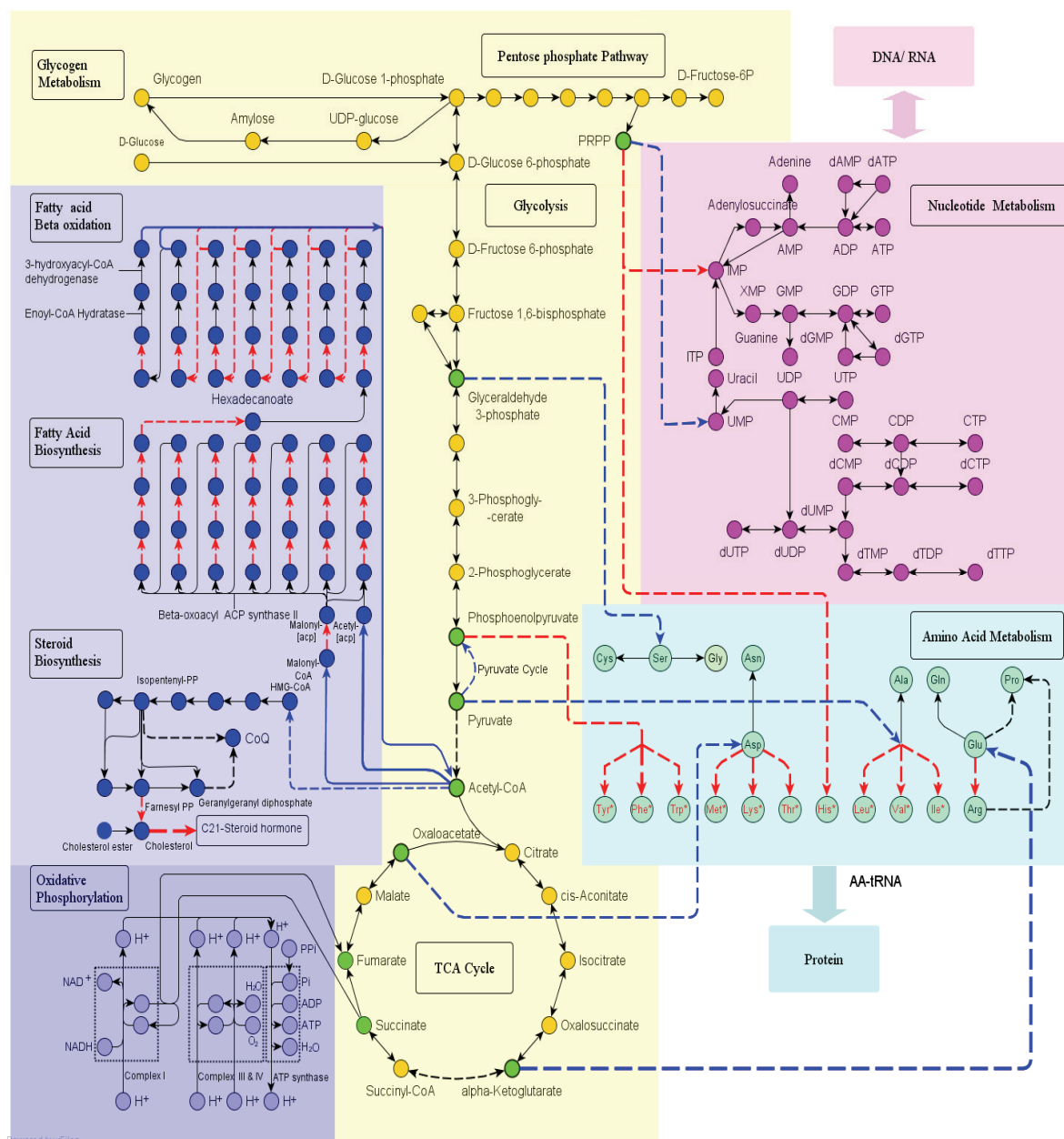
Supplementary Figure 5. Stepwise annotation of *S. japonicum* genes

A total of 13,469 proteins were classified into 6 groups in a stepwise manner, retrieving reliable and functional information at higher priority. Each gene is only classified into one group, with corresponding labels stamped to describe the reliability level of functional information that the annotation carries. For exemplar, there are 2,516 *S. japonicum* genes mapped to 1,546 KEGG orthology (refer to Supplementary Table 12), naming as ‘KO defined protein’.



Supplementary Figure 6. Protein domain analysis in metazoan

The distribution of 3,834 domains detected in vertebrates, insects, *C. elegans*, and *N. vectensis*, but not in *S. japonicum*. A total of 1,140 protein domains, including 582 domains shared by all other taxa, and 558 domains by three of the four taxa, were possibly lost by *S. japonicum*.



Supplementary Figure 7. An overview of basic metabolic pathways in *S. japonicum*

Major metabolic pathways, including glycolysis/gluconeogenesis, TCA cycle, glycogen metabolism, pentose phosphate shunt, fatty acid and cholesterol biosynthesis and metabolism, oxidative phosphorylation, nucleotide and amino acid metabolisms are illustrated with different colored backgrounds. Compounds in or between individual pathways are indicated with circles and linked by different types of lines. Black solid line, direct chemical reaction; red dashed line, chemical reaction cannot be completed due to apparent absence of appropriate gene (coding for enzymes);

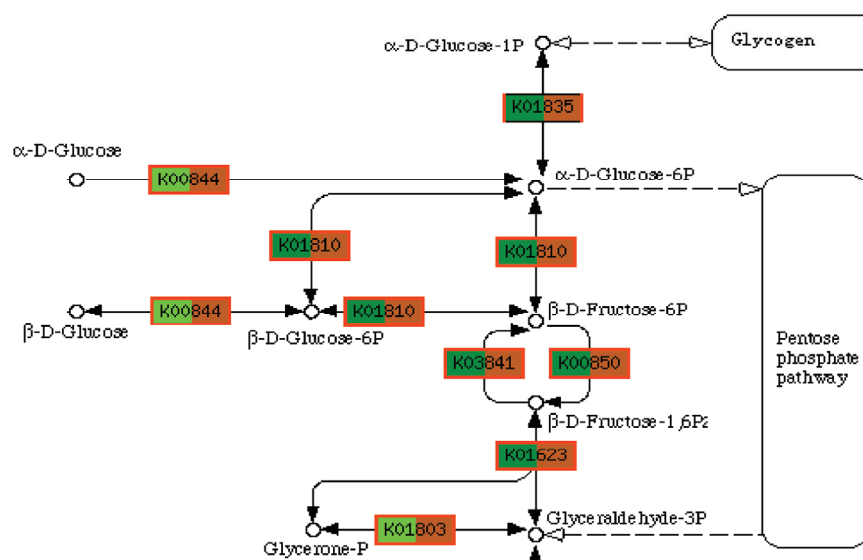
arrowhead, reversible or irreversible reaction; Blue solid lines; cross-talk between pathways; blue dashed line, pathway linkage with multi-step reaction. Green circle, shared intermediates between the carbohydrate metabolism and other pathways.

Supplementary Figure 8. Pathways of carbohydrate metabolism (include five sub-figures)

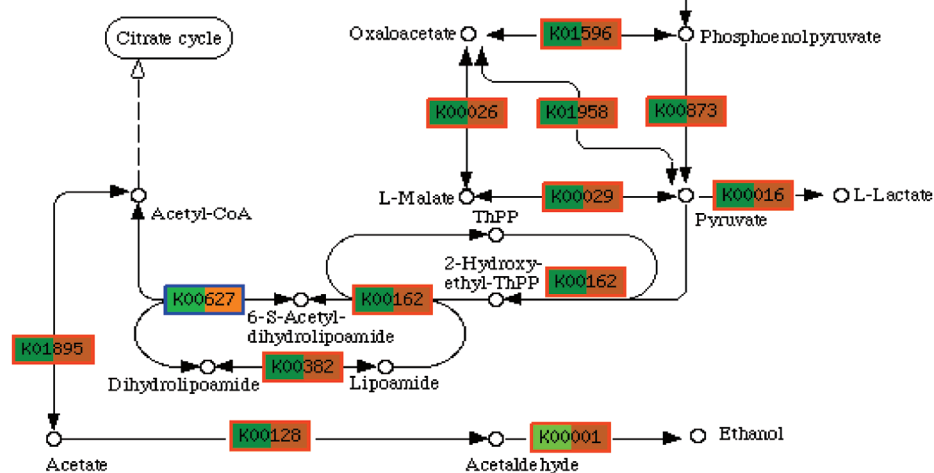
The boxes and the colors represent the enzymes annotated in the *S. japonicum* genome, based on BLAST searches of predicted gene sequences in the database (left, different level of green) or BLAST searches against the KEGG database of the *S. japonicum* ESTs or flcDNAs (right, different levels of pink or brown), while the circles show the named metabolic intermediates.

1e-5 1e-12 1e-25 1e-50 1e-100 Animals Plants Fungi Protists Bacteria Archaea

GLYCOLYSIS



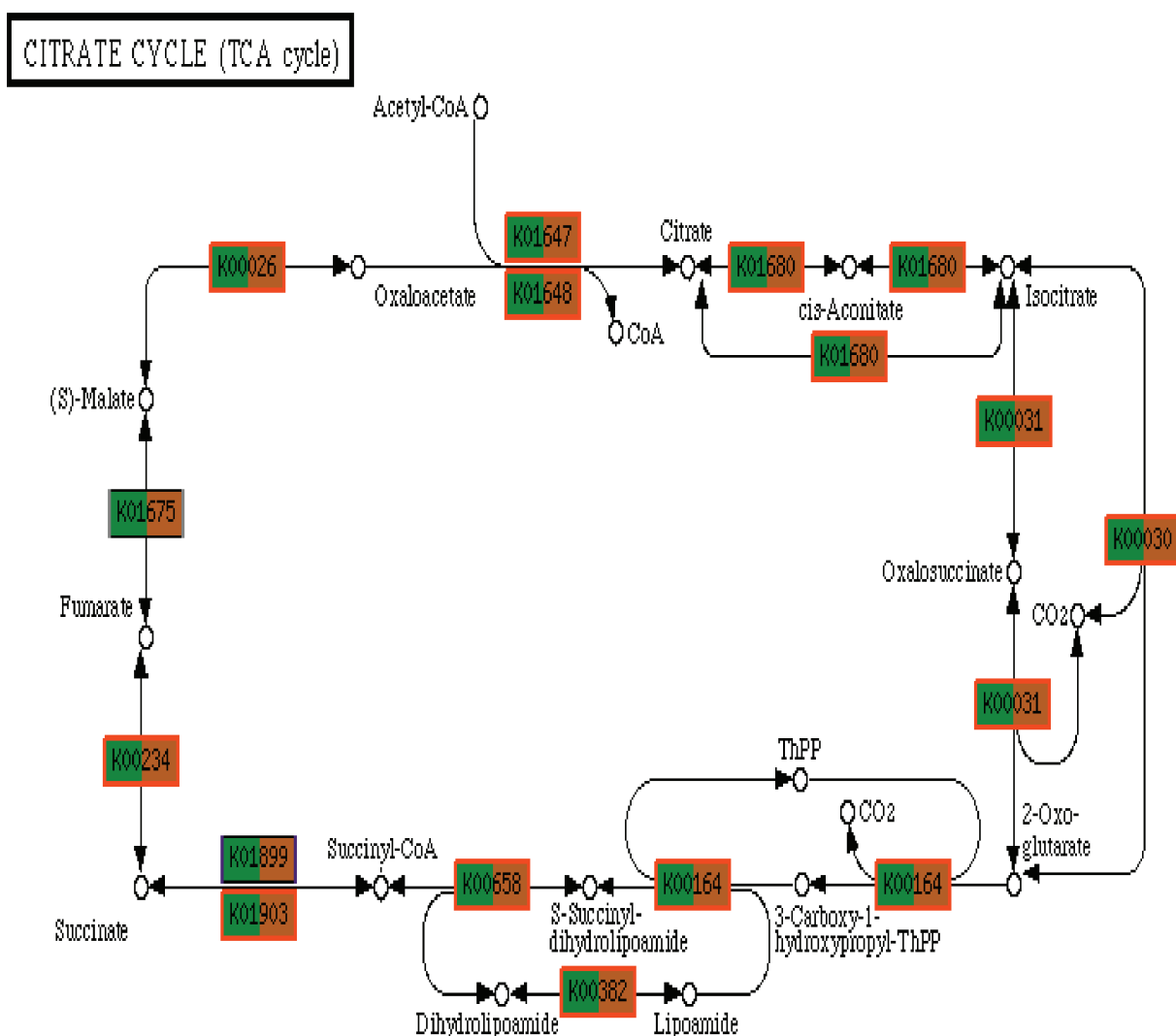
GLUCONEOGENESIS



Supplementary Figure 8.1. Glycolysis and gluconeogenesis

All the requisite enzymes for glycolysis and gluconeogenesis are annotated in *S. japonicum* indicating that glucose can be used as a carbon and energy source.

1e-5 1e-12 1e-25 1e-50 1e-100 ■ Animals ■ Plants ■ Fungi ■ Protists ■ Bacteria ■ Archaea

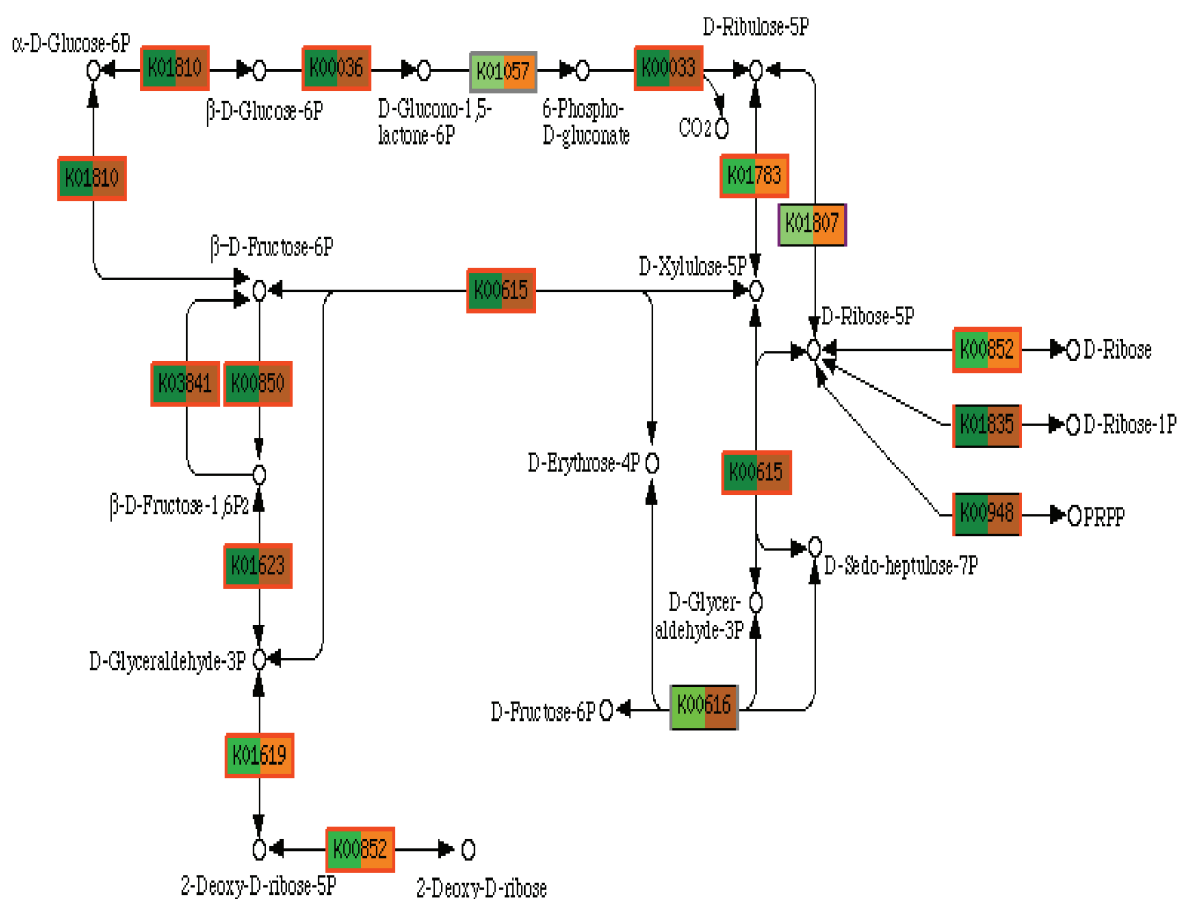


Supplementary Figure 8.2. Citrate cycle (TCA cycle)

A fully annotated citric acid cycle is present in *S. japonicum* which provides metabolic intermediates that connect with other pathways.



PENTOSE PHOSPHATE PATHWAY



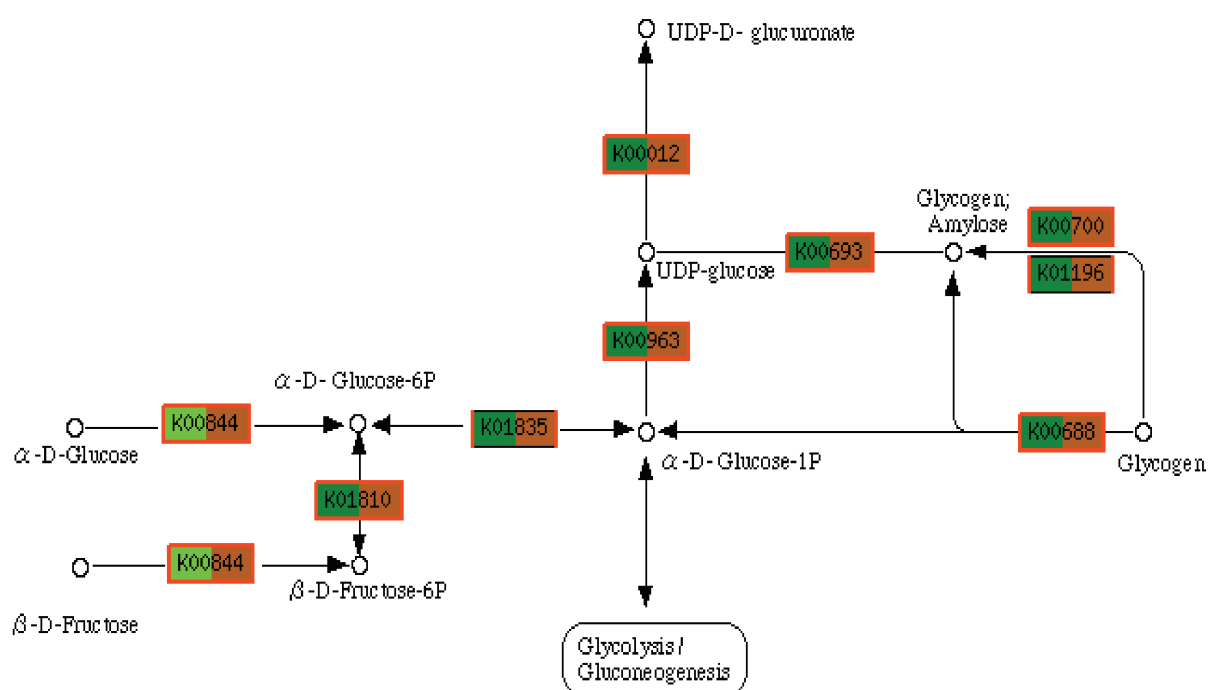
Supplementary Figure 8.3. Pentose phosphate pathway

The pentose phosphate pathway provides ribose and 5-deoxy-D-ribose which are essential for nucleotide biosynthesis.

1e-5 1e-12 1e-25 1e-50 1e-100

Animals Plants Fungi Protists Bacteria Archaea

Glycogen Metabolism



Supplementary Figure 8.4. Glycogen metabolism

S. japonicum is able to both synthesize glycogen from glucose as an energy store and utilize glycogen to produce glucose.



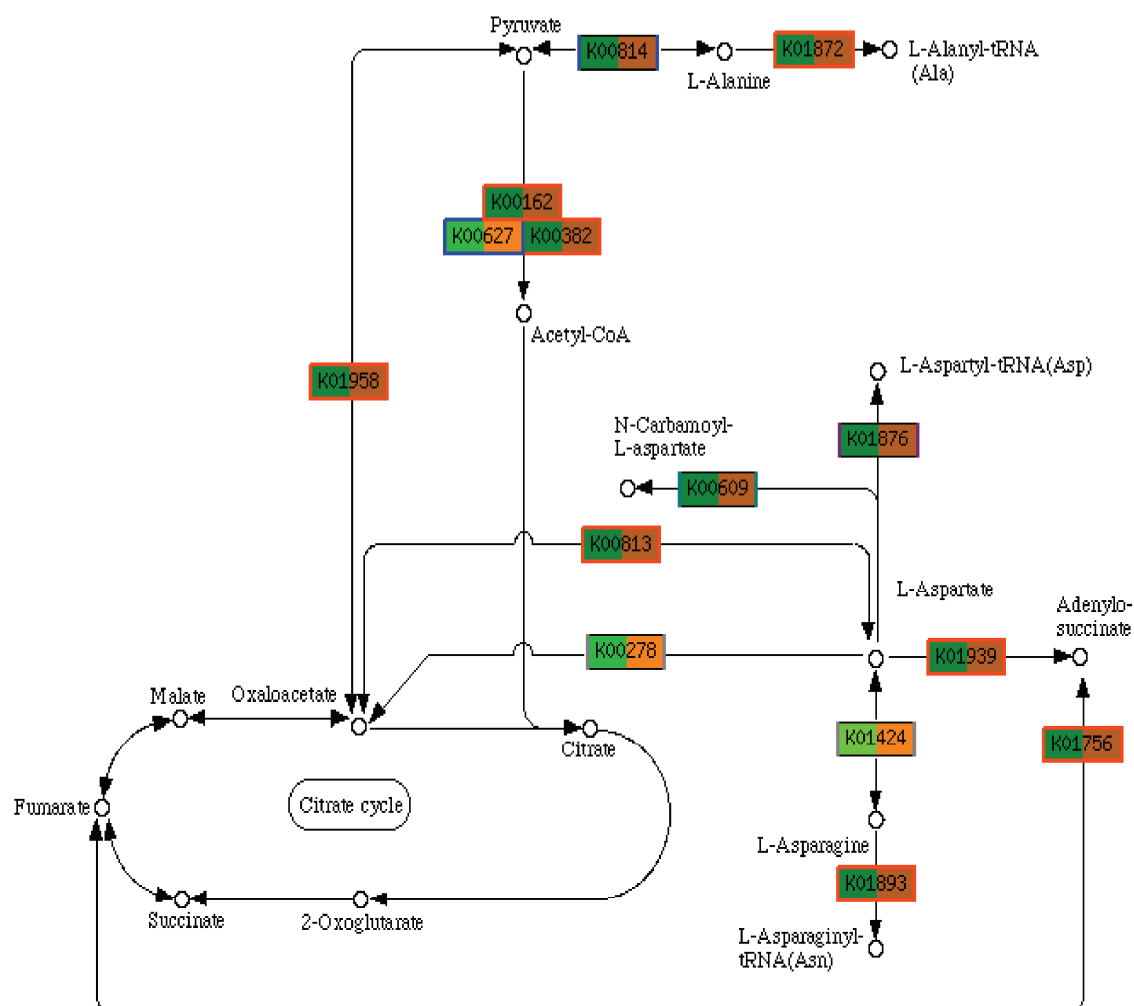
S. japonicum has a full component of annotated enzymes for the generation of ATP by oxidative phosphorylation.

Supplementary Figure 9. Amino acid metabolism (include 10 sub-figures)

The annotated pathways show that for *S. japonicum*, alanine, aspartate, cysteine, glutamate, glycine, serine, asparagine and glutamine can be generated from metabolic intermediates. In contrast, as key enzymes are missing, threonine, valine, leucine, isoleucine, methionine, phenylalanine, tyrosine, tryptophane, histidine, lysine and arginine cannot be generated or synthesized via intermediates of other metabolic pathways such as the citric acid cycle. All enzymes for aminoacyl tRNA synthesis are present, and these contribute to overall protein synthesis in *S. japonicum*.

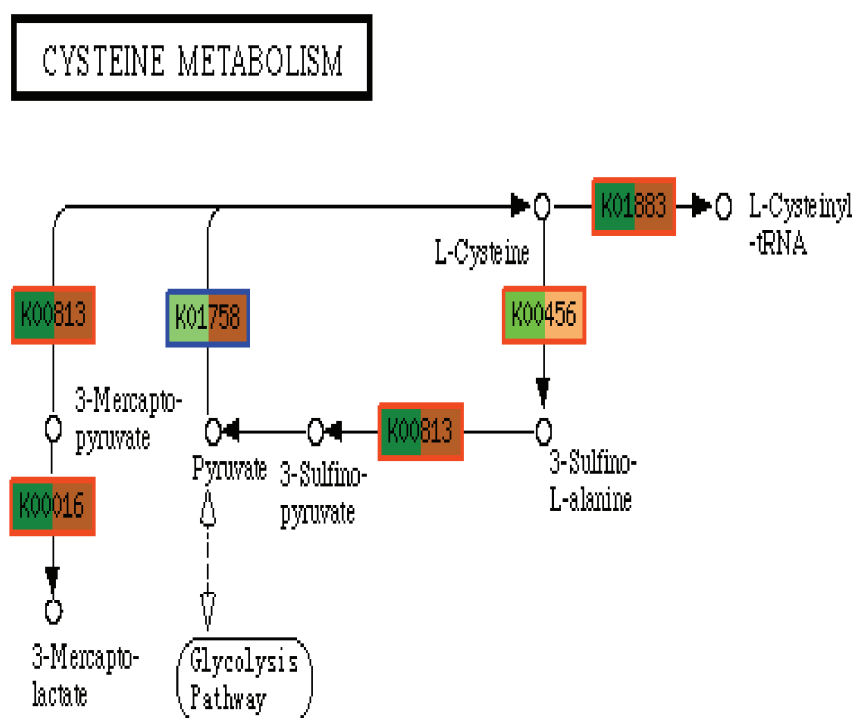
1e-5 1e-12 1e-25 1e-50 1e-100 Animals Plants Fungi Protists Bacteria Archaea

ALANINE AND ASPARTATE METABOLISM



Supplementary Figure 9.1. Alanine and aspartate/asparagine metabolism

In *S. japonicum*, alanine can be generated from pyruvate through the action of an amino-transferase enzyme; similarly, aspartate can be generated from oxaloacetate by transamination, and is interconvertible with asparagine.

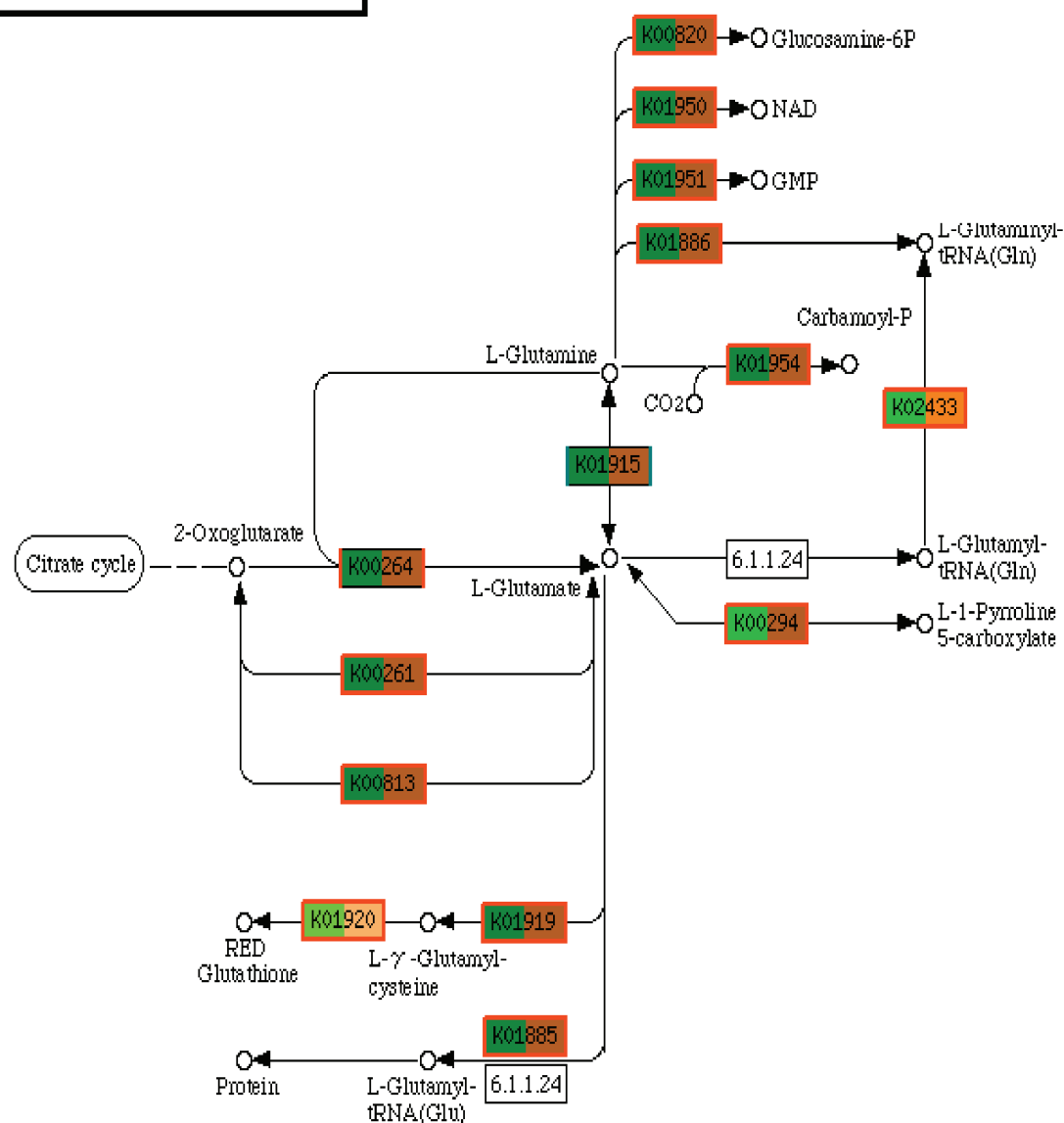


Supplementary Figure 9.2. Cysteine metabolism

L-cysteine can be generated from pyruvate, catalyzed by the enzyme cystathionine gamma-lyase (K01758).

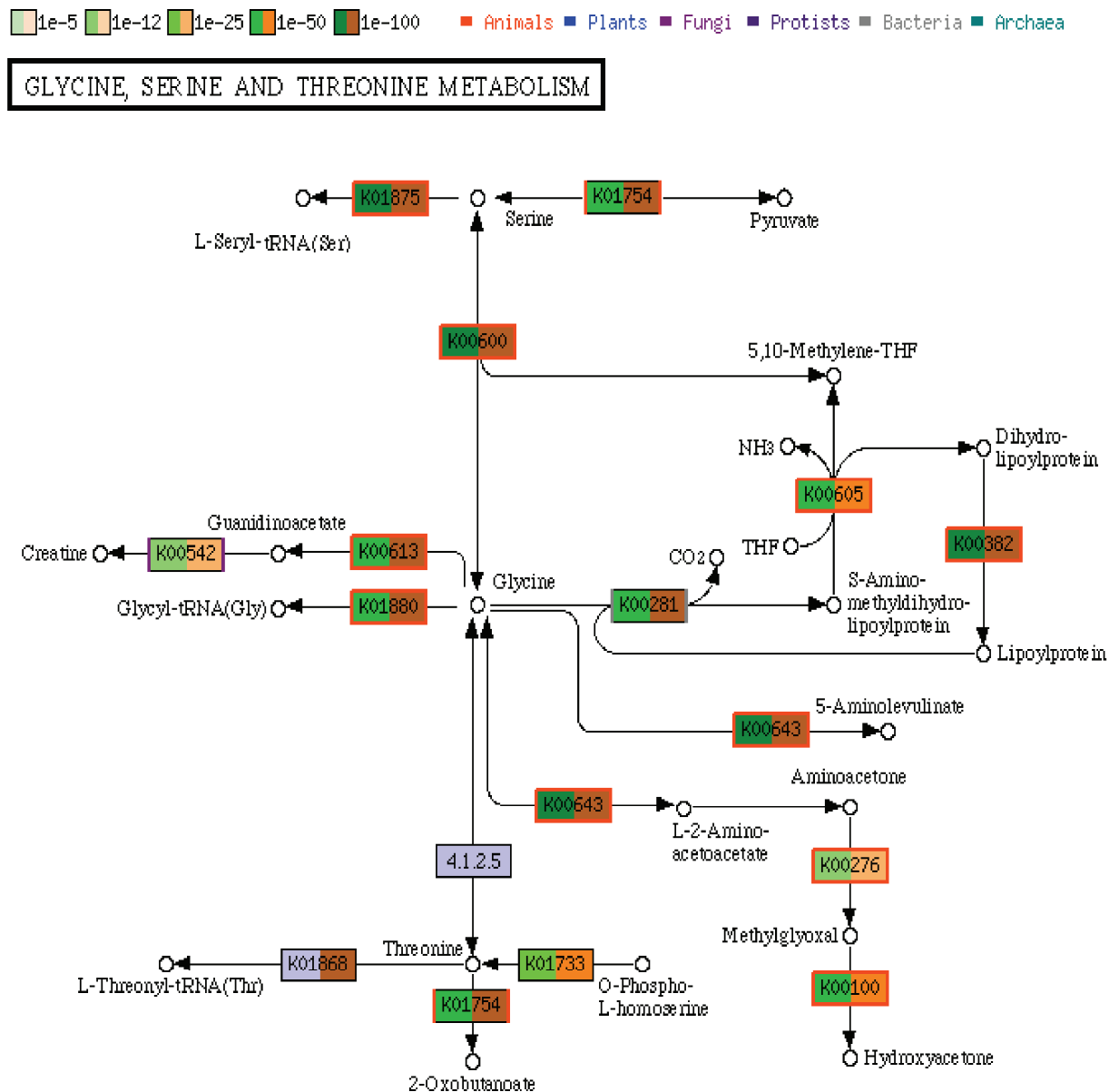
1e-5 1e-12 1e-25 1e-50 1e-100 Animals Plants Fungi Protists Bacteria Archaea

GLUTAMATE METABOLISM



Supplementary Figure 9.3. Glutamate and glutamine metabolism

Glutamic acid can be generated from the citric acid cycle, and is interconvertible with glutamine.

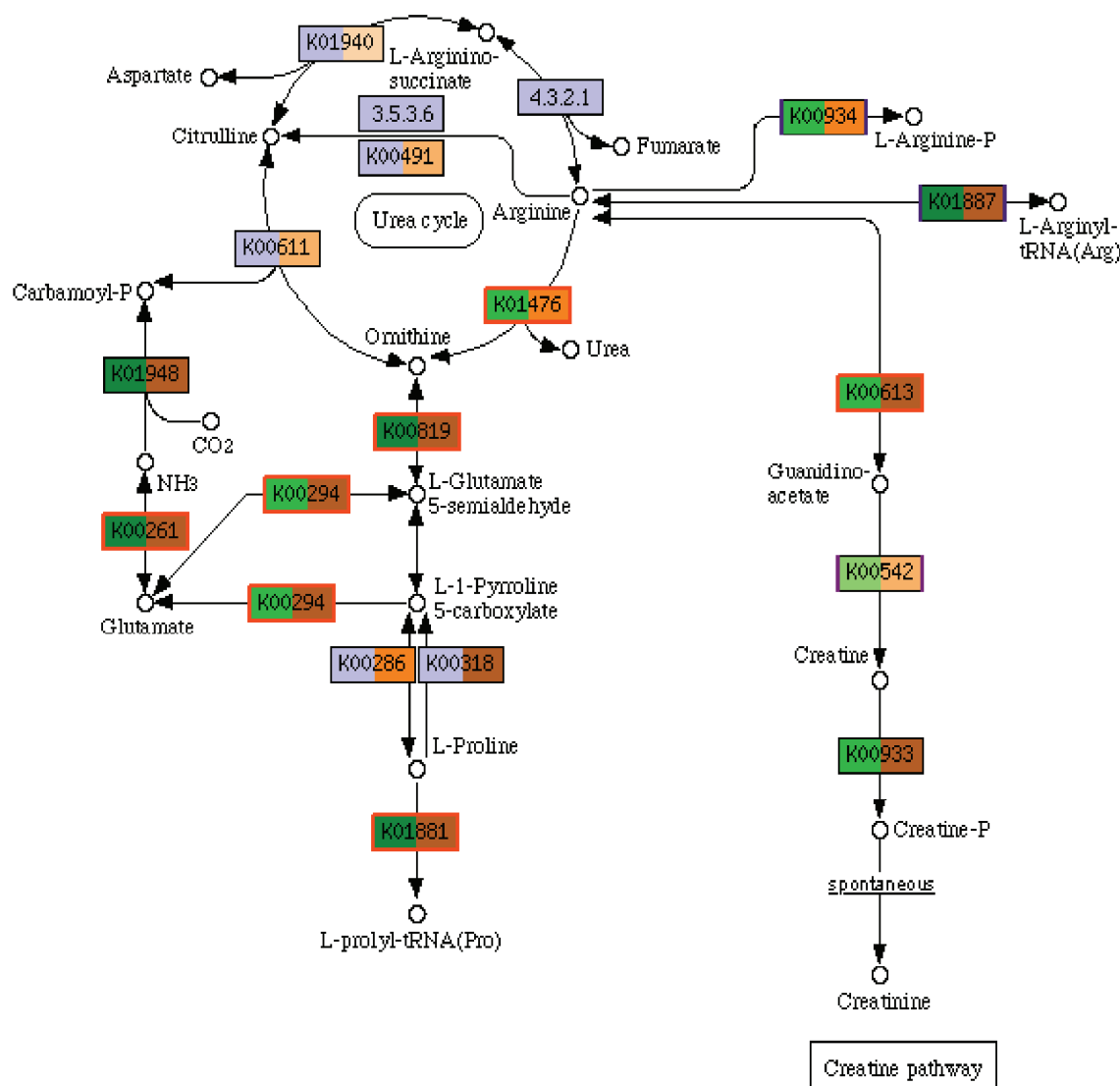


Supplementary Figure 9.4. Glycine, serine and threonine metabolism

Serine is interchangeable with pyruvate via the enzyme threonine dehydratase (K01754), and glycine can be generated from serine through serine hydroxymethyltransferase (K00600). However, threonine cannot be produced due to the absence of threonine aldolase (EC: 4.1.2.5).

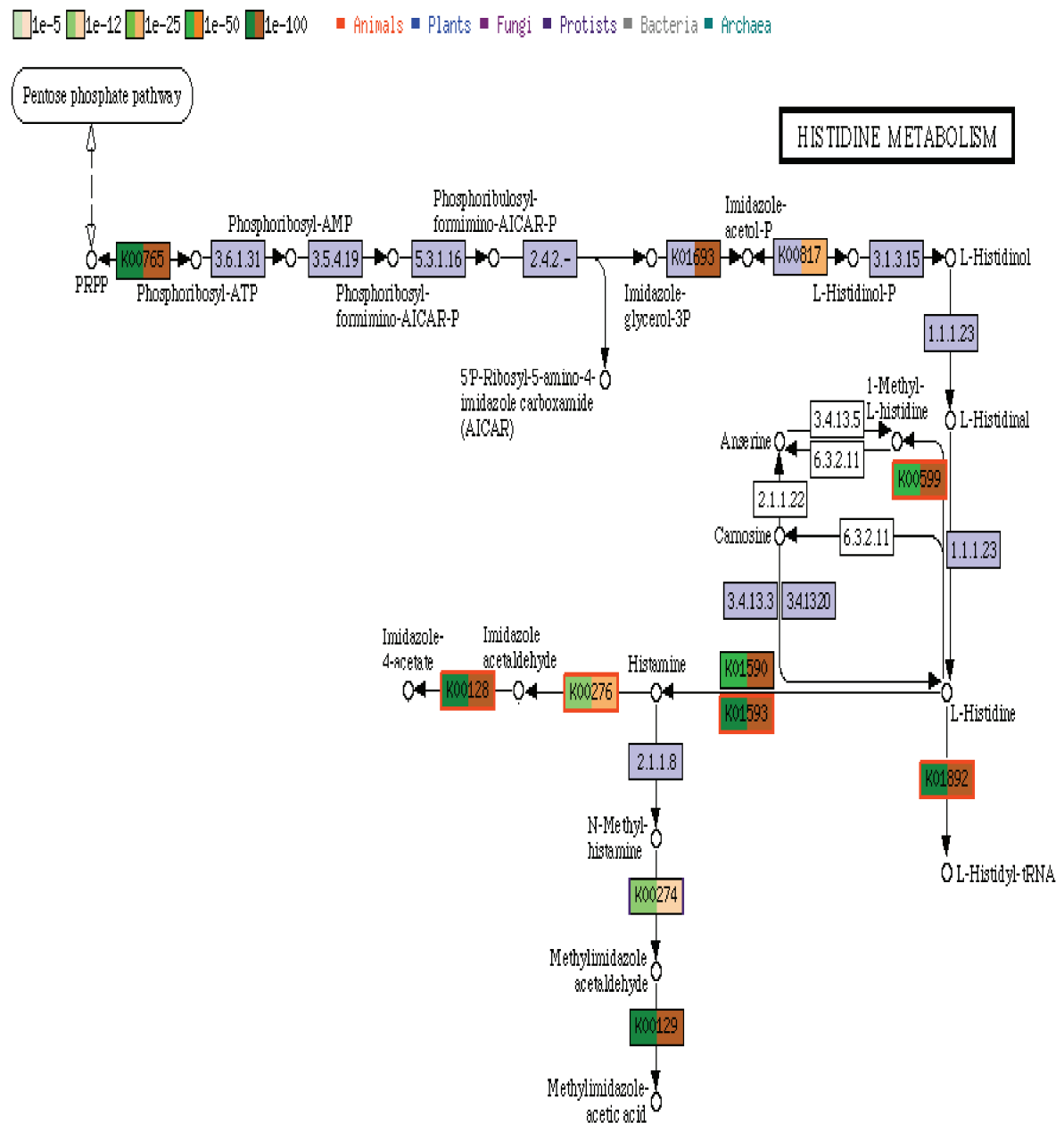
1e-5 1e-12 1e-25 1e-50 1e-100 ■ Animals ■ Plants ■ Fungi ■ Protists ■ Bacteria ■ Archaea

ARGININE AND PROLINE METABOLISM



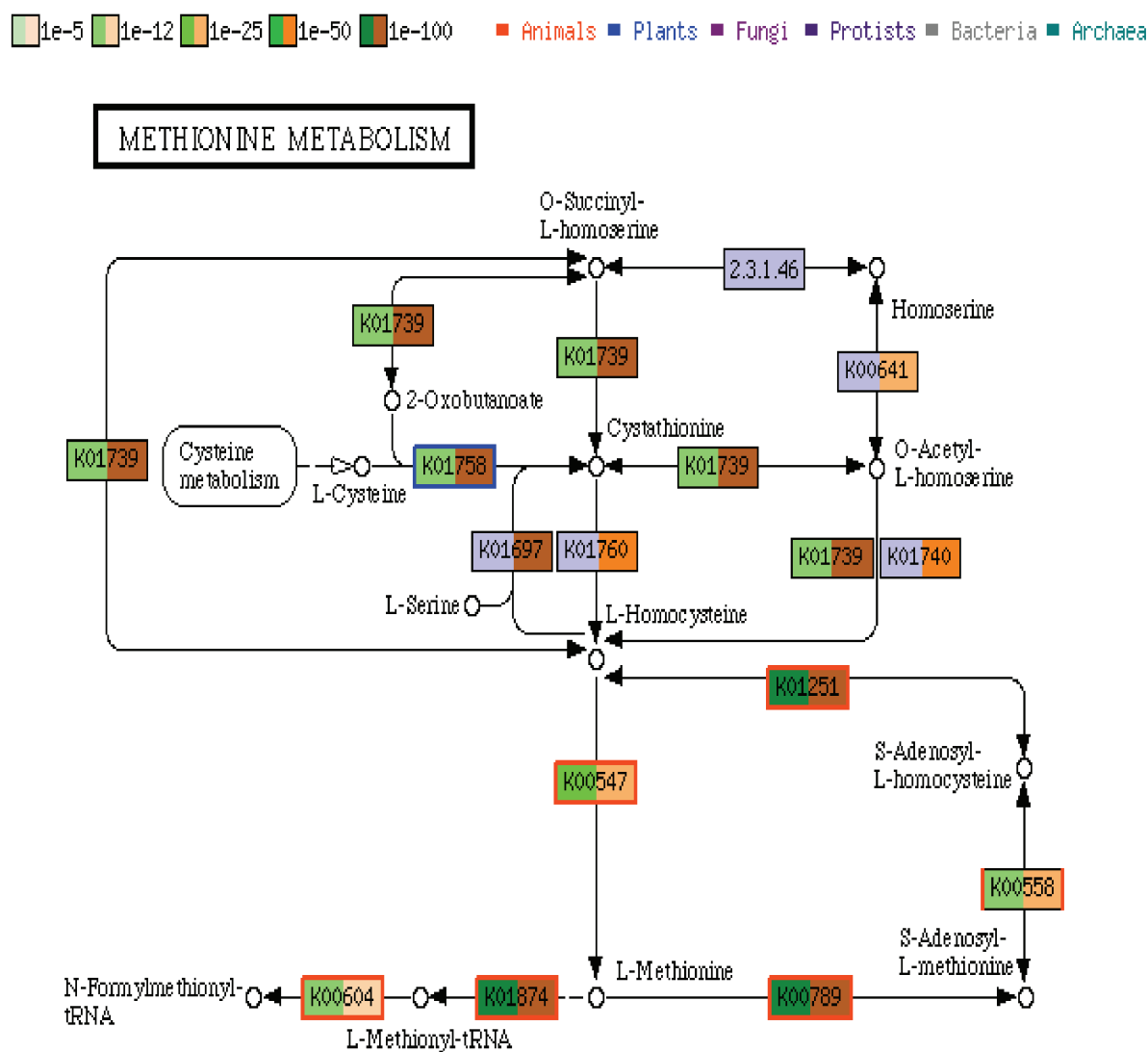
Supplementary Figure 9.5. Arginine and proline metabolism

The urea cycle appears incomplete, indicating *S. japonicum* does not excrete nitrogen through the urea pathway. Moreover, arginine could not be generated from this cycle. The key enzyme for proline generation from glutamic acid (E1.5.1.2) also appears to be missing.



Supplementary Figure 9.6. Histidine metabolism

Histidine cannot be synthesized due to the absence of the enzyme for histidine *de novo* biosynthesis from the pentose, 5-phospho-D-ribosyl-1-pyrophosphate (PRPP).



Supplementary Figure 9.8. Methionine metabolism

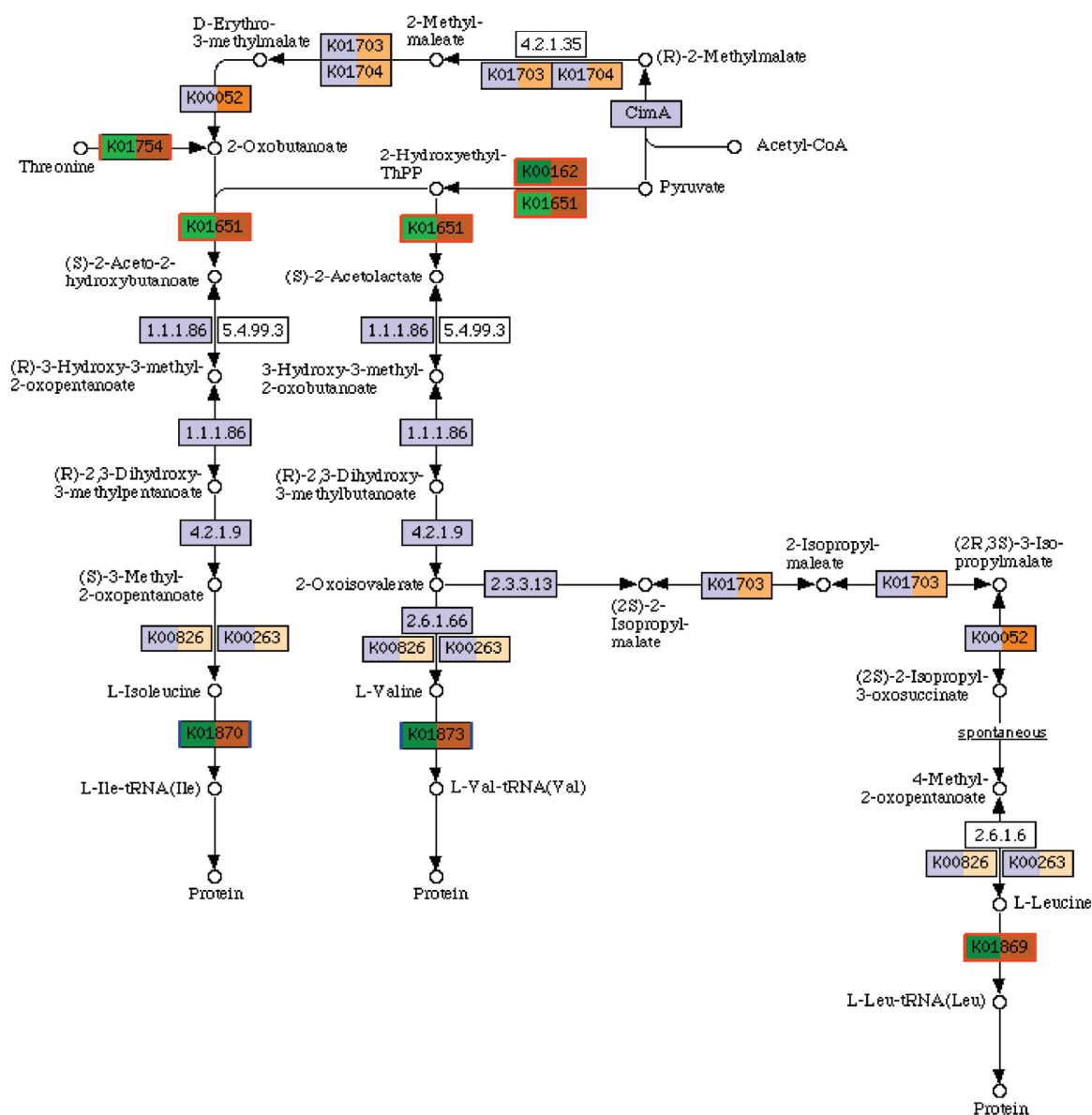
Methionine cannot be synthesized *de novo* by *S. japonicum* because its precursor L-homocysteine is not generated. The worm can use the S-adenosyl-L-methionine cycle when methionine or L-homocysteine is available.



None of the key enzymes required for biosynthesis of phenylalanine, tyrosine or tryptophan were identified by genome annotation.

1e-5 1e-12 1e-25 1e-50 1e-100 Animals Plants Fungi Protists Bacteria Archaea

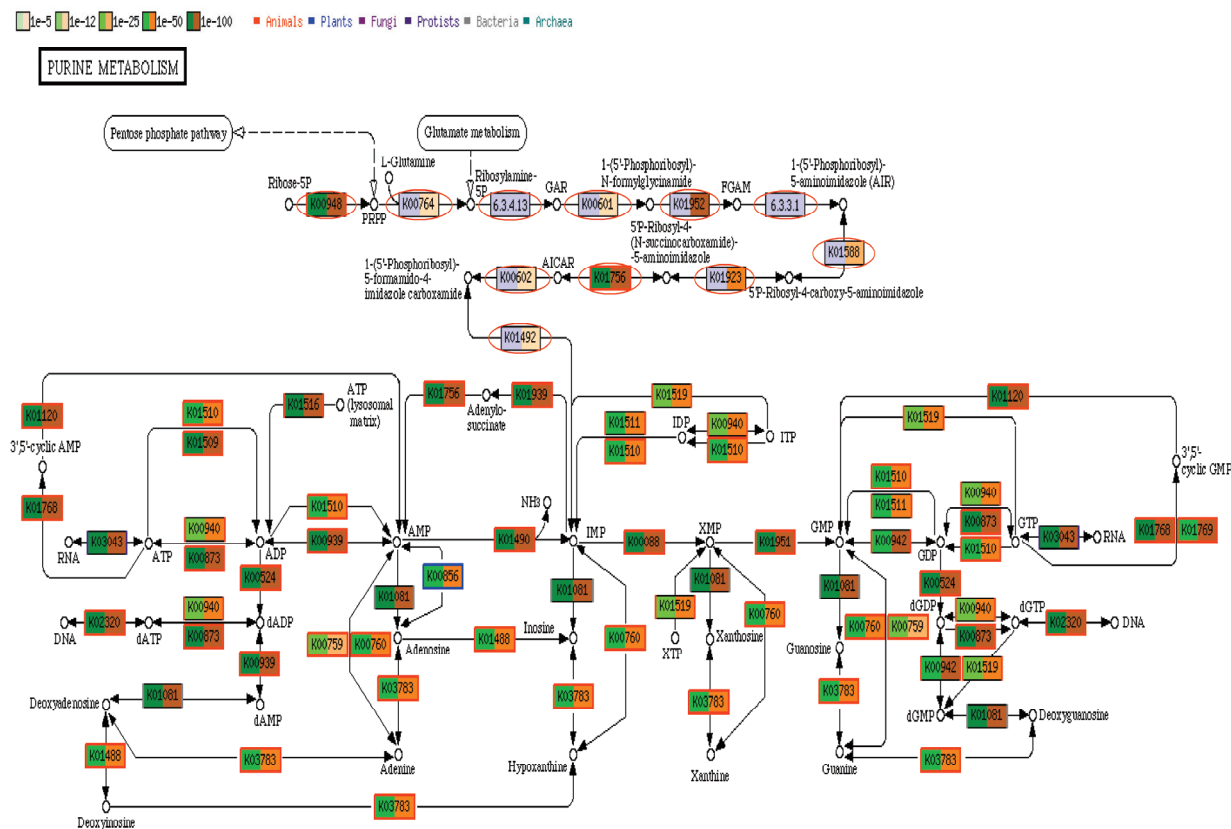
VALINE, LEUCINE AND ISOLEUCINE BIOSYNTHESIS



Supplementary Figure 9.10. Valine, leucine and isoleucine metabolism

None of the key enzymes required for biosynthesis of valine, leucine or isoleucine were identifiable by genome annotation.

Supplementary Figure 10. Nucleotide metabolism (include two sub-figures)



Supplementary Figure 10.1. Purine metabolism

Purine bases cannot be synthesized *de novo* because nine of 10 enzymes for synthesis of purines from PRPP to inosine monophosphate (IMP) are absent; intriguingly only adenylosuccinate lyase (EC:4.3.2.2, K01756) is present. However, the genome annotation revealed that the interconversion between purine bases and nucleosides could occur owing to the presence of all enzymes, suggesting the blood fluke can use the salvage pathway for purine synthesis. The red ovals mean a BLAST value less than $1e-10$ between human and *Schmidtea* enzyme genes, indicating the existence of this gene and that the pathway is complete in *Schmidtea*.

1e-5 1e-12 1e-25 1e-50 1e-100 Animals Plants Fungi Protists Bacteria Archaea

FATTY ACID METABOLISM

Acetyl-CoA

CoA

Acetoacetyl-CoA

(S)-3-Hydroxybutanoyl-CoA

trans-But-2-enoyl-CoA

Butanoyl-CoA

CoA

3-Oxo-hexanoyl-CoA

(S)-3-Hydroxyhexanoyl-CoA

trans-Hex-2-enoyl-CoA

Hexanoyl-CoA

CoA

3-Oxo-octanoyl-CoA

(S)-3-Hydroxyoctanoyl-CoA

trans-Oct-2-enoyl-CoA

Octanoyl-CoA

CoA

3-Oxo-decanoyl-CoA

(S)-3-Hydroxydecanoyl-CoA

trans-Dec-2-enoyl-CoA

Decanoyl-CoA

CoA

3-Oxo-dodecanoyl-CoA

(S)-3-Hydroxydodecanoyl-CoA

trans-Dodec-2-enoyl-CoA

Dodecanoyl-CoA

CoA

3-Oxo-tetradecanoyl-CoA

(S)-3-Hydroxytetradecanoyl-CoA

trans-Tetradec-2-enoyl-CoA

Tetradecanoyl-CoA

CoA

3-Oxo-hexadecanoyl-CoA

(S)-3-Hydroxyhexadecanoyl-CoA

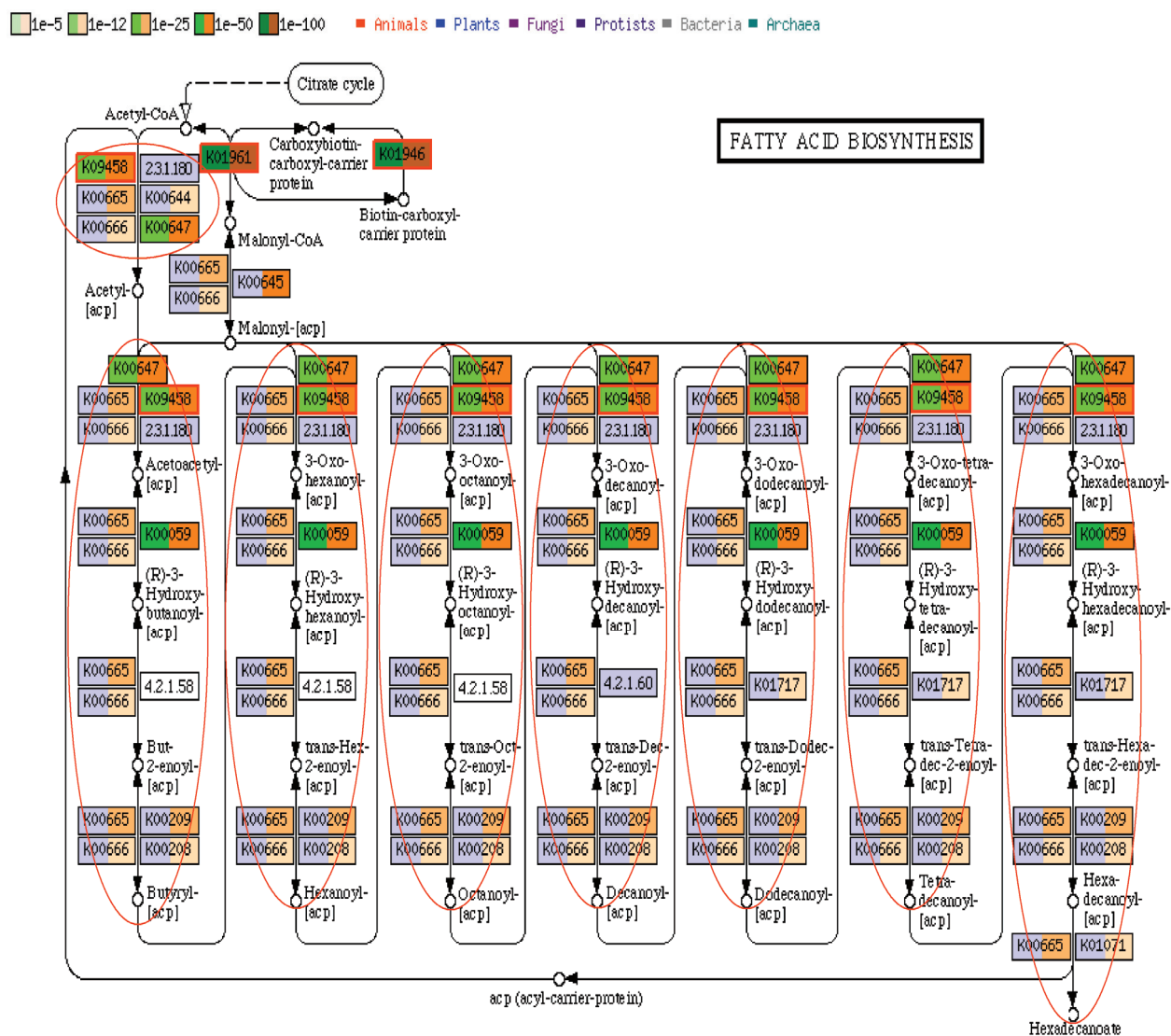
trans-Hexadec-2-enoyl-CoA

Hexadecanoyl-CoA

Hexadecanoate (Fatty acid)

EC numbers: 1.2.1.74, 2.3.1.41, 2.3.1.42, 2.3.1.43, 2.3.1.44, 2.3.1.45, 2.3.1.46, 2.3.1.47, 2.3.1.48, 2.3.1.49, 2.3.1.50, 2.3.1.51, 2.3.1.52, 2.3.1.53, 2.3.1.54, 2.3.1.55, 2.3.1.56, 2.3.1.57, 2.3.1.58, 2.3.1.59, 2.3.1.60, 2.3.1.61, 2.3.1.62, 2.3.1.63, 2.3.1.64, 2.3.1.65, 2.3.1.66, 2.3.1.67, 2.3.1.68, 2.3.1.69, 2.3.1.70, 2.3.1.71, 2.3.1.72, 2.3.1.73, 2.3.1.74, 2.3.1.75, 2.3.1.76, 2.3.1.77, 2.3.1.78, 2.3.1.79, 2.3.1.80, 2.3.1.81, 2.3.1.82, 2.3.1.83, 2.3.1.84, 2.3.1.85, 2.3.1.86, 2.3.1.87, 2.3.1.88, 2.3.1.89, 2.3.1.90, 2.3.1.91, 2.3.1.92, 2.3.1.93, 2.3.1.94, 2.3.1.95, 2.3.1.96, 2.3.1.97, 2.3.1.98, 2.3.1.99, 2.3.1.100, 2.3.1.101, 2.3.1.102, 2.3.1.103, 2.3.1.104, 2.3.1.105, 2.3.1.106, 2.3.1.107, 2.3.1.108, 2.3.1.109, 2.3.1.110, 2.3.1.111, 2.3.1.112, 2.3.1.113, 2.3.1.114, 2.3.1.115, 2.3.1.116, 2.3.1.117, 2.3.1.118, 2.3.1.119, 2.3.1.120, 2.3.1.121, 2.3.1.122, 2.3.1.123, 2.3.1.124, 2.3.1.125, 2.3.1.126, 2.3.1.127, 2.3.1.128, 2.3.1.129, 2.3.1.130, 2.3.1.131, 2.3.1.132, 2.3.1.133, 2.3.1.134, 2.3.1.135, 2.3.1.136, 2.3.1.137, 2.3.1.138, 2.3.1.139, 2.3.1.140, 2.3.1.141, 2.3.1.142, 2.3.1.143, 2.3.1.144, 2.3.1.145, 2.3.1.146, 2.3.1.147, 2.3.1.148, 2.3.1.149, 2.3.1.150, 2.3.1.151, 2.3.1.152, 2.3.1.153, 2.3.1.154, 2.3.1.155, 2.3.1.156, 2.3.1.157, 2.3.1.158, 2.3.1.159, 2.3.1.160, 2.3.1.161, 2.3.1.162, 2.3.1.163, 2.3.1.164, 2.3.1.165, 2.3.1.166, 2.3.1.167, 2.3.1.168, 2.3.1.169, 2.3.1.170, 2.3.1.171, 2.3.1.172, 2.3.1.173, 2.3.1.174, 2.3.1.175, 2.3.1.176, 2.3.1.177, 2.3.1.178, 2.3.1.179, 2.3.1.180, 2.3.1.181, 2.3.1.182, 2.3.1.183, 2.3.1.184, 2.3.1.185, 2.3.1.186, 2.3.1.187, 2.3.1.188, 2.3.1.189, 2.3.1.190, 2.3.1.191, 2.3.1.192, 2.3.1.193, 2.3.1.194, 2.3.1.195, 2.3.1.196, 2.3.1.197, 2.3.1.198, 2.3.1.199, 2.3.1.200, 2.3.1.201, 2.3.1.202, 2.3.1.203, 2.3.1.204, 2.3.1.205, 2.3.1.206, 2.3.1.207, 2.3.1.208, 2.3.1.209, 2.3.1.210, 2.3.1.211, 2.3.1.212, 2.3.1.213, 2.3.1.214, 2.3.1.215, 2.3.1.216, 2.3.1.217, 2.3.1.218, 2.3.1.219, 2.3.1.220, 2.3.1.221, 2.3.1.222, 2.3.1.223, 2.3.1.224, 2.3.1.225, 2.3.1.226, 2.3.1.227, 2.3.1.228, 2.3.1.229, 2.3.1.230, 2.3.1.231, 2.3.1.232, 2.3.1.233, 2.3.1.234, 2.3.1.235, 2.3.1.236, 2.3.1.237, 2.3.1.238, 2.3.1.239, 2.3.1.240, 2.3.1.241, 2.3.1.242, 2.3.1.243, 2.3.1.244, 2.3.1.245, 2.3.1.246, 2.3.1.247, 2.3.1.248, 2.3.1.249, 2.3.1.250, 2.3.1.251, 2.3.1.252, 2.3.1.253, 2.3.1.254, 2.3.1.255, 2.3.1.256, 2.3.1.257, 2.3.1.258, 2.3.1.259, 2.3.1.260, 2.3.1.261, 2.3.1.262, 2.3.1.263, 2.3.1.264, 2.3.1.265, 2.3.1.266, 2.3.1.267, 2.3.1.268, 2.3.1.269, 2.3.1.270, 2.3.1.271, 2.3.1.272, 2.3.1.273, 2.3.1.274, 2.3.1.275, 2.3.1.276, 2.3.1.277, 2.3.1.278, 2.3.1.279, 2.3.1.280, 2.3.1.281, 2.3.1.282, 2.3.1.283, 2.3.1.284, 2.3.1.285, 2.3.1.286, 2.3.1.287, 2.3.1.288, 2.3.1.289, 2.3.1.290, 2.3.1.291, 2.3.1.292, 2.3.1.293, 2.3.1.294, 2.3.1.295, 2.3.1.296, 2.3.1.297, 2.3.1.298, 2.3.1.299, 2.3.1.300, 2.3.1.301, 2.3.1.302, 2.3.1.303, 2.3.1.304, 2.3.1.305, 2.3.1.306, 2.3.1.307, 2.3.1.308, 2.3.1.309, 2.3.1.310, 2.3.1.311, 2.3.1.312, 2.3.1.313, 2.3.1.314, 2.3.1.315, 2.3.1.316, 2.3.1.317, 2.3.1.318, 2.3.1.319, 2.3.1.320, 2.3.1.321, 2.3.1.322, 2.3.1.323, 2.3.1.324, 2.3.1.325, 2.3.1.326, 2.3.1.327, 2.3.1.328, 2.3.1.329, 2.3.1.330, 2.3.1.331, 2.3.1.332, 2.3.1.333, 2.3.1.334, 2.3.1.335, 2.3.1.336, 2.3.1.337, 2.3.1.338, 2.3.1.339, 2.3.1.340, 2.3.1.341, 2.3.1.342, 2.3.1.343, 2.3.1.344, 2.3.1.345, 2.3.1.346, 2.3.1.347, 2.3.1.348, 2.3.1.349, 2.3.1.350, 2.3.1.351, 2.3.1.352, 2.3.1.353, 2.3.1.354, 2.3.1.355, 2.3.1.356, 2.3.1.357, 2.3.1.35

Beta-oxidation of fatty acids cannot be carried out, which is in agreement with the fact that *S. japonicum* does not use fatty acids as an energy source. The presence of enoyl-CoA hydratase (EC:4.2.1.17, K01692) and 3-hydroxyacyl-CoA dehydrogenase (EC:1.1.1.35, K00022) could contribute to the modification and conversion of fatty acids. The red ovals represent the genes which are absent in *S. japonicum*. The BLAST values were less than 1e-10 between the human and *Schmidtea* enzyme genes, suggesting this pathway is complete in *Schmidtea*.

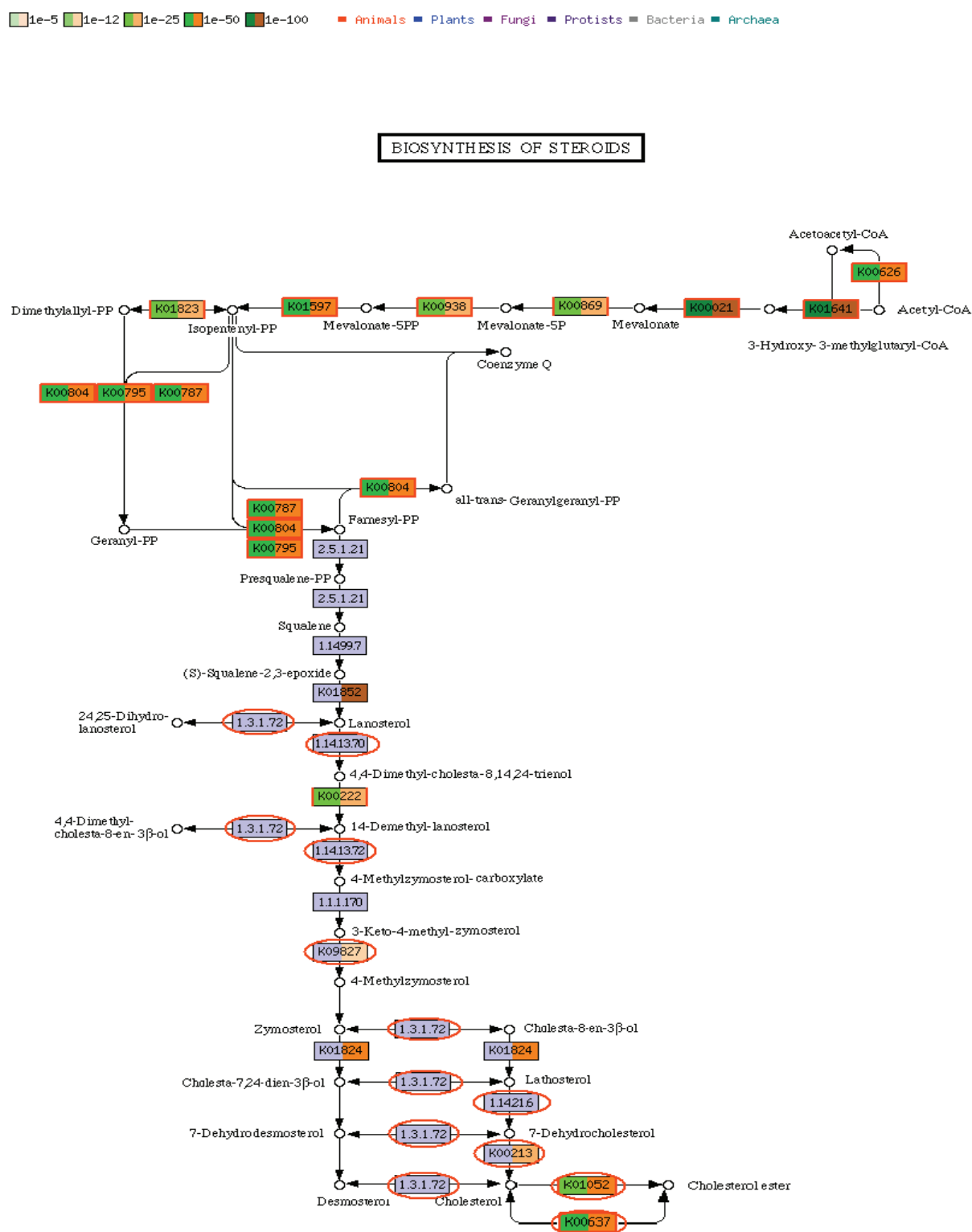


Supplementary Figure 11.2. Fatty acid biosynthesis

S. japonicum cannot carry out fatty acid biosynthesis *de novo*, as only 3-oxoacyl-[acyl-carrier-protein] synthase II, EC:2.3.1.179, K09458) - which can transfer Acetyl-CoA to Acetyl-ACP - among the 7 members of the enzyme complex is present. The BLAST values were less than 1e-10 between the human and *Schmidtea* enzyme genes, suggesting this pathway is complete in *Schmidtea*.



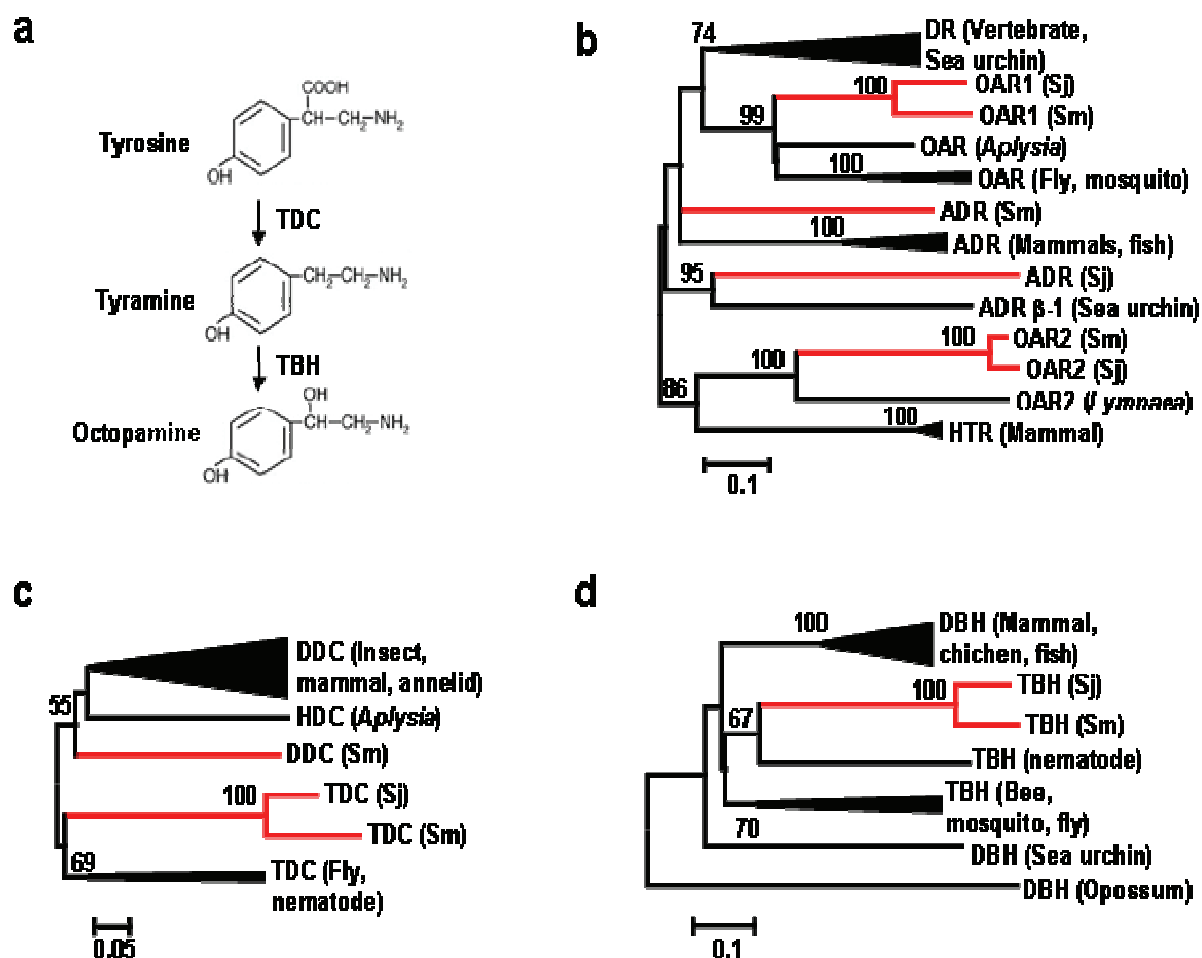
S. japonicum can generate glycerophospholipids such as phosphocholine, phosphatidyl-L-serine and phosphatidylethanolamine, which are essential cell membrane components.



Supplementary Figure 11.4. Cholesterol biosynthesis pathway

S. japonicum cannot carry out the *de novo* biosynthesis of cholesterol, but can convert (host derived) cholesterol ester to cholesterol given the presence of the enzymes cholesterol esterase (K01052, EC:3.1.1.13 3.1.1.3) and sterol O-acyltransferase

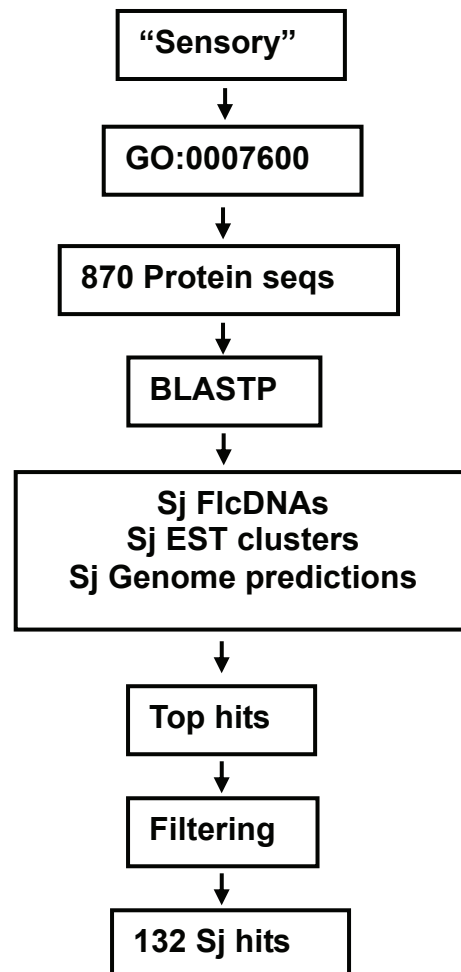
(K00637, EC:2.3.1.26). Of note, the metabolism of 3-hydroxy-3-methylglutaryl-CoA to farnesyl-diphosphate, which could then contribute to the biosynthesis of CoQ, should be possible since the enzymatic machinery is present. However, no requisite enzymes are present to convert farnesyl-diphosphate to cholesterol. The red ovals represent a BLAST value less than $1e-10$ between the human and *Schmidtea* enzymes; the blue ovals represent an E-value greater than $1e-10$, suggesting the pathway is also incomplete in *Schmidtea*.



Supplementary Figure 12. Octopamine synthesis and phylogenetic analysis of octopamine receptors and synthesizing enzymes

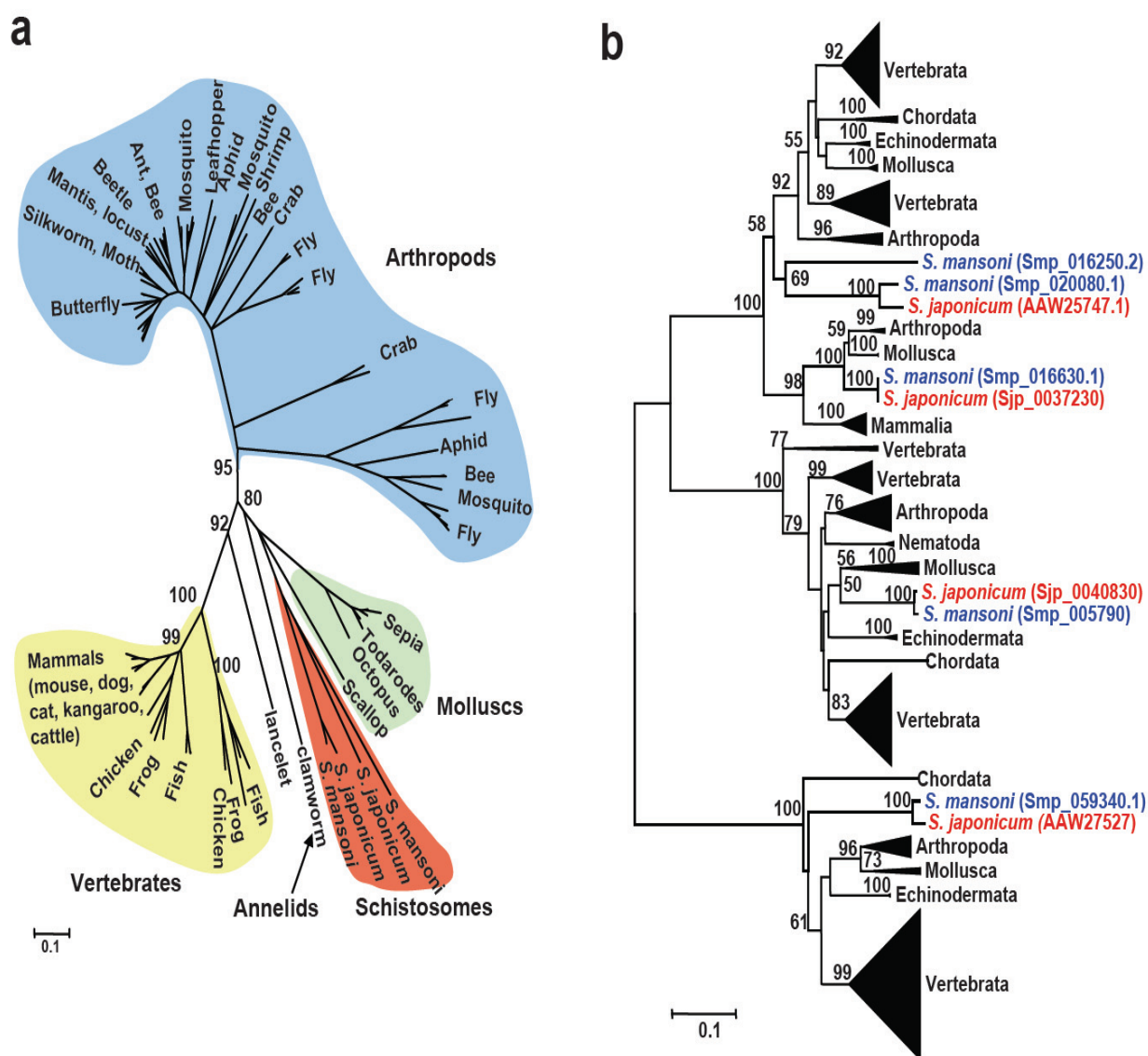
a, The biosynthesis of octopamine (OA). Tyrosine is converted into tyramine by tyrosine decarboxylase (TDC), and tyramine is then converted into OA by tyramine β-hydroxylase (TBH). b, The putative octopamine receptors (OARs) of *S. japonicum* were grouped with OARs from snails and insects. *Schistosoma* OAR1 are from Sjp_0214550 and Smp_contig006303, respectively. *Schistosoma* OAR2 are from Sjp_0007620 of *S. japonicum* and Smp_134820 of *S. mansoni*, respectively. Putative adrenergic receptors (ADRs) are from *S. japonicum* (Sjp_0206980) and *S. mansoni* (Smp_150180). c, The phylogenetic relationship of putative *S. japonicum* TDC with

the dopa decarboxylase (DDC) family. *Schistosoma* TDCs are SJCHGC04754 for *S. japonicum* and Smp_135230 for *S. mansoni*, respectively. A putative DDC is from *S. mansoni* (Smp_171580). d, Phylogenetic analysis of *Schistosoma* TBH or dopamine beta hydroxylase (DBH). *Schistosoma* TBHs are Sjp_0056290 for *S. japonicum* and Smp_163900 for *S. mansoni*, respectively. The *S. japonicum* proteins are highlighted with bold type and the branches leading to the schistosomes are shown in red. HDC, histidine decarboxylase; DR, dopamine receptor; HTR, 5-hydroxytryptamine (serotonin) receptor. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 3.1 using the Neighbour-joining method followed by 1000-bootstraps test.



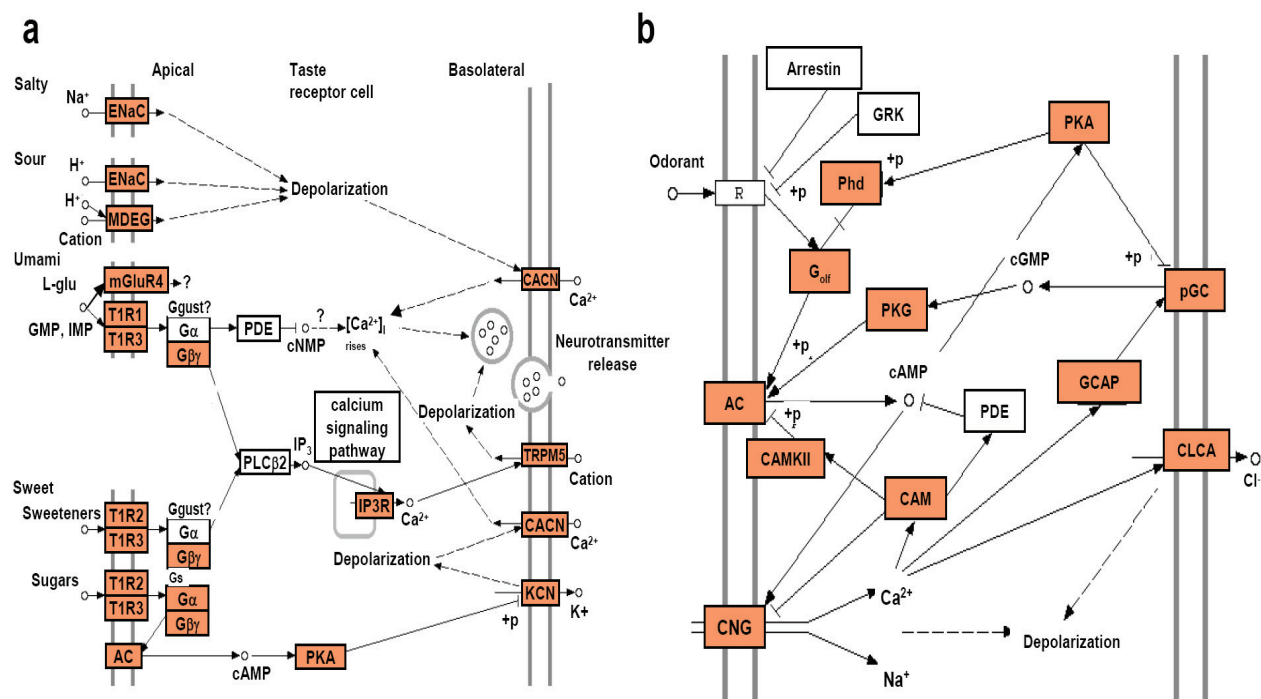
Supplementary Figure 13. Profile the sensory components of *S. japonicum* using a top-down strategy

A bottom-up GO annotation strategy was used for searching sensory components.



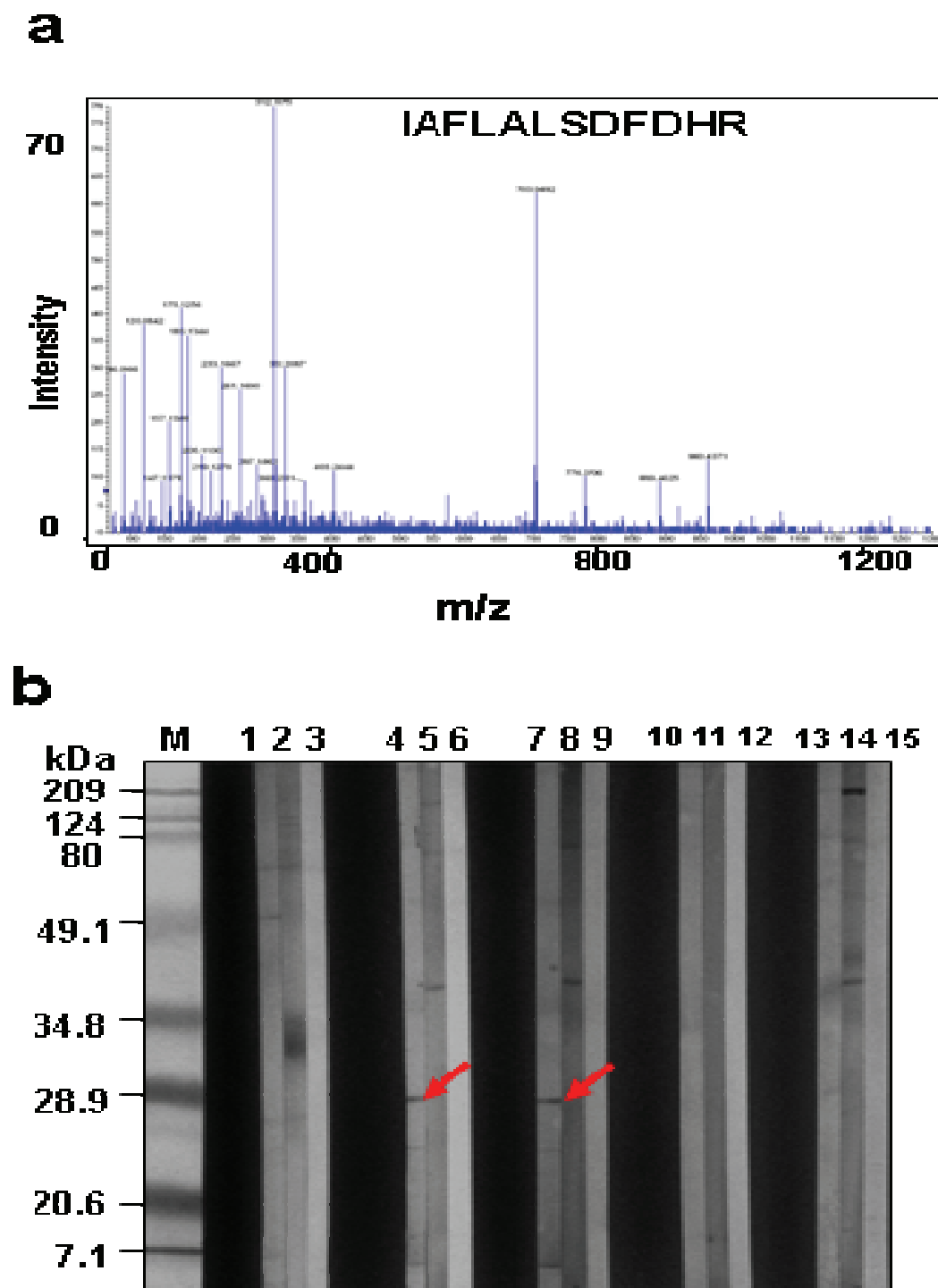
Supplementary Figure 14. Phylogenetic analysis of light sensory components in *S. japonicum*

a, Phylogenetic analysis of members of the rhodopsin superfamily in schistosomes and other organisms. **b**, Phylogenetic analysis performed on schistosome transducins within the transducin superfamily in vertebrates, chordates, echinoderms, molluscs and arthropods, suggesting that schistosome transducins comprise at least four members of this superfamily, where each member represents a divergent subtype. The transducins of *S. japonicum* and *S. mansoni* are colored red and blue, respectively. The trees were generated with Mega v3.1 by the neighbor-joining method and tested with 1000 bootstraps.



Supplementary Figure 15. Putative taste and olfactory pathways in *S. japonicum*

a, Putative receptors and downstream components for a variety of taste. **b**, Putative receptors and downstream components for odorants. The components encoded by *S. japonicum* genome are colored in orange. Abbreviations used are: ENaC, epithelial Na(+) channel; MDEG, mammalian degenerin homolog; mGluR4, glutamate receptor, metabotropic 4; Gbg, guanine nucleotide binding protein (G protein), beta 1; Ga, guanine nucleotide binding protein (G protein), alpha; IP3R, inositol 1,4,5-triphosphate receptor, type 3; PKA, protein kinase A; CACN, calcium channel, voltage-dependent, P/Q type, alpha 1A subunit; TRPM5, transient receptor potential cation channel, subfamily M, member 5; KCN, potassium voltage-gated channel, Shab-related subfamily B, member 1; AC, adenylate cyclase 3; CNG, cyclic nucleotide gated channel alpha 3; Phd, phosducin; Golf, guanine nucleotide binding protein; CAMKII, CaM kinase; PKG, protein kinase, cGMP-dependent; CAM, calmodulin; GCAP, guanylate cyclase activator 1; CLCA, chloride channel, calcium activated, family member 2.



Supplementary Figure 16. Identification of *S. japonicum* elastase

a, Mass spectrum analysis of the proteome of cercariae of *S. japonicum* identifies a peptide fragment of SjCE. The letters shown indicate the amino acid sequence of the

peptide fragment. b, Immunoblotting analysis of *S. japonicum* extracts. Lanes 1-3, 4-6, 7-9, 10-12 and 13-15 represent protein extracts from eggs, sporocysts, cercariae, schistosomula and adults, respectively. Immunoblotting of lanes 1, 4, 7, 10 and 13 was performed using rabbit anti-recombinant SjCE antibodies. Lanes 2, 5, 8, 11 and 14 were probed with serum taken from rabbits 42 days after challenge infection with *S. japonicum*. Lanes 3, 6, 9, 12 and 15 were probed with normal rabbit serum. M: molecular mass markers. The position of SjCE is indicated by red arrows.