Whole-genome sequence of Schistosoma haematobium

Author

Young, Neil D, Jex, Aaron R, Li, Bo, Liu, Shiping, Yang, Linfeng, Xiong, Zijun, Li, Yingrui, Cantacessi, Cinzia, Hall, Ross S, Xu, Xun, Chen, Fangyuan, Wu, Xuan, Zerlotini, Adhemar, Oliveira, Guilherme, Hofmann, Andreas, Zhang, Guojie, Fang, Xiaodong, Kang, Yi, Campbell, Bronwyn E, Loukas, Alex, Ranganathan, Shoba, Rollinson, David, Rinaldi, Gabriel, Brindley, Paul J, Yang, Huanming, Wang, Jun, Wang, Jian, Gasser, Robin B

Published

2012

Journal Title

Nature Genetics

DOI

10.1038/ng.1065

LETTER

nature genetics

NG-LE30434R.R2 Gasser (15.09.11; 4 display items: 1 table and 3 figures)

The *Schistosoma haematobium* genome - in response to the wake-up call

Neil D. Young[1*], Aaron R. Jex[1*], Bo Li[2*], Shiping Liu[2], Linfeng Yang[2], Zijun Xiong[2], Yingrui Li[2], Cinzia Cantacessi[1], Ross S. Hall[1], Xun Xu[2], Fangyuan Chen[2], Xuan Wu[2], Adhemar Zerlotini[3], Guilherme Oliveira[3], Andreas Hofmann[4], Guojie Zhang[2], Xiaodong Fang[2], Yi Kang[2], Bronwyn E. Campbell[1], Alex Loukas[5], Shoba Ranganathan[6], David Rollinson[7], Gabriel Rinaldi[8.9], Paul J. Brindley[9], Huanming Yang[2], Jun Wang[2**], Jian Wang[2**] & Robin B. Gasser[1**]

[1]The University of Melbourne, Parkville, Victoria 3010, Australia. [2]BGI, Shenzhen 518083, PR China. [3]Centro de Excelência em Bioinformática, Laboratório de Parasitologia Celular e Molecular, Instituto Nacional de Ciência e Tecnologia em Doenças Tropicais Instituto de Pesquisa René Rachou-Fiocruz, Belo Horizonte, Brasil. [4]Eskitis Institute for Cell & Molecular Therapies, Griffith University, Brisbane, Queensland, Australia. [5]Queensland Tropical Health Alliance, James Cook University, Cairns, Queensland 4878, Australia. [6]Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales 2109, Australia. [7]Wolfson Wellcome Biomedical Laboratories, Department of Zoology, Natural History Museum, London SW7 5BD, UK. [8]Departamento de Genética, Facultad de Medicina 11800, Universidad de la República (UDELAR), Montevideo, Uruguay. [9]Department of Microbiology, Immunology and Tropical Medicine, The George Washington University Medical Center, Washington DC, 20037 USA.

*equal first authors.
**corresponding authors.

Schistosomiasis is a neglected tropical disease affecting 200 million people worldwide and causing 300,000 deaths each year[1]. The intestinal and urinary forms of this debilitating disease are caused by chronic infections with *Schistosoma* species (blood flukes or schistosomes). No vaccines are available, and treatment relies on the drug praziquantel[2]. Of the three main species that infect people, *Schistosoma haematobium* has come into the spotlight because of its high prevalence (> 110 million people infected), as a cause of urogenital disease, its proven link to malignant bladder cancer[1,3] and as a predisposing factor for HIV/AIDS[4,5] in endemic countries. Through a complex two-host life cycle, *S. haematobium* is transmitted from a freshwater snail (*Bulinus* spp.) to humans[1]. Adult schistosomes dwell as pairs in the blood vessels of the urinary bladder, where female worms release eggs that become embedded in the bladder wall to elicit chronic immune-mediated disease[6] and serious complications, including squamous cell carcinoma[7]. Here, we report the 385 million-base (Mb) draft genome of *S. haematobium* and compare it with sequences from related parasites[8,9]. This genome provides an unprecedented resource for fundamental genomic, genetic, evolutionary, biological and epidemiological research and exciting prospects for the design of new interventions against a largely unknown disease complex of humans.

From 200 ng of genomic DNA isolated from one single pair of adult (male and female) worms a single, mated pair (i.e. one male and one female worm) of adult *S. haematobium*, we produced 33.5 Gb of useable sequence data (**Supplementary Data 1**). We showed consistently low sequence heterozygosity and estimated the genome size to be 431-452 Mb. We then assembled the data and used local assemblies to close most (96.1%) of the remaining gaps, achieving a final assembly of 385 Mb (365 contigs; scaffold-N50 of 307 kb; 74-fold coverage) (**Table 1** and **Supplementary Data 1**). The GC content (mean: 34.3%) was similar to that of *S. mansoni* and *S. japonicum* (**Table 1**). Following assembly, all useable reads were realigned to scaffolds to assess single-base accuracy of the assembled genome sequence (**Supplementary Data 2**). We neither found evidence of GC-biased non-random sampling[10] nor of multiple displacement amplification (MDA)-induced artefacts, consistent with published information[11,12] (**Supplementary Data 2).**

Comparison of the *S. haematobium* and *S. mansoni* genomes revealed a similar percentage and composition of repetitive elements (**Table 1** and **Supplementary Data 3**). Using both homology-based and *de novo* predictions, we estimated that 43% of the genome comprises repetitive elements, consistent with the *S. mansoni* genome (40%)[8]. More than half (58.5%) of the repeats were retrotransposons (at least 20 types, including LINE/RTE-BovB and LTR/*Gypsy*), followed by unknown repeats (37.5%), satellites (1.9%), simple repeats (1.2%) and DNA transposons (five types; < 1%). We inferred 13,073 protein-coding genes from the genome, based on homology and *de novo* predictions as well as evidence of transcription (in adult and egg stages), and included data for *S. mansoni* and *S. japonicum* for comparisons (**Supplementary Data 4**). The number of *S. haematobium* genes was consistent with those of *S. mansoni* (13,184) and *S. japonicum* (13,469), as were the gene structures. Most (9,714) *S. haematobium* genes were supported by the RNA-seq data from adult and egg stages. Comparative analyses of the complete gene set showed higher nucleotide sequence identity (mean: 92%) and length match for individual coding domains between *S. haematobium* and *S. mansoni* than between *S. haematobium* and *S. japonicum* (86%) or *S. mansoni* and *S. japonicum* (86%).

For the protein-encoding genes (~4.4% of the *S. haematobium* genome), 96.3% had matches in non-redundant databases, 52.8% had conserved protein domains and 43% could be mapped to known biological pathways (**Supplementary Data 5**). These data allowed 44% of genes to be classified by gene ontology (GO), providing a list of terms that was consistent with *S. mansoni* and *S. japonicum*. A small percentage (2.6%) of *S. haematobium* genes were predicted to be excreted/secreted (ES), including those in the egg (omega-1, interleukin-4-inducing protein) and/or adult (e.g., cathepsin B, heat shock proteins, thioredoxin peroxidase, superoxide dismutase, protein disulfide isomerase and venom allergen-like proteins)[13,14]. High-stringency genetic networking from the entire genomic dataset identified major hubs of connectivity for conserved molecules associated with nucleotide and protein synthesis, degradation and signal transduction (**Supplementary Data 6**).

A genome-wide analysis revealed a significantly higher synteny between *S. haematobium* and *S. mansoni* (89.4%) than between *S. haematobium* and *S. japonicum* (51.7%) or *S. mansoni* and *S. japonicum* (67.0%). Intrachromosomal rearrangements of orthologues within large genome scaffolds (> 1 Mb) were not observed between *S. haematobium* and *S. mansoni* but were detected between *S. japonicum* and *S. mansoni* (**Supplementary Data 7**), consistent with current knowledge of evolutionary relationships[15] and karyotypes[16]. Given the close relationship between *S. haematobium* and *S. mansoni*, and size and quality of the draft genome for *S. mansoni*[8], we aligned *S. haematobium* to *S. mansoni* scaffolds that mapped to chromosomes (2n = 16; ZZ)[17]. Overall, rearrangements in *S. haematobium*, with respect to *S. mansoni*, were rare, with 11 inversions of syntenic blocks linked to four chromosomes (nos. 1, 3, 4, 6 and Z) (**Fig. 1** and **Supplementary Data 7)**.

S. haematobium shared proteins with *S. mansoni* (n = 1,333) and *S. japonicum* (235), respectively (**Supplementary Data 8)**, of which only a minor portion could be assigned functional categories (KEGG BRITE) linked to a wide array of different molecular groups. Of the former (1,333), 91 represented mainly enzymes (such as kinases, glycosyltransferases and peptidases) and cytoskeletal, DNA repair, replication, recombination and spliceosome proteins, and elements of the ubiquitin system. Of the latter (235), 33 were linked to various metabolic enzymes, cytoskeletal proteins, transcription factors and also some proteins in the ubiquitin complex. A subset of 73 molecules was unique to *S. haematobium* and not detected in the other two human schistosomes (**Supplementary Data 8**); despite these molecules containing structural elements, such as alpha-helices and beta-sheets, none of them had homology to any presently known eukaryotic proteins or contained conserved motifs. Of 10,880 proteins common among *S. haematobium*, *S. mansoni* and *S. japonicum*, we identified 6,142 homologues in other flatworms, including *Fasciola hepatica*, *F. gigantica*, *Clonorchis sinensis* and *Opisthorchis viverrini*. Utilizing concatenated protein sequence data inferred from a subset of 59 single-copy gene homologues (**Supplementary Data 9**), we were able to provide a robust inference of the genetic relationships of socio-economically important trematodes, in which *S. haematobium* and *S. mansoni* were most closely related, followed by *S. japonicum*, to the exclusion of other trematodes (**Fig. 2**). The relationship of the schistosomes was in accordance with previous studies using mitochondrial and/or small nuclear DNA markers[15]. The present phylogenetic analysis extends our understanding of the evolution of key trematodes, and the approach used provides a sound basis for future, large-scale evolutionary analyses when extensive genomic and transcriptomic datasets become available for a wide range of flatworms.

Having assessed genetic relationships, we then proceeded to explore transcription in the adult male, female and egg stages of *S. haematobium*, which cause disease in the human host (**Fig. 3** and **Supplementary Data 9**). The number of transcribed genes was greater in adult females (10,002) and males (10,559) compared with eggs (7,794). Egg-enriched transcription was observed for 1,631 genes, including those encoding venom allergen-like and cd63 tetraspanin-like proteins (reported to be transcribed exclusively in the miracidium of *S. mansoni*)[18] as well as two encoding CP391S-like egg proteins. Adult-enriched transcription was shown for 2,975 genes, of which 965 were significantly up-regulated in females and 2,010 in males. These findings are similar to those of recent microarray studies of *S. mansoni* and *S. japonicum*[18,19], although our interpretation is guarded, at this stage, due to the use of different animal hosts for parasite production and analytical methods among studies. Indeed, given the substantial depth of the present RNA-seq dataset (compared with microarray), we were able to accurately profile enriched biological pathways in the different stages/sexes of *S. haematobium*, which clearly reflect its biology and life-style. Notable in the adult female, which lives within the gynaecophoric canal of the male, is a major enrichment in pathways linked to haematophagia (superoxide dismutase, saposin and cathepsin B and ferritin) and egg production (including lipid metabolism, protein synthesis and egg shell-specific proteins), whereas the pathway enrichment in the adult male reflects its major need to maintain a highly developed musculature (pertaining largely to myosin, paramyosin, tropomyosin and troponin) for motility, feeding, attachment and grasping the female *in copula*. In contrast, pathway enrichment in the egg stage relates predominantly to egg-specific proteins and an interleukin-4-inducing factor, which is a secreted glycoprotein that induces IgE-dependent interleukin-4 production in basophils and likely relates to immune modulation and/or alteration in the host, as proposed for *S. mansoni*[20]. Having explored transcription, we then interaction-networked all genes inferred to be essential and transcribed constitutively or in a developmentally-regulated manner in *S. haematobium*[21] and prioritized 6 molecules (See **Fig. 3** and **Supplementary Data 10**) as prime targets for the design of novel trematocides. Although a small number of drug targets was predicted in *S. haematobium* using our bioinformatic pipeline, under very stringent selection criteria, all 72 'druggable' molecules inferred previously for *S. mansoni*[21] were represented in the *S. haematobium* proteome.

Schistosomes have adapted to their mammalian hosts to such an extent that they can survive for decades in an intricate host-parasite interplay without succumbing. They achieve their longevity by suppression, diversion and alteration of immune responses[22]. Chronic infections induce key changes in immune cell populations, including a dominance of the T-helper 2 (Th2) cells and selective loss of effector T cell activity, against a background of regulatory T cells, alternatively activated macrophages, and Th2-inducing dendritic cells[22]. Much of the immunomodulatory capacity of schistosomes is attributable particularly to ES products[23]. In the *S. haematobium* proteome, we identified 55 molecules (20 of which were predicted to be ES proteins) with known immunomodulatory roles in other helminths (**Supplementary Data 11**); these include molecules linked to inhibition of antigen processing and presentation via binding (Sjc23 tetraspanin), cleavage (cysteine and serine proteases) or inhibition of posttranslational modification (cystatins) of host immunoglobulins, known inducers of Th2 responses (IPSE/alpha 1, omega-1, peroxiredoxin and sm16/SPO-1) and host defence mimicry molecules (e.g., C-type lectins). Interestingly, we also identified an estradiol 17beta dehydrogenase orthologue (**Supplementary Data 11**), which has a known role in the synthesis of

estradiol. Intriguingly, *S. haematobium* but not *S. mansoni* ES products stimulate wound healing, mitosis and cell migration and also down-regulate apoptosis, all properties expected to be conducive to tumourigenesis, and one or more estradiol-like molecules have been implicated in this genesis[24]. Even though orthologues of 17beta dehydrogenase exist in *S. mansoni* and *S. japonicum*, the specific spatial and temporal expression associated with the synthesis of estradiol-like molecules in *S. haematobium* eggs in situ, in the bladder, might contribute to carcinogenesis, warranting detailed exploration. Moreover, in spite of limited proteomic differences among human schistosomes, substantial variation in splicing[25], differential methylation[26], regulatory RNAs[27] and other epigenetic processes is likely. These are areas that can now be tackled readily using genome information for all three human schistosomes.

Clearly much remains unknown about fundamental biology and pathogenesis of schistosomes, which cause considerable morbidity to many millions of people and animals worldwide[1,28]. Given the challenges in propagating these parasites in the laboratory, particularly *S. haematobium*[29], the ability to sequence the genome from a single pair of worms, as realized here, represents an unprecedented step towards deciphering the genomes of a diverse range of other schistosomes and neglected tropical disease pathogens, including the food-borne flukes, and to addressing fundamental and controversial questions regarding their genetics, evolution, ecology, epidemiology, pathogenesis and host-parasite relationships. The genome provides a solid foundation for future large-scale and integrated studies of gene function and essentiality, employing tools such as RNA interference and transgenesis[28], and will also facilitate urgently-needed proteomic explorations. Published findings[30] show that developmental stages of *S. haematobium* can be manipulated genetically and that effective gene silencing can be achieved, which now provides enormous scope for future large-scale functional genomic analyses. Unlocking the molecular biology of this and related, neglected and globally important disease pathogens using a raft of ‘omic technologies, particularly functional genomics, proteomics and metabolomics, will provide entirely novel and unique insights into schistosome development, host-parasite affiliations, disease and schistosomiasis-associated bladder cancer, and will underpin the design of new diagnostic tools, anti-schistosome drugs and vaccines.

METHODS

We sequenced the genome of *S. haematobium* by Illumina technology from whole-genome amplified genomic DNA from one single pair of adult (male and female) worms. From six paired-end sequencing libraries (insert sizes: ~ 0.17 to 10 kb), we generated 33.5 Gb of useable short-read sequence data, equating to > 74-fold coverage of the 385 Mb genome. We assembled the short-reads, constructed scaffolds, in a step-by-step manner, and then closed intra-scaffold gaps. Transposable elements and the proteome were inferred using *de novo*-, homology- and/or evidence-based approaches. RNA-seq was used to sequence the transcriptomes of the adult (female and male) and egg stages, and expression profiles were compared between stages and sexes. Inferred proteins were classified according to function, gene ontology and pathway mapping, genetic networking, phylogenetic analyses as well as essentially, and drug target predictions were conducted using established or in-house methods.

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

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

ACKNOWLEDGEMENTS

This project was funded by the Australian Research Council (ARC), the National Health and Medical Research Council (NH&MRC) of Australia, and BGI. Other support from the Australian Academy of Science, the Australian-American Fulbright Commission, Melbourne Water Corporation, the Victorian Life Sciences Computation Initiative (VLSCI) and the IBM Collaboratory is gratefully acknowledged. We thank staff of BGI, Shenzhen, including Qin Nan, Pei Na, Bo Min and Peixiang Ni, whose names were not included in the present authorship but who contributed to this study. *Schistosoma haematobium*-infected hamsters were provided by Drs Fred A. Lewis and Yung-San Liang, Biomedical Research Institute, Rockville, MD, USA, under NIAID-NIH contract HHSN272201000005I. G.O. is supported by NIH-Fogarty TW007012, FAPEMIG CBB-1181/08 and CNPq 573839/2008-5.

AUTHOR CONTRIBUTIONS

R.B.G. conceived and led the project, with support from X.W., J.W. and H.M.Y. N.D.Y. and A.R.J designed the experimental plan, and executed and guided the bioinformatic analyses. P.J.B. and G.R. provided parasite material. Guided by N.D.Y., A.R.J. and R.B.G., B.L. coordinated genome assembly and annotation. Y.L. performed the genome assembly, X.Z. conducted comparative genomic analysis, A.H., L.S., Y.K. and C.F. undertook the genome annotation, X.X. and his team performed the whole genome amplification step, and C.C. and B.E.C. conducted the essentiality and drug target predictions. A.Z. and R.S.H. provided bioinformatic support. N.D.Y, A.R.J. and R.B.G. wrote the manuscript, with critical inputs from P.J.B., G.O., S.R., A.L. and D.R., and comments from the other coauthors.

AUTHOR INFORMATION

All data have been released for public use (http://gasser-research.vet.unimelb.edu.au/GasserData/Schistosoma_haematobium_data.html) and are accessible in GeneDB (www.genedb.org/Homepage/Shaematobium). Correspondence and requests regarding different aspects of the work can be addressed to Robin B. Gasser (robinbg@unimelb.edu.au), Jun Wang (wangj@genomics.org.cn) or Jian Wang (wangjian@genomics.org.cn).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Rollinson, D. A wake up call for urinary schistosomiasis: reconciling research effort with public health importance. *Parasitology* **136**, 1593-1610 (2009).
2. Doenhoff, M. J., Cioli, D. & Utzinger, J. Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Curr. Opin. Infect. Dis.* **21**, 659-667 (2008).
3. Bouvard, V. *et al.* A review of human carcinogens-Part B: biological agents. *Lancet Oncol.* **10**, 321-322 (2009).
4. Hotez, P. J., Fenwick, A. & Kjetland, E. F. Africa's 32 cents solution for HIV/AIDS. *PLoS Negl. Trop. Dis.* **3**, e430 (2009).
5. Kjetland, E. F. *et al.* Association between genital schistosomiasis and HIV in rural Zimbabwean women. *AIDS* **20**, 593-600 (2006).
6. Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. Human schistosomiasis. *Lancet* **368**, 1106-1118 (2006).
7. Palumbo, E. Association Between Schistosomiasis and Cancer: A Review. *Infectious Diseases in Clinical Practice* **15**, 145-148 (2007).
8. Berriman, M. *et al.* The genome of the blood fluke *Schistosoma mansoni*. *Nature* **460**, 352-358 (2009).
9. *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* **460**, 345-351 (2009).
10. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59 (2008).
11. Paez, J. G. *et al.* Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res.* **32**, e71 (2004).
12. Valentim, C. L., LoVerde, P. T., Anderson, T. J. & Criscione, C. D. Efficient genotyping of *Schistosoma mansoni* miracidia following whole genome amplification. *Mol. Biochem. Parasitol.* **166**, 81-84 (2009).
13. Liu, F. *et al.* Excretory/secretory proteome of the adult developmental stage of human blood fluke, *Schistosoma japonicum*. *Mol. Cell. Proteomics* **8**, 1236-1251 (2009).
14. Mathieson, W. & Wilson, R. A. A comparative proteomic study of the undeveloped and developed *Schistosoma mansoni* egg and its contents: the miracidium, hatch fluid and secretions. *Int. J. Parasitol.* **40**, 617-628 (2010).
15. Webster, B. L., Southgate, V. R. & Littlewood, D. T. A revision of the interrelationships of *Schistosoma* including the recently described *Schistosoma guineensis*. *Int. J. Parasitol.* **36**, 947-955 (2006).
16. Short, R. B. & Menzel, M. Y. Chromosomes of nine species of schistosomes. *J. Parasitol.* **46**, 273-287 (1960).
17. Criscione, C. D., Valentim, C. L., Hirai, H., Loverde, P. T. & Anderson, T. J. Genomic linkage map of the human blood fluke *Schistosoma mansoni*. *Genome Biol.* **10**, R71 (2009).
18. Fitzpatrick, J. M. *et al.* Anti-schistosomal intervention targets identified by lifecycle transcriptomic analyses. *PLoS Negl. Trop. Dis.* **3**, e543 (2009).
19. Gobert, G. N., Moertel, L., Brindley, P. J. & McManus, D. P. Developmental gene expression profiles of the human pathogen *Schistosoma japonicum*. *BMC Genomics* **10**, 128 (2009).
20. Schramm, G. *et al.* Cutting edge: IPSE/alpha-1, a glycoprotein from *Schistosoma mansoni* eggs, induces IgE-dependent, antigen-independent IL-4 production by murine basophils *in vivo*. *The Journal of Immunology* **178**, 6023-6027 (2007).
21. Caffrey, C. R. *et al.* A comparative chemogenomics strategy to predict potential drug targets in the metazoan pathogen, *Schistosoma mansoni*. *PLoS ONE* **4**, e4413 (2009).

22. Allen, J. E. & Maizels, R. M. Diversity and dialogue in immunity to helminths. *Nat. Rev. Immunol.* **11**, 375-388 (2011).
23. Hewitson, J. P., Grainger, J. R. & Maizels, R. M. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. *Mol Biochem Parasitol* **167**, 1-11 (2009).
24. Botelho, M. C., Machado, J. C. & Correia da Costa, J. M. *Schistosoma haematobium* and bladder cancer: What lies beneath? *Virulence* **1**, 84-87 (2010).
25. Verjovski-Almeida, S. & DeMarco, R. Gene structure and splicing in schistosomes. *Journal of Proteomics* **74**, 1515-1518 (2011).
26. Geyer, K. K. *et al.* Cytosine methylation regulates oviposition in the pathogenic blood fluke *Schistosoma mansoni*. *Nat. Commun.* **2**, 424 (2011).
27. de Souza Gomes, M., Muniyappa, M. K., Carvalho, S. G., Guerra-Sa, R. & Spillane, C. Genome-wide identification of novel microRNAs and their target genes in the human parasite *Schistosoma mansoni*. *Genomics* **98**, 96-111 (2011).
28. Brindley, P. J., Mitreva, M., Ghedin, E. & Lustigman, S. Helminth genomics: The implications for human health. *PLoS Negl. Trop. Dis.* **3**, e538 (2009).
29. Mann, V. H., Morales, M. E., Rinaldi, G. & Brindley, P. J. Culture for genetic manipulation of developmental stages of *Schistosoma mansoni*. *Parasitology* **137**, 451-462 (2010).
30. Rinaldi, G. *et al.* Genetic manipulation of *Schistosoma haematobium,* the neglected schistosome. *PLoS Negl. Trop. Dis.* **in press** (2011).

Table 1 Comparison of the features of the genomes of *Schistosoma haematobium* with those of two other schistosomes

Genomic features	*Schistosoma haematobium*	*Schistosoma mansoni*	*Schistosoma japonicum*
Estimate of genome size (megabases)	385	381	403
Chromosome number (2n)	8[a]	8	8
Total number of base pairs within assembled contigs	361,903,089	374,944,597	369,039,322
N50 contig (length (bp); total number > 500 bp)	21,744; n = 36,826	16,320; n = 50,292	6,121; n = 95,265
Total number of base pairs within assembled scaffolds	385,110,549	381,096,674	402,705,545
N50 scaffold (length (bp); total number > 1000 bp in length)	306,738; n = 7,475	832,5415; n = 19,022	176,869; n = 25,048
Proportion of genome that is coding (%)	4.43	4.72	4.32
Number of putative coding genes	13,073	13,184	13,469
Gene size (average bp ± standard deviation; range)	11,952±16,273; 30-204,220	13,397±18,029; 84-240,193	10,003±12,980; 150-173,394
Average coding domain length (average bp ± standard deviation; range)	1,319±1,502; 30-28,212	1,344±1,447; 60-22,983	1,179±1,201; 147-24,180
Average exon number per gene (average bp ± standard deviation; range)	5.4±5.80; 1-136	6.2±6.24; 1-94	5.3±4.70; 1-65
Gene exon length (average bp ± standard deviation; range)	246±287; 1-9,737	218±236; 1-9,291	223±256; 6-6,326
Gene intron length (average bp ± standard deviation; range)	2,442±2,958; 1-68,754	2,331±3,200; 1-67,221	2,058±2,679; 15-59,770
Total GC content (%)	34.25%	34.71%	33.47%
Repeat rate (%)	47.2	45	40.1

[a] Estimate derived from a karyological study of *S. haematobium*[19]

Figure 1 Synteny inferred between the *Schistosoma haematobium* and *Schistosoma mansoni* genomes. (a) Large blocks of *S. haematobium* genomic scaffolds mapped physically to chromosome 3 of *S. mansoni*. (b) Evidence of an inversion in *S. haematobium* with reference to *S. mansoni*, supported by paired-end sequence data.

Figure 2 Genetic relationship of *Schistosoma haematobium* with other members of class Trematoda. *S. haematobium* was related to trematodes (for which genomic or transcriptomic data are available) using Bayesian inference (BI) analysis of amino acid sequence data inferred from 59 genes. The topology of this tree is the same as that achieved using maximum parsimony (MP) analysis. Nodal support was absolute (100%) for each branch.

Figure 3 Differential transcription among adult female (F), adult male (M) and egg (E) stages of *Schistosoma haematobium*. Transcripts were mapped to each gene in two dimensions, and their relative abundance displayed. Genes transcribed in a constitutive manner are shown within a central 100-pixel radius, and the top-20 genes transcribed in a gender- or stage-enriched manner are shown within a 25-pixel radius of each node. Molecules inferred to be essential (1-8) are indicated, and those representing druggable targets are in yellow. Transcription is expressed as $\log_{10}$-transformed reads per kilobase per million reads (RPKM).