

# SCIENTIFIC REPORTS



OPEN

## Defining the *Schistosoma haematobium* kinome enables the prediction of essential kinases as anti-schistosome drug targets

Received: 06 May 2015  
Accepted: 26 August 2015  
Published: 04 December 2015

Andreas J. Stroehlein<sup>1</sup>, Neil D. Young<sup>1</sup>, Aaron R. Jex<sup>1</sup>, Paul W. Sternberg<sup>2</sup>, Patrick Tan<sup>3,4</sup>, Peter R. Boag<sup>5</sup>, Andreas Hofmann<sup>1,6</sup> & Robin B. Gasser<sup>1</sup>

The blood fluke *Schistosoma haematobium* causes urogenital schistosomiasis, a neglected tropical disease (NTD) that affects more than 110 million people. Treating this disease by targeted or mass administration with a single chemical, praziquantel, carries the risk that drug resistance will develop in this pathogen. Therefore, there is an imperative to search for new drug targets in *S. haematobium* and other schistosomes. In this regard, protein kinases have potential, given their essential roles in biological processes and as targets for drugs already approved by the US Food and Drug Administration (FDA) for use in humans. In this context, we defined here the kinome of *S. haematobium* using a refined bioinformatic pipeline. We classified, curated and annotated predicted kinases, and assessed the developmental transcription profiles of kinase genes. Then, we prioritised a panel of kinases as potential drug targets and inferred chemicals that bind to them using an integrated bioinformatic pipeline. Most kinases of *S. haematobium* are very similar to those of its congener, *S. mansoni*, offering the prospect of designing chemicals that kill both species. Overall, this study provides a global insight into the kinome of *S. haematobium* and should assist the repurposing or discovery of drugs against schistosomiasis.

Schistosomiasis is a neglected tropical disease caused by blood flukes of the genus *Schistosoma* (phylum Platyhelminthes; class Trematoda)<sup>1,2</sup>. The three main species, *Schistosoma haematobium*, *S. mansoni* and *S. japonicum*, affect around 230 million people worldwide<sup>1</sup>. The former two flukes predominate, infecting almost 200 million humans in sub-Saharan Africa alone<sup>3,4</sup>. *S. haematobium* causes the urogenital form of this disease, and *S. mansoni* leads to hepato-intestinal illness<sup>1</sup>. These flukes have a complex life cycle, involving aquatic snails (family Planorbidae) as intermediate hosts. In freshwater, the infective larvae (cercariae) leave the snail and infect the definitive, human host by penetrating skin. Upon penetration, the cercariae lose their tails, and the larvae (schistosomules) migrate through the circulatory system and lung to the portal system, after which they mature and mate. Subsequently, paired adult worms migrate to their site of predilection and start to reproduce. *S. mansoni* adults live mainly in the portal system and/or the mesenteric venules of the small intestine, where they produce eggs that pass through the intestinal wall and are excreted in faeces. *S. haematobium* adults usually inhabit the blood vessels around the urinary bladder and genital system; here, the parasite produces eggs that pass through the bladder wall and are released in urine. Once eggs are released into freshwater, they immediately hatch to release miracidia

<sup>1</sup>Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia. <sup>2</sup>HHMI, Division of Biology, California Institute of Technology, Pasadena, California, USA. <sup>3</sup>Genome Institute of Singapore, Republic of Singapore. <sup>4</sup>Cancer and Stem Cell Biology, Duke-NUS Graduate Medical School, Republic of Singapore. <sup>5</sup>Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Victoria, Australia. <sup>6</sup>Structural Chemistry Program, Eskitis Institute, Griffith University, Brisbane, Australia. Correspondence and requests for materials should be addressed to R.B.G. (email: robinbg@unimelb.edu.au)

(free-living larvae), which then invade a molluscan intermediate host<sup>1</sup>. *S. haematobium* infects snails of the genus *Bulinus*<sup>5</sup>, whereas *S. mansoni* prefers snails of the genus *Biomphalaria*<sup>6</sup>.

Disease in humans is precipitated by eggs that become entrapped in tissues, where they induce a chronic immune-mediated response, followed by granulomatous changes and ensuing fibrosis<sup>1</sup>. Eggs of *S. mansoni* become lodged mainly in the liver and intestinal wall, leading to egg-induced hepatitis, enteritis and/or associated complications<sup>7</sup>. In contrast, *S. haematobium* eggs are deposited mainly in the vasculature of the urinary bladder, ureter and/or genital tract (particularly in female individuals), although they can be disseminated to other sites in the body. Entrapped eggs induce considerable inflammation and subsequent fibrosis and/or calcification of the bladder. In addition, chronic *S. haematobium* infection can increase the risk of secondary bacterial infections<sup>7</sup>, is a predisposing factor for HIV/AIDS<sup>8</sup> and can, together with other factors, induce malignant bladder cancer<sup>9</sup>. As there is no effective vaccine against schistosomiasis, current treatment relies on a single drug, praziquantel<sup>10</sup>. With increased efforts to control this disease by mass treatment, the possibility of praziquantel resistance developing is a serious concern<sup>11,12</sup>. Thus, there is a need for sustained research toward developing alternative chemotherapeutic compounds against schistosomiasis.

Recent research efforts to identify new molecular targets for chemotherapeutic intervention have focused on protein kinases<sup>13,14</sup>, because they are involved in signalling cascades of essential regulatory and developmental processes<sup>15–17</sup>, particular kinase groups have relatively conserved structures<sup>18</sup>, and also because drugs targeting these enzymes in humans have shown particular potential for the treatment of cancers and other diseases<sup>19,20</sup>. Protein kinases are enzymes (transferases) that phosphorylate a substrate by transferring a phosphoryl group from an energy-rich molecule, such as adenosine triphosphate (ATP), to a target protein. This phosphorylation induces a modification of the substrate, leading to changes in conformation and activity<sup>21</sup>. Substrates are phosphorylated at an amino acid residue that has a free hydroxyl group. Kinases can be subdivided into serine/threonine-phosphorylating kinases (STKs), tyrosine-phosphorylating kinases (TKs) and kinases that phosphorylate either of these residues (called ‘dual-specificity’ or ‘hybrid’ kinases). The conserved, catalytic domain of kinases is a protein fold consisting of an amino-terminal lobe comprised of  $\beta$ -strands and a carboxy-terminal lobe that contains  $\alpha$ -helices<sup>22</sup>. A polypeptide linker functions as a hinge and connects the two lobes, allowing for rotation. This lobe structure forms a catalytic cleft for substrate and ATP binding<sup>15,22,23</sup>.

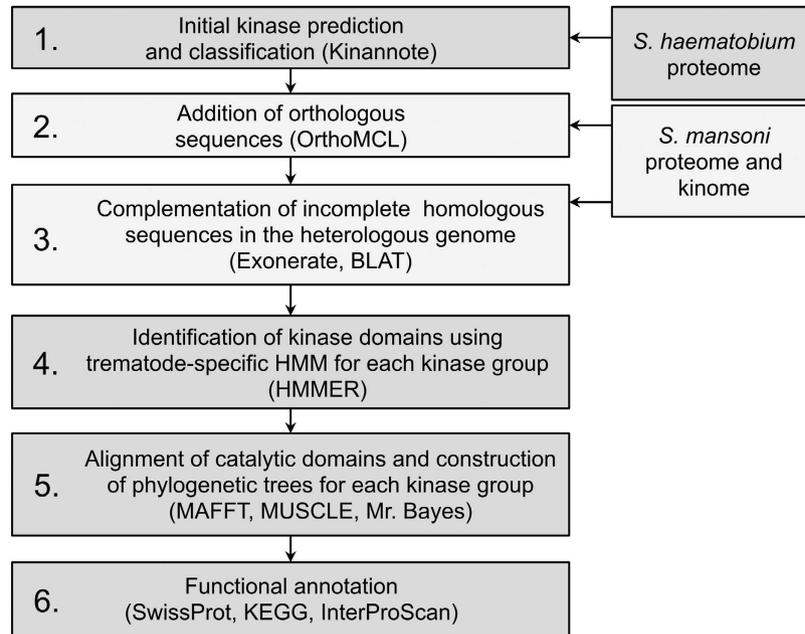
Eukaryotic protein kinases (ePKs) represent the largest class of enzymes that share the same protein kinase-like (PKL) fold<sup>24</sup>. Kinases that have catalytic activity but are not structurally similar to the PKL fold are classified as atypical kinases (aPKs)<sup>15</sup>. Protein kinases can be assigned to groups, families and subfamilies based on sequence similarity in their catalytic domains and the presence of accessory domains. The established classification scheme for kinases (<http://kinase.com/kinbase>)<sup>16</sup> is based on that originally proposed by Hanks and Hunter<sup>23</sup>, and defines nine ePK groups.

Recognising their essential role in a range of regulatory processes and relatively conserved structure and function<sup>25–27</sup>, more than 20 ePKs have been investigated experimentally in *S. mansoni*<sup>13,25</sup>. Some of these kinases have been shown to assume essential functions in the parasite<sup>17,28–31</sup>. For example, the targeting of multiple receptor kinases of *S. mansoni* with a single inhibitor led to a fatal impact on schistosome morphology and physiology<sup>32</sup>. The fact that human protein kinases are involved in cancer and numerous compounds which inhibit these enzymes are available and approved for therapeutic use offers a unique prospect of repurposing such chemicals to schistosomes<sup>25</sup>. In this context, the *in silico* prediction of the kinome of *S. mansoni* provides a basis for the investigation of schistosome kinases as drug targets<sup>33</sup>.

In contrast to the situation for *S. mansoni*, there is no detailed information on the kinome of *S. haematobium* or any other schistosome. Given that *S. haematobium* is the causative agent of schistosomiasis in approximately two thirds of all humans infected by schistosomes and therefore has a substantial socioeconomic impact, in terms of disability-adjusted life years and morbidity<sup>4</sup>, there is a major need to work toward identifying drug targets in *S. haematobium* and designing new treatments<sup>34,35</sup>. In the present study, we defined the kinome of *S. haematobium*. Employing the *S. mansoni* kinome as a reference<sup>33</sup>, we: (i) curated the full complement of predicted kinases of *S. haematobium* using a comparative genomic-phylogenetic approach; (ii) assessed levels of transcription of genes encoding these kinases in the adult and egg stages of *S. haematobium*, and (iii) prioritised a panel of kinases as potential drug targets as well as chemicals inferred to bind to them using an integrated bioinformatic pipeline. We discuss the findings in the context of drug discovery and with regard to the distinctive biologies of *S. haematobium* and *S. mansoni*.

## Results

**The *S. haematobium* kinome.** Here, we employed an integrative bioinformatic pipeline (Fig. 1). First, we predicted 223 kinases in the *S. haematobium* genome, 111 and 93 of which could be assigned to subfamilies and families, respectively; 10 could be assigned exclusively to a group, and nine remained unclassified. Subsequently, we identified 46 additional kinase sequences. Following this curation, the number of unclassified sequences decreased to four, and an improved classification of kinases into subfamilies (n = 134), families (n = 129) and groups (n = 2) was achieved (Fig. 2; Supplementary Table 1).



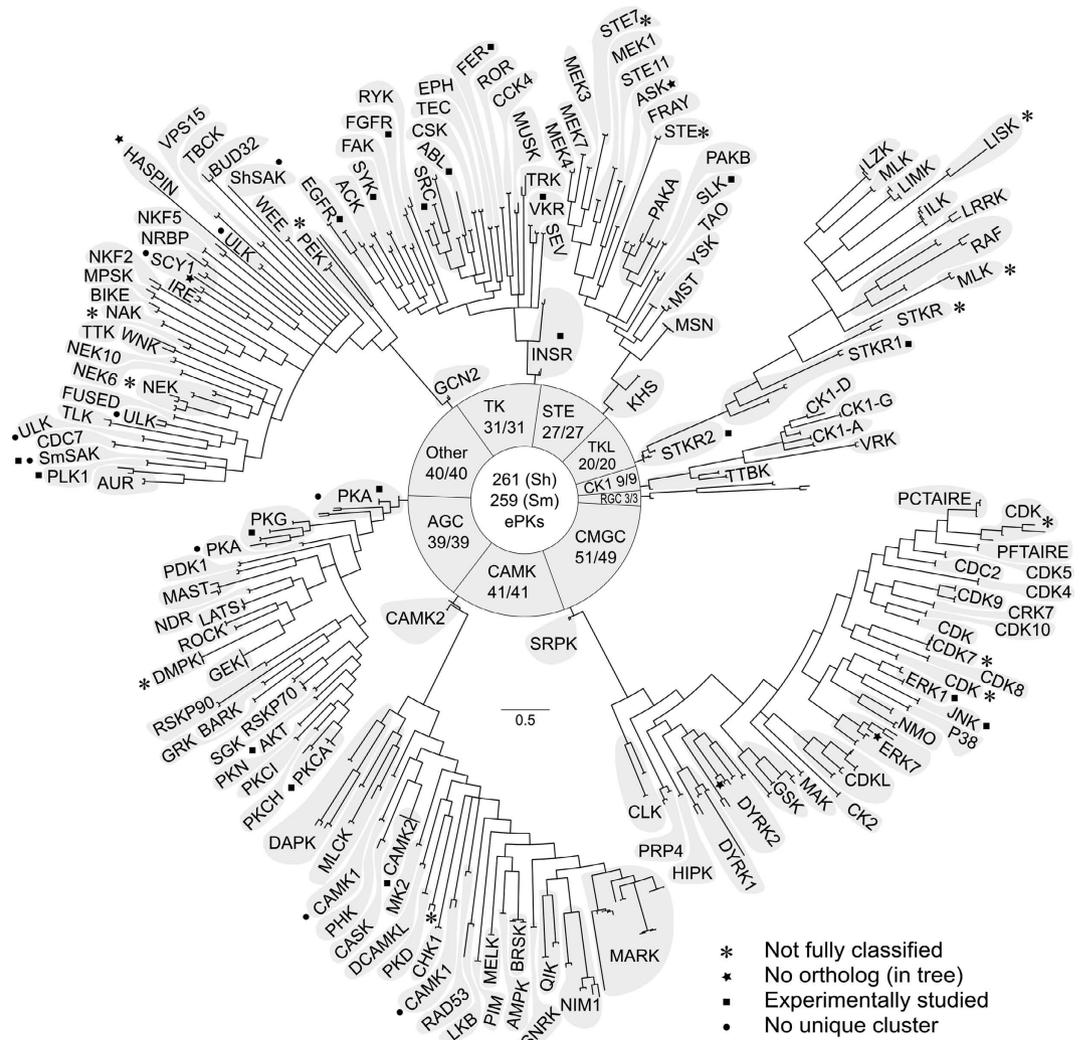
**Figure 1. Bioinformatic pipeline used to characterize and curate kinases in *Schistosoma haematobium*.**

In step 1, we predicted and classified kinases in *S. haematobium*. In steps 2–3, additional sequences were identified employing the proteome<sup>51,88</sup> and kinome<sup>33</sup> inferred from the *S. mansoni* genome; incomplete or missing sequences were complemented using orthologous full-length sequences, which resulted in the final set of predicted kinase sequences. In steps 4 and 5, the catalytic domains in the kinase sequences were identified using trematode-specific HMMs for individual kinase groups, and then aligned (according to group) for subsequent phylogenetic analysis. In step 6, all kinases identified were functionally annotated employing SwissProt, KEGG and InterProScan databases.

Thus, the curated *S. haematobium* genome was inferred to encode 269 kinases, including both ePKs and PKLs.

A total of 261 ePKs representing all nine major kinase groups were identified in *S. haematobium* (Fig. 2; Supplementary Table 1). The largest group represented CMGCs ( $n = 51$ ), including 17 cyclin-dependent kinases (CDKs), four CDK-like kinases (CDKLs), 10 mitogen-activated protein kinases (MAPKs), 11 dual-specificity tyrosine-regulated kinases (DYRKs), four glycogen synthase kinases (GSKs), two CDC-like kinases (CLKs), and one member of each of the families CK2, RCK and SRPK. The second largest group was CAMK, representing 41 kinases including CAMKs, CAMK-related kinases, MARKs and death-associated protein kinases (DAPKs). Only slightly smaller was the ‘Other’ group, which included 40 kinases representing 20 families that do not belong to any of the eight other ePK groups; this group included NEK, AUR (Aurora kinase), BUD32, HASPIN, two Polo-like kinases (PLKs), PEK (pancreatic eIF-2 $\alpha$  kinase), SCY1 and ULK (Unc-51-like kinase). The AGC group represented 39 kinases, including the cyclic nucleotide-dependent kinase families PKA ( $n = 6$ ) and PKG ( $n = 4$ ), and PKCs ( $n = 5$ ), RSKs ( $n = 5$ ) and DMPKs ( $n = 7$ ). Of the 31 members of the TK group, 13 were receptor tyrosine kinases (RTKs), including epidermal growth factor receptors (EGFRs), fibroblast growth factor receptors (FGFRs), insulin receptors (INSRs or IRs) and two Venus kinase receptors (VKRs). The other 18 members were cytoplasmic tyrosine kinases (CTKs) and were assigned to 11 families (ABL, ACK, CSK, FAK, FER, RYK, SEV, SYK, TEC, TRK and SRC). The STE group contained 18 members of the STE20 family (MAP4Ks), two STE11 kinases (MAP3Ks) and six STE7 family members (MAP2Ks). The 20 representatives of the TKL group belonged to the families STKR ( $n = 7$ ), MLK ( $n = 6$ ), RAF ( $n = 3$ ), LRRK ( $n = 1$ ) and LISK ( $n = 3$ ). We also identified nine kinases belonging to the CK1 group, including three members of the Tau tubulin kinase family (TTBK) and one vaccinia-related kinase (VRK). Finally, with only three members, the receptor guanylate cyclases (RGCs) represented the smallest group of ePKs in the *S. haematobium* kinome.

In addition to ePKs, we identified four PKLs: two right open reading frame kinases, *Sh*-RIOK-1 (A\_06019) and *Sh*-RIOK-2 (A\_01816), and two representing the ABC1 family (A\_02560 and A\_01324). The *S. haematobium* genome also encodes four unclassified serine/threonine kinases, to which we assigned the following annotations based on similarity searches against the protein database Swiss-Prot: A\_05753 - Cell cycle serine/threonine-protein kinase CDC5; A\_08069 - Kinase suppressor of Ras 1 (KSR); C\_01296 - Serine/threonine-protein kinase WNK1; *Sh*\_Smp\_017900.1 - Ribosomal protein S6 kinase (RSK).



**Figure 2. Phylogenetic analysis of eukaryotic protein kinases (ePKs) of *Schistosoma haematobium* and *S. mansoni*.** Following the alignment of amino acid sequences representing individual kinase groups, phylogenetic trees were constructed. High resolution figures of individual trees including nodal support values and sequence identifiers are given in Supplementary Figs 2–9.

All remaining kinase sequences ( $n = 265$ ) were assigned to families and/or subfamilies, except for two sequences (A\_03674 and A\_04152) that could be classified only to a group level (i.e. CAMK and STE, respectively). In a phylogenetic analysis, sequence A\_03674 clustered with A\_07692 (predicted PKD kinase), albeit with a low nodal support (61%; Supplementary Fig. 2), and thus could not be assigned with confidence to any particular family. The homolog of sequence A\_04152 (STE family member) in *S. mansoni* (Smp\_146290.1) has been classified previously as a STE7 kinase<sup>33</sup>, but, according to the present analysis, it clustered with a kinase of the STE20 family and FRAY subfamily with 72% nodal support (Supplementary Fig. 6). Thus, sequence A\_04152 was not classified to a family or subfamily level.

For 267 of the 269 kinases defined in *S. haematobium*, orthologs were identified in *S. mansoni* based on a comparative genomic approach and subsequent phylogenetic analyses. For two *S. haematobium* kinase sequences, no ortholog was found, in spite of exhaustive searching of the *S. mansoni* genome (A\_01970; CMGC/MAPK/ERK7 and A\_07508; CMGC/DYRK/DYRK2), suggesting their uniqueness to *S. haematobium*.

A comparison of the kinomes of *S. haematobium* and *S. mansoni* revealed a high overall sequence identity (82–92%), similarity (87–94%) and a relatively conserved length (0–7% difference) between pairs of kinases (Table 1). The degree of sequence similarity among individual kinase groups differed considerably, with kinases from the groups CK1 and RGC, and unclassified and PKL kinases, being, on average, more dissimilar compared with the other groups (Table 1). A pairwise sequence comparison of kinases of *S. haematobium* with human homologs revealed an average sequence similarity ranging from 60.9% (PKL) to 76.3% (CK1) for kinases that could be classified. For unclassified kinase sequences, we observed low sequence identity (35.1% on average) to their closest human homologs.

Kinase group <sup>a</sup>	Length ratio <i>Sh/Sm</i> [SD]	<i>S. mansoni</i> % identity [SD]	<i>S. mansoni</i> % similarity [SD]	<i>H. sapiens</i> % identity [SD]	<i>H. sapiens</i> % similarity [SD]
CMGC	1.00 [0.08]	91.87 [6.10]	94.42 [4.96]	61.04 [8.93]	75.91 [7.15]
CAMK	1.00 [0.08]	87.89 [8.48]	91.48 [6.55]	54.57 [13.91]	71.31 [10.79]
AGC	0.99 [0.19]	91.20 [5.91]	93.78 [4.94]	52.40 [13.22]	68.66 [11.15]
Other	0.98 [0.17]	87.18 [8.46]	90.28 [7.77]	43.54 [12.06]	62.11 [10.87]
TK	0.99 [0.11]	87.66 [8.18]	91.12 [6.92]	45.74 [6.50]	63.68 [6.46]
STE	1.04 [0.22]	90.75 [6.89]	93.21 [6.22]	56.94 [12.14]	72.62 [10.73]
TKL	1.01 [0.08]	88.59 [6.33]	92.01 [5.67]	45.29 [9.95]	63.58 [8.68]
CK1	0.98 [0.14]	85.05 [8.21]	87.85 [7.94]	62.61 [11.20]	76.30 [8.58]
RGC	1.02 [0.19]	85.30 [6.00]	87.53 [6.86]	50.43 [5.36]	66.70 [5.38]
PKL	1.07 [0.12]	88.45 [3.92]	91.53 [4.01]	43.55 [6.34]	60.90 [5.82]
Unclassified	1.06 [0.11]	82.50 [14.84]	88.92 [9.30]	35.12 [7.76]	58.10 [3.93]

**Table 1. Pairwise comparisons of *Schistosoma haematobium* (*Sh*) kinase sequences with orthologs in *S. mansoni* (*Sm*) and human.** Average length ratios, identity and similarity values are indicated. Amino acid sequence conservation between *S. haematobium* and *S. mansoni* was observed for all kinase groups. Predicted sequences had very similar lengths. The comparison with human homologs showed moderate to low identities and similarities. Average and standard deviation [SD] values were calculated based on the number of predicted *S. haematobium* sequences in each kinase group. <sup>a</sup>CMGC = Cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAP kinases), glycogen synthase kinases (GSKs) and CDK-like kinases; CAMK = Ca<sup>2+</sup>/calmodulin-dependent kinases; AGC = Nucleoside-regulated kinases; TK = Tyrosine kinases; STE = MAPK cascade kinases; TKL = Tyrosine kinase-like kinases; CK1 = Casein kinase 1 kinases;

Subsequent phylogenetic analyses of ePKs of both *S. haematobium* and *S. mansoni* supported the orthology found between pairs of kinases of these two species. With the exception of the Polo-like kinase *Sh*-SAK, and representatives of the ULK, SCY1, PKA and CAMK1 families/subfamilies (Fig. 2; Supplementary Figs 2–4), orthologous sequences formed pairs in individual trees (Fig. 2; Supplementary Figs 1–11), consistent with their classification using an approach based on hidden Markov models (HMMs). Seven kinase sequences were excluded from phylogenetic analysis, because the catalytic domain of one or both representatives of the orthologous pair did not match the trematode-specific HMM. Six of these sequences were members of the family SCY1 (A\_01858, Smp\_176440.1 and Sh\_Smp\_156890.1) or HASPIN (Smp\_Sh\_A\_07473, Sh\_Smp\_158950.1 and Smp\_158950.1), which are part of the ‘Other’ kinase group. The seventh sequence (Smp\_Sh\_A\_06810) was a member of the STE group, STE11 family and ASK subfamily.

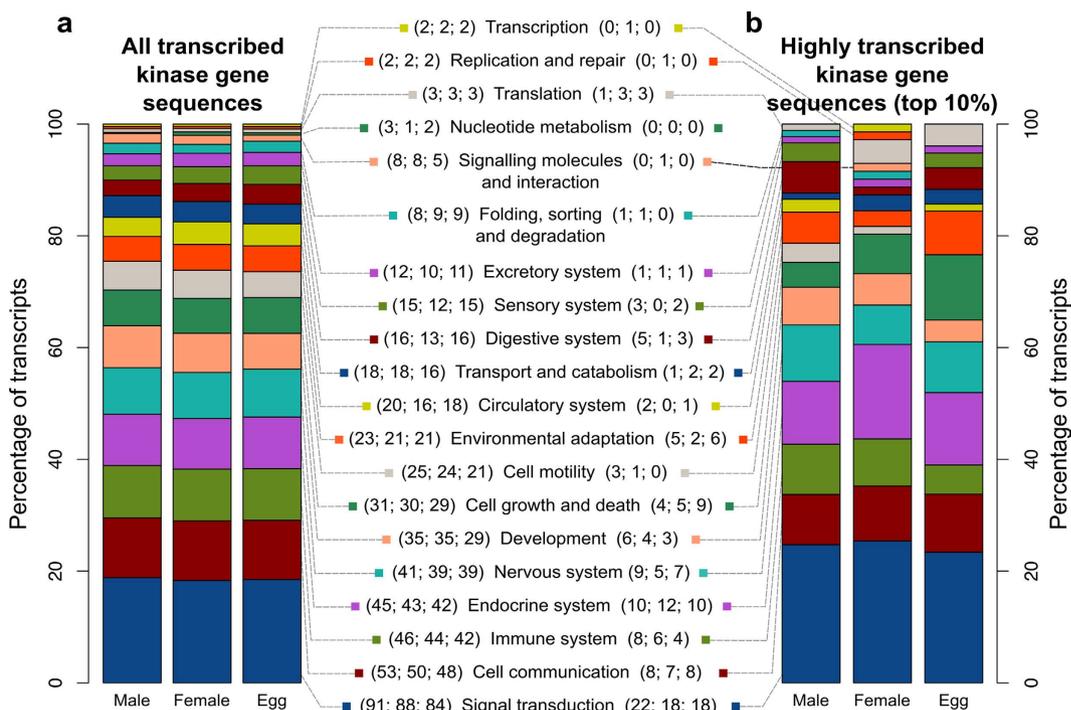
Taken together, the 269 protein kinases of *S. haematobium* and 267 orthologs in *S. mansoni* were shown to represent all nine recognised kinase groups, 88 families and 79 subfamilies. However, we did not detect representatives of 19 kinase families and subfamilies (Table 2) in these two schistosomes (Lophotrochozoa; Protostomia) that are present in members of both the Ecdysozoa (Protostomia, represented by *Caenorhabditis elegans* and *Drosophila melanogaster*) and Deuterostomia (represented by *Homo sapiens*). Finally, we functionally annotated *S. haematobium* kinase sequences identified herein and linked them to 20 conserved functional categories (Fig. 3; Supplementary Table 1). Most kinases were predicted to have functional roles in signal transduction, cell communication, cell growth and the immune and/or nervous systems (Fig. 3; Supplementary Table 1).

**Transcription profiles.** Following the curation and annotation of kinase sequences, we assessed transcription levels of respective genes in different developmental stages and genders of *S. haematobium* (adult male, adult female and egg). Of the 274 sequences encoding kinases identified in *S. haematobium*, 214 were transcribed in all three stages (Fig. 4). By contrast, 13 kinase genes were transcribed exclusively in the male and egg stages, 21 kinase genes were uniquely transcribed in the two adult stages, and one gene was transcribed in the female and egg stages, to the exclusion of the male stage (Fig. 4). One and eight kinase genes were transcribed exclusively in the egg and male stages, respectively. Among the eight male-specific genes were orthologs of the testis-expressed gene 14 (*tex-14*, Sh\_Smp\_131630.1\_p1) and a gene coding for an atrial natriuretic peptide receptor (A\_02682), a kinase belonging to the RGC group that regulates cardiovascular and body fluid homeostasis<sup>36</sup>. For 16 kinase genes, there was no evidence of transcription in any of the life cycle stages studied here (Fig. 4; Supplementary Table 2).

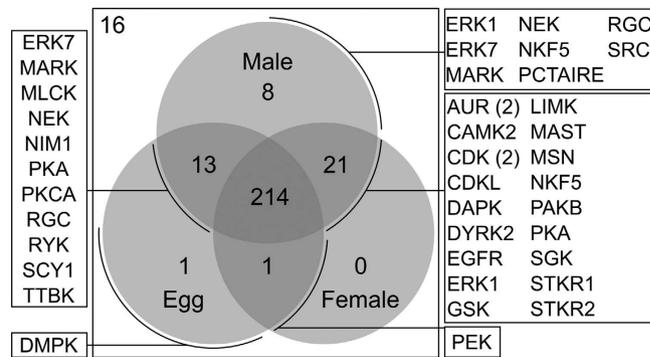
We also assessed transcription levels for the four unclassified *S. haematobium* kinase genes. For the sequence A\_05753, we did not observe transcription in any of the life stages studied; A\_08069 was lowly transcribed in the adult female only (TPM: 0.06) and C\_01296 was moderately transcribed in both adult

Name	Kinase classification
Novel (Nua) kinase family	CAMK/CAMKL/NUAK
MAPK-integrating or -interacting kinase	CAMK/MAPKAPK/MNK
Testis-specific serine/threonine kinase	CAMK/TSSK
RSK-like kinase	AGC/RSKL
Mitogen- and stress-activated protein kinase	AGC/RSK/MSK
RSK-related kinase	AGC/RSKR
Yet another novel kinase	AGC/YANK
Budding uninhibited by benzimidazoles kinase	OTHER/BUB
New kinase family 1	OTHER/NKF1
Anaplastic lymphoma kinase	TK/ALK
Discoidin domain receptor kinase	TK/DDR
IL1 receptor-associated kinase	TKL/IRAK
Serine/threonine-like kinase	STE/STE20/STLK
Eukaryotic elongation factor 2 kinase	ATYPICAL/ALPHA/EEF2K
Bromodomain-containing kinases	ATYPICAL/BRD
Pyruvate dehydrogenase kinase	ATYPICAL/PDHK
Phosphatidylinositol 3 kinase-related kinase	ATYPICAL/PIKK
Right open reading frame kinase 3	ATYPICAL/RIO/RIO3
TATA-binding protein-associated factor 1	ATYPICAL/TAF1

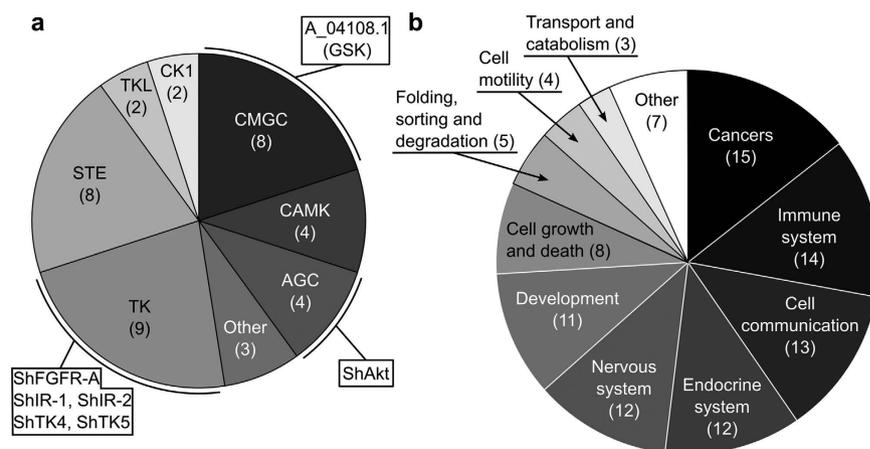
**Table 2.** Kinase families and subfamilies absent from the kinomes of *Schistosoma haematobium* and *S. mansoni* (Lophotrochozoa; Protostomia). Members of these families and subfamilies are found in both Ecdysozoa (Protostomia) and Deuterostomia.



**Figure 3.** Functional annotation and levels of transcription of *Schistosoma haematobium* kinase genes. (a) All kinase genes transcribed in different sexes/developmental stages (male, female and egg). (b) Top 10% of transcribed kinase genes. Proteins inferred from these transcripts were associated with biochemical pathways. Numbers of inferred sequences in the respective functional category are indicated in parentheses for each sex/developmental stage.



**Figure 4.** Venn diagram indicating the number of kinase genes selectively transcribed in the three developmental stages of *Schistosoma haematobium* studied. A total of 214 kinase genes were constitutively transcribed in all three developmental stages. Of the 274 coding regions, 16 were not transcribed. Kinase families/subfamilies assigned to transcribed kinase genes are indicated (boxed).



**Figure 5.** Kinases prioritised as targets in *Schistosoma haematobium* and associated pathways.

(a) Numbers of predicted targets in individual kinase groups. Kinases that have already been investigated or prioritised in *S. mansoni* are indicated. (b) Pathway associations of prioritised targets.

stages (TPM female: 2.64; TPM male: 9.80); Sh\_Smp\_017900.1 was most highly transcribed in the egg stage (TPM: 50.97), but was also transcribed at varying levels in both adult stages (TPM female: 5.32; TPM male: 23.57).

Although most kinase genes were transcribed in all developmental stages of *S. haematobium* (Figs 3a and 4), there were differences in transcription levels, depending on their functional category (Fig. 3b). Notably, almost twice as many genes of kinases associated with cell growth and death were highly transcribed in the egg stage compared with either gender of the adult stage. In addition, kinase genes associated with cell motility were more abundantly transcribed in the male adult. We also found increased levels of transcription for kinase genes associated with environmental adaptation and the sensory system in the egg and male adult compared with the female adult stage.

**Druggable kinases and their prioritisation.** Following the transcriptional analysis, we prioritised *S. haematobium* kinases as potential drug targets. First, we inferred the essentiality of *S. haematobium* kinase genes based on lethal gene knock-down or knock-out phenotypes linked to one-to-one orthologs in *C. elegans*, *D. melanogaster* and/or *Mus musculus* (Supplementary Table 3). In total, 219 of 269 (81%) *S. haematobium* kinases matched orthologs inferred to be associated with lethal phenotypes in at least one of the three organisms (Supplementary Table 3). Of these 219 kinases, 57 mapped (at amino acid level) to unique chokepoints in key biological pathways (Supplementary Table 3). Of these 57 kinases, 40 were predicted to bind chemical ligands listed in Kinase SARfari and DrugBank, 11 of which were present in both databases (Supplementary Tables 4–5). These 40 kinases represented all recognised groups, except RGC, and had human orthologs, some of which related to the nervous system, development and/or cancer (Fig. 5b).

Then, we showed that genes encoding these 40 kinases were transcribed in both adult and egg stages ( $n = 38$ ), and that two (i.e. A\_06570 and A\_07448) were specific to adults (Supplementary Table 2). Amongst them were two casein kinases (A\_08312.1 and Sh\_Smp\_099030.1) with  $>90\%$  sequence similarity to human orthologs; four other kinases in this group (i.e. A\_03569 (FAK), A\_00551 (GCN2), m.56516 (RAF) and A\_03539 (CHK1)) had  $\leq 50\%$  sequence similarity to human counterparts (Supplementary Table 1).

Of the 40 prioritised kinases, tyrosine kinases were the most highly represented group ( $n = 9$ ), including a fibroblast growth factor receptor (*Sh-FGFR-A*), two insulin receptors (*Sh-IR-1* and *Sh-IR-2*) and kinases SYK (*Sh-TK4*) and FYN (*Sh-TK5*), orthologs of which have been experimentally evaluated as drug targets in one or more schistosomes other than *S. haematobium*<sup>13,29,32,37–39</sup>. Two other targets, namely *Sh-Akt* (AGC group) and A\_04108.1 (CMGC group; GSK family), were inferred, both of which have also been predicted to be promising drug targets in *S. mansoni*<sup>40,41</sup> (Fig. 5a).

Taken together, we predicted that all 40 essential kinases represent targets, and therefore interrogated key databases for chemicals. We identified 42 drugs predicted to bind one or more of these targets, 17 of which are already approved by the FDA for the treatment of cancers or other diseases of humans (Table 3). These 17 drugs include four ABL kinase inhibitors (imatinib<sup>42</sup>, dasatinib<sup>43</sup>, bosutinib<sup>44</sup> and ponatinib<sup>45</sup>), one JAK kinase inhibitor (tofacitinib), one GSK3 inhibitor (lithium carbonate), one protein kinase C inhibitor (ingenol mebutate) and 10 other drugs that inhibit multiple (receptor) kinases.

## Discussion

Here, we established an integrated bioinformatic pipeline to identify, classify and curate full-length kinase sequences encoded in the genome of *S. haematobium* for subsequent comparison with orthologs in *S. mansoni* and humans. This workflow enabled high-confidence predictions of anti-schistosome drug targets and compounds, and should be applicable to various schistosome species and, following modification, also to other flatworms as well as roundworms. In the future, we propose to gradually enhance the workflow by integrating tools for the prediction of binding sites of ligands, structural comparisons of prioritised targets and/or comparative analyses of parasite and host kinases into this pipeline.

In most previous studies, the identification of kinase sequences has relied on searches using HMMs from databases such as Pfam<sup>46</sup> or Kinomer<sup>47</sup>, or position-specific scoring matrices (PSSMs)<sup>48</sup>. However, the combination of several of these methods can achieve enhanced predictions and classification compared with a single method. The program Kinannotate uses such a combined approach, thereby increasing sensitivity and precision for kinase identification<sup>49</sup>, and was thus employed by us to produce a draft kinome in the first step of our workflow. Subsequently, an orthology-based approach<sup>50</sup>, using the published kinome<sup>33</sup> and draft genome of *S. mansoni*<sup>51</sup> as a reference, identified pairs of kinase orthologs, which facilitated the improvement of gene models for both schistosomes. This step also increased the number of kinases identified in *S. haematobium* by 17%, and their classification into families/subfamilies by 30%. Independent phylogenetic analyses verified the pairs of orthologs and functional subfamilies. Since the construction of reliable phylogenetic trees requires meticulous alignment of homologous characters, we restricted multiple alignments to the catalytic domains of kinases, because some sequence regions external to the catalytic domain can vary considerably. Phylogenetic trees calculated from these alignments can be used to sub-classify kinases, as sequence divergence in catalytic domains of kinases is recognised to reflect variation of function and/or mode of regulation of protein kinases<sup>23,52</sup>. The boundaries of kinase catalytic domains, such as *Pkinase* (Pfam identifier PF00069) or *Pkinase\_Tyr* (Pfam identifier PF07714), are usually defined by HMMs. However, the sequences used to construct these two HMMs ( $n = 54$  and  $n = 145$ , respectively) did not represent any lophotrochozoans, and thus, might not accurately represent the catalytic kinase domains of trematodes, which are clearly evolutionarily very distinct from those of Ecdysozoa and Deuterostomia<sup>53</sup>. In contrast to the alignment made using these Pfam HMMs, we obtained an improved alignment of homologous characters (with less gaps) by constructing a HMM from high-confidence kinase predictions for four trematode species.

Using the present bioinformatic workflow, we identified 269 full-length kinases that represent the kinome of *S. haematobium*. An assessment of transcription levels revealed transcription of 258 sequences, 214 (79.5%) of which were constitutively transcribed in all developmental stages/sexes studied, indicating essential roles for these kinases in signalling processes throughout the parasite's life cycle. This statement is supported by the constitutive transcription of 83 of the 108 kinase genes (77%) assigned to the functional categories 'signal transduction' and/or 'cell communication'. In contrast, only 11 (10%) kinase genes assigned to these general categories had variable transcription profiles. Although a small number of kinase sequences identified ( $n = 16$ ;  $<6\%$ ) were not transcribed in either the egg or adult stage, they are likely to be transcribed in other developmental stages (including the miracidium, cercaria and/or schistosomulum) not investigated here. The validity of these sequences was supported by pairwise orthologs in *S. mansoni* that are transcribed in the cercarial and/or schistosomule stages<sup>51</sup>.

Sex-specifically transcribed kinase genes were more frequently assigned to specialised functional categories; among them was the male-specifically transcribed testis-expressed gene 14 (*Tex14*, Sh\_Smp\_131630.1), which we hypothesize is critical for chromosome segregation associated with mitosis and meiosis during spermatogenesis. This proposal is supported by findings in mice, showing that *Tex14* is highly expressed during spermatogenesis, and localises to intracellular bridges of germ cells, where it plays an integral role in the establishment and maintenance of male fertility<sup>54,55</sup>. Other evidence

Name or code of compound	Number of target kinases	Indicated for treatment of	Status of approval
AG-13736 (axitinib)	1	Cancer	A
Dasatinib	32	Cancer	A
Pazopanib	30	Cancer	A
Erlotinib	30	Cancer	A
Imatinib	30	Cancer	A
Gefitinib	30	Cancer	A
Sorafenib	32	Cancer	A
Sunitinib	32	Cancer	A
Vandetanib	30	Cancer	A
CP-690550 (tofacitinib)	30	Rheumatoid arthritis, psoriasis, inflammatory bowel disease	A
Bosutinib	4	Cancer	A
Cabozantinib	1	Cancer	A
Ingenol mebutate	1	Cancer, actinic keratosis	A
Ponatinib	3	Cancer	A
Regorafenib	2	Cancer	A
Trametinib	2	Cancer	A
Lithium carbonate	1	Bipolar disorder	A
ABT-869 (linifanib)	31	Cancer	III
Vatalanib	30	Cancer	III
AMG-706 (motesanib)	30	Cancer	III
PD-184352	4	Cancer	II
PHA-739358 (danusertib)	6	Cancer	II
Seliciclib	30	Cancer	II
SNS-032	30	Cancer	I
Fasudil	4	Cerebral vasospasm, pulmonary hypertension	II
Ruboxistaurin	33	Diabetic retinopathy	III
CHEMBL1173486	2	Unknown	N/A
CHEMBL1230122	1	Unknown	N/A
CHEMBL150504	1	Unknown	N/A
AT7519	1	Cancer	II
AZD2171 (cediranib)	1	Cancer	III
CYC116	1	Cancer	I
Ellagic acid	2	Cancer	N/A
XL228	1	Cancer	I
XL518 (cobimetinib)	2	Cancer	III
XL820	1	Cancer	II
XL844	2	Cancer	I
XL880 (foretinib)	1	Cancer	II
XL999	2	Cancer	II
CEP-1347	1	Asthma, Parkinson's disease	III
KC706	1	Rheumatoid arthritis, psoriasis, inflammatory bowel disease	II
TG100801	1	Macular degeneration, diabetic retinopathy	II

**Table 3.** List of prioritised chemical compounds as drug candidates against *Schistosoma haematobium*.

For each compound, the number of target kinases, its indicated therapeutic use(s) and the status of FDA approval for use in humans are given (A = approved; I, II or III = phase of clinical trial). Additional information and chemical structures are given in Supplementary Tables 4–6.

from a study of human cells lines shows that TEX14 is regulated by the kinase Plk-1 and is crucial for kinetochore-microtubule attachment during mitosis<sup>56</sup>.

A second gene encoding a protein kinase R (PKR)-like endoplasmic reticulum kinase (PERK; A\_03220) was transcribed exclusively in female and egg stages of *S. haematobium*. The human ortholog of this kinase phosphorylates the eukaryotic translation initiation factor 2 alpha (eIF2 $\alpha$ ) and mediates the response to endoplasmic reticulum (ER) stress (represented by an accumulation of misfolded or unfolded proteins in the ER) which, among other factors, is induced by glucose deprivation<sup>57,58</sup> and/or an excessive requirement for proteins<sup>59</sup>. The transcription of this additional, stress-mitigating kinase in eggs and female worms might thus be a mechanism to cope with increased ER stress due to the energy- and protein-demanding processes of reproduction, which are sustained by glucose metabolism. This specific transcription might also relate to stress on female worms, induced by separating them from their male partner (on which they rely, in terms of nutrient supply, such as sugar uptake from the host)<sup>60</sup> prior to RNA-sequencing.

A third kinase gene encoding a myotonic dystrophy protein kinase (A\_05067) of the DMPK family was transcribed exclusively in the egg stage of *S. haematobium*. Since different muscle types are already established in the miracidium within the egg, and a transformation of these muscle structures takes place during metamorphosis from sporocysts to cercariae<sup>61</sup>, we propose that this kinase-encoding gene is specifically transcribed in the miracidium in the egg, and is involved in muscle development and/or locomotion/motility. Evidence from other invertebrates, such as *D. melanogaster*, shows that DMPKs are involved in establishing correct muscle morphology and functionality in third instar larvae<sup>62</sup>. This aspect warrants further exploration when RNA-sequencing data for the miracidium stage of *S. haematobium* become available.

Comparative analysis showed that the *S. haematobium* kinome contains all recognised eukaryotic kinase groups, including 79 of the 144 (55%) subfamilies found in other metazoans studied<sup>16,63</sup>. The *S. haematobium* kinome has approximately half of the 518 kinases found in humans<sup>15</sup> and has a similar number to that (n = 438) of the *C. elegans* kinome, to the exclusion of known specific expansions in this free-living nematode<sup>16,63</sup>. Nonetheless, we did not detect any members of 19 kinase families/subfamilies present in *C. elegans*, *D. melanogaster* or *H. sapiens*. The lack of evidence for kinases of these families/subfamilies, including RIO3 (which has been lost from numerous flatworms<sup>64</sup>), suggests their absence from schistosomes or a substantial diversification of their sequences that precluded their identification. Since there are presently no curated kinomes for flatworms other than *S. haematobium* and *S. mansoni*, it is not known whether such kinase families or subfamilies have been lost from all lophotrochozoans or only from schistosomes during evolution. A preliminary exploration of the flatworms *Clonorchis sinensis*, *Opisthorchis viverrini* and *Fasciola hepatica* (Stroehlein *et al.*, unpublished) suggests that these families and subfamilies (except the PIKK family) are absent from lophotrochozoans. Future studies should focus on defining and curating the kinomes of a range of socioeconomically important parasitic flatworms and roundworms (nematodes), in order to undertake detailed comparative analyses, explore kinome evolution and investigate contractions and expansions of particular kinase groups in relation to worm phylogeny as well as biology.

The global comparison of the kinomes of *S. haematobium* and its close relative, *S. mansoni*, did not detect any major expansions or contractions in kinase groups, families or subfamilies, but did reveal two kinase genes of the CMGC group (ERK7 and DYRK2 subfamilies) that are present exclusively in the former species. Given the quality of the draft genome and transcriptome of *S. mansoni*, there is only a remote possibility that these two genes were not detected. It is more plausible that they are indeed uniquely present in *S. haematobium* and encode kinases that may relate indirectly to this pathogen's unique biology and site predilection in the human host. Published evidence indicating that ERKs are involved in parasite-host interactions<sup>65,66</sup> supports this hypothesis. Although very little is known about the function of the second *S. haematobium*-specific kinase (DYRK2), in human and murine cell lines, a DYRK homolog interacts with the MAPK kinase MKK3 (an upstream activator of p38), which is involved in a growth factor-mediated signalling pathway<sup>67</sup>. The fact that both *S. haematobium*-specific kinases are part of receptor-activated signalling pathways advocates a role in pathogen-host interactions, as has been suggested previously for other receptor kinase pathways<sup>68,69</sup>.

Despite this difference of two kinases, the comparison of the kinomes of *S. haematobium* and *S. mansoni* showed a relatively high level of conservation of kinase sequences. Although such conservation has been reported previously for small numbers of kinases<sup>32,38,70</sup>, here we report the first global comparison of these kinomes. The conservation between the kinomes of the two most medically important species of schistosomes is considered to provide opportunities for the repurposing of existing, safe drugs against both species<sup>25</sup>. Thus, we focused on 40 *S. haematobium* kinase genes with (relatively) conserved orthologs in *S. mansoni* and *S. japonicum* (not shown) as well as human, whose gene products are inferred to be essential and to bind drugs available for treating human diseases.

A functional annotation of these 40 kinases showed that 37.5% (n = 15) were linked to human orthologs that are involved in cancer pathways, and a similar number of kinases (n = 14; 35%) were linked to roles in the immune system (Fig. 5b). Based on these findings, we suggest that associated anti-cancer/anti-inflammatory compounds should now be assessed as to their ability to disrupt normal schistosome growth, development and/or viability *in vitro*. In this context, a recent study has shown that blood components (such as serum albumin and  $\alpha$ -1 acid glycoprotein) impede the deleterious effect of the drug

imatinib on schistosomes *in vitro*, which should be considered in the experimental design of *in vitro* or *in vivo* experiments<sup>71</sup>.

A list of compounds (Table 3) revealed promising candidates for repurposing as schistosome kinase inhibitors. Many of these compounds have been predicted to target multiple kinases (targeted poly-pharmacology), a property that can increase the deleterious effect of a drug, thereby overcoming limited efficacy (due to redundancies in signalling pathways) associated with some single-targeted drugs<sup>72,73</sup>. Among the selected compounds were the anti-cancer drugs imatinib and dasatinib, the latter of which is assumed to target the Src/Fyn kinase SmTK5 in *S. mansoni*<sup>13</sup>. The orthologous kinase in *S. haematobium* (*Sh*-TK5) is one of the 40 prioritised targets in this study. Other selected targets of particular interest (Fig. 5a) include a Syk kinase (*Sh*-TK4), four receptor kinases (*Sh*-IR1, *Sh*-IR2, *Sh*-FGFR-A and B\_00871), two members of the AGC group (*Sh*-Akt and A\_01385) and a GSK3 kinase (A\_04108.1). These kinases have either already been computationally predicted as drug targets in *S. mansoni*, or there is some experimental evidence indicating that orthologs in *S. mansoni* are essential and/or can be inhibited *in vitro*<sup>13,27,29,32,37,38,40,41,65,68</sup>, which lends additional support to our predictions. Furthermore, we predicted 32 additional kinases as potential targets for which no experimental information is yet available for schistosomes, including a TTK kinase (*Sh*\_Smp\_171610.1) and an eIF2 $\alpha$  kinase ortholog (A\_00551). *Sh*\_Smp\_171610.1 is an ortholog of a human kinetochore kinase, also known as Mps1 (Monopolar spindle 1), which plays an essential role in the spindle assembly checkpoint (SAC) pathway<sup>74</sup>. The prioritised eIF2 $\alpha$  kinase ortholog is involved in mediating stress-response pathways, and several members of this kinase family are essential in *Plasmodium falciparum* (malaria parasite)<sup>75</sup>. Taken together, the high sequence similarity between schistosome kinases and the availability of kinase inhibitors for human orthologs offer great prospect with regard to the development of new anti-schistosome drugs.

In addition to the conserved kinase complement, there is also considerable merit in exploring selective kinase targets, namely those that are specific to schistosomes but absent from the mammalian host. For instance, the two genes encoding VKRs are specific to schistosomes and other Protostomia<sup>32,76</sup>, but absent from humans. Some functional studies of *S. mansoni* have shown that the compound trypanothin AG1024 kills schistosomula and adults *in vitro*<sup>32,77</sup> by targeting schistosome VKRs and IRs. Given the sequence conservation of VKRs and IRs between *S. haematobium* and *S. mansoni* (97.3% and 93.8% similarity, respectively), this compound is likely to also kill the former species. In the context of identifying further schistosome-specific targets, four pairs of unclassified schistosome kinases identified here (Supplementary Table 1) were of interest, as they exhibited substantially lower sequence similarity to their human orthologs compared with *S. mansoni* orthologs. Three of these kinase-encoding genes were transcribed at varying levels in at least one of the sexes of the adult stage. We suggest that these results might assist in designing inhibitors for schistosomes, particularly if the premise is to target less conserved structural regions in a kinase outside of the conserved catalytic domain. This hypothesis warrants testing.

The curated set of kinases for *S. haematobium* as well as for its close relative, *S. mansoni*, might provide a stepping stone to fundamental studies of the biology of selected kinases in these worms. For instance, gene knockdown experiments by double-stranded RNA interference (RNAi)<sup>78</sup> could be conducted on adult worms to validate the essentiality of subsets of kinases as drug targets in schistosomes. Combined with transcriptomic, proteomic and metabolomic investigations<sup>79–81</sup> of treated *versus* untreated schistosomes, such studies could provide insights into the biological (e.g., signalling) pathways affected in the schistosome and also verify the specific knockdown of kinase genes and gene products. Moreover, in a similar manner, chemical knockdown experiments could confirm the specificity of the predicted and prioritised ligands *in vitro*<sup>82</sup>. Concordance between RNAi and chemical knockdown results would then provide confidence regarding the bioinformatic drug target/drug predictions made. Subsequently, compounds for which one or multiple targets have been validated and that have shown efficacy *in vitro* could then be investigated further in a hit-to-lead phase. At this point, chemical analogs could be produced to optimise target selectivity and minimize side effects on the host organism. Selected chemicals with specific binding to a kinase target but with limited selectivity (e.g., because of activity in mammalian host cells) might still serve as probes<sup>14</sup> to explore kinase biology in the parasite.

In conclusion, we believe that the present bioinformatic investigation represents a step forward in the characterisation and curation of worm kinomes. The concordance in results between *S. mansoni* and *S. haematobium* (Fig. 2; Supplementary Table 1) as well as known lethal/adverse effects of some inhibitors against *S. mansoni* kinases<sup>13,27,29,32,37,38,40,41,65,68</sup> suggest that some of our target and drug predictions are promising. However, we acknowledge that the prediction of drug targets and associated ligands represents a humble beginning to an often long and challenging route to validate new chemical entities (NCEs), to assess them in a preclinical context by administration, distribution, metabolism, excretion and toxicity (ADMET) testing<sup>83–85</sup>, and, *via* clinical trials (phases I–III; <http://www.phrma.org/innovation/clinical-trials>)<sup>86</sup>, to develop one or more safe, effective and specific anti-schistosomal drugs. We hope that our bioinformatic pipeline will assist, at least in part, at the very beginning of this long and expensive discovery and development process.

## Methods

**Defining the *S. haematobium* kinome.** We predicted, curated and annotated the protein kinase complement encoded in the published draft genome<sup>87</sup> using an integrated bioinformatic pipeline in six steps (Fig. 1):

- (1) First, we identified ePKs and PKLs of *S. haematobium* using the program Kinannotate<sup>49</sup> employing the *-m* (metazoan) option. Predicted kinase sequences were then classified according to group, family and/or subfamily<sup>16,63</sup>. Sequences that could not be unequivocally classified using this approach were retained for subsequent curation.
- (2) Orthologous kinase sequences from both *S. haematobium* and *S. mansoni* were predicted by pairwise sequence comparison using the program OrthoMCL<sup>50</sup>, employing publicly accessible (SchistoDB v.3.0; <http://schistodb.net/schisto/> and GeneDB v.5.2; <http://www.genedb.org/>) genomic and transcriptomic datasets<sup>51,87,88</sup>. Amino acid sequences that grouped with classified kinases, but were not predicted to be kinases using Kinannotate, were added to a kinase group, family or subfamily based on their respective orthologous sequence (in the heterologous species) and included in subsequent analyses.
- (3) Then, we exhaustively searched all of the genomic and transcriptomic data available for *S. haematobium* and *S. mansoni*, to be able to complement any incomplete sequences and also to retrieve kinase-encoding sequences that had not been predicted previously for either or both schistosome species. If a full-length ortholog could not be inferred for the heterologous species, the kinase amino acid sequence was aligned to the genomic scaffold coding for the incomplete gene using the program BLAT<sup>89</sup>. This genomic region was then exhaustively searched for a full-length orthologous coding domain using the program Exonerate<sup>90</sup> employing the multi-pass suboptimal alignment algorithm and the protein2genome:bestfit model. Refined gene predictions and protein translations were named according to their ortholog identifier (e.g., Sh\_Smp\_123456.1 and Smp\_A\_12345).
- (4) To increase the sensitivity of identification of kinase domains of schistosomes, we constructed HMMs for individual kinase groups based on the catalytic domains of high-confidence trematode kinase sequences (assigned to a subfamily by Kinannotate) using the program HMMER v.3.1b1 (<http://hmmer.janelia.org/>). These HMMs were constructed using the inferred proteomic datasets of *S. japonicum*, *C. sinensis*, *O. viverrini* and *F. hepatica*<sup>91–94</sup>, and were then employed to query kinase sequences of individual groups of *S. haematobium* and *S. mansoni* and to identify catalytic kinase domains.
- (5) The catalytic domain sequences of all predicted kinases representing individual groups were aligned using the program MAFFT v.6.864b, employing the L-INS-i option<sup>95</sup>. Alignments were improved using the program MUSCLE v.3.7 (-refine option)<sup>96</sup> and by subsequent manual adjustment, to optimise the alignment of homologous characters. The aligned sequences were then subjected to Bayesian inference (BI) analysis in the program MrBayes v.3.2.2 (ref. 97). Posterior probabilities (pp) were calculated, as recommended, using a mixture of models with fixed rate matrices, generating 1,000,000 trees and sampling every 100th tree. The initial 25% of trees were discarded as burn-in, and the others were used to construct a majority rule tree. Phylogenetic trees were drawn using the program FigTree v.1.4.1 (<http://tree.bio.ed.ac.uk/software/figtree/>).
- (6) Curated kinase sequences were functionally annotated by searching the databases Swiss-Prot (database release 01/2014)<sup>98</sup> and KEGG BRITE (database release 03/2014)<sup>99</sup> using BLASTP v.2.2.28+ (ref. 100) and an e-value cut-off of  $10^{-05}$ . Pfam domains and PANTHER families were predicted using the program InterProScan v.5–44.0 (ref. 101). In addition, sequence identities and similarities to *S. mansoni* and human kinase homologs (sequences accessed from KinBase, <http://kinase.com/kinbase/FastaFiles/>) were determined for *S. haematobium* kinases by pairwise comparison using the program EMBOSS Matcher v.6.3.1 (ref. 102).

**Transcription analysis.** We assessed transcription in male and female adults as well as eggs of *S. haematobium* using publicly available RNA-seq data<sup>87</sup>. Data were filtered using the program Trimmomatic<sup>103</sup> and aligned to the final sequences encoding kinases using Bowtie v.2.1.0 (ref. 104). Levels of transcription (numbers of transcripts per million, TPMs) were calculated using the software package RSEM v.1.2.11 (ref. 105). Kinase genes were considered as transcribed if at least 5 read pairs mapped to their coding regions and they had a TPM of  $>0$ . For each kinase gene, a relative measure of transcription was inferred by ranking individual genes from *S. haematobium* by their TPM values. The top and bottom 10% of transcribed genes were defined as being highly and lowly transcribed, respectively.

**Drug target prediction and prioritisation.** To assess the druggability of individual predicted kinases and to prioritise them as potential targets in *S. haematobium*, essentiality was inferred by selecting *S. haematobium* proteins homologous (BLASTP; e-value  $\leq 10^{-5}$ ) to *C. elegans*, *D. melanogaster* and/or *M. musculus* kinases with a lethal phenotype upon gene perturbation - listed in WormBase<sup>106</sup>, FlyBase<sup>107</sup> and MGI<sup>108</sup>. Essential kinases were considered to represent metabolic chokepoints if only one gene was assigned to one KEGG orthologous gene (KO) term for a KEGG pathway. These kinases were then matched to homologous kinase sequences in the databases Kinase SARfari<sup>109</sup> and DrugBank v.3.0 (ref. 110) using PSI-BLAST v.2.2.26+ employing an e-value cut-off of  $10^{-30}$  (ref. 111). If both query and target sequence had the same kinase classification (using Kinannotate), the sequence in the database had one or more ligands that met the Lipinski rule-of-five<sup>112</sup> and was flagged as “medicinal chemistry friendly”, salient information on associated ligands (chemicals or small molecules) was extracted from the two databases and used to assess the druggability of the target. Prioritised kinases predicted to bind compounds approved by the FDA for use in humans or assessed in clinical trials, as indicated in Kinase

SARfari (<https://www.ebi.ac.uk/chembl/sarfari/kinasesarfari>), were considered to have potential as drug targets. Kinases with entries in DrugBank were prioritised as drug targets if at least one associated small molecule (with a description of its properties) was found in this database.

## References

- Colley, D. G., Bustinduy, A. L., Secor, W. E. & King, C. H. Human schistosomiasis. *Lancet* **383**, 2253–2264 (2014).
- World Health Organization. Research priorities for helminth infections: technical report of the TDR disease reference group on helminth infections. *WHO Technical report series; no. 972*. WHO Press (2012).
- Rollinson, D. *et al.* Time to set the agenda for schistosomiasis elimination. *Acta Trop.* **128**, 423–440 (2013).
- van der Werf, M. J. *et al.* Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Trop.* **86**, 125–139 (2003).
- Rollinson, D., Stothard, J. R. & Southgate, V. R. Interactions between intermediate snail hosts of the genus *Bulinus* and schistosomes of the *Schistosoma haematobium* group. *Parasitology* **123**, 245–260 (2001).
- Morgan, J. A., Dejong, R. J., Snyder, S. D., Mkoji, G. M. & Loker, E. S. *Schistosoma mansoni* and *Biomphalaria*: past history and future trends. *Parasitology* **123**, 211–228 (2001).
- Burke, M. L. *et al.* Immunopathogenesis of human schistosomiasis. *Parasite Immunol.* **31**, 163–176 (2009).
- Kjetland, E. F. *et al.* Association between genital schistosomiasis and HIV in rural Zimbabwean women. *AIDS* **20**, 593–600 (2006).
- Palumbo, E. Association between schistosomiasis and cancer: a review. *Infect. Dis. Clin. Pract.* **15**, 145–148 (2007).
- 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).
- Greenberg, R. M. New approaches for understanding mechanisms of drug resistance in schistosomes. *Parasitology* **140**, 1534–1546 (2013).
- Chai, J. Y. Praziquantel treatment in trematode and cestode infections: an update. *Infect. Chemother.* **45**, 32–43 (2013).
- Beckmann, S., Leutner, S., Gouignard, N., Dissous, C. & Grevelding, C. G. Protein kinases as potential targets for novel anti-schistosomal strategies. *Curr. Pharm. Des.* **18**, 3579–3594 (2012).
- Knapp, S. *et al.* A public-private partnership to unlock the untargeted kinome. *Nat. Chem. Biol.* **9**, 3–6 (2013).
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
- Manning, G. Genomic overview of protein kinases. *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.60.1 (2005).
- de Saram, P. S. *et al.* Functional mapping of protein kinase A reveals its importance in adult *Schistosoma mansoni* motor activity. *PLoS Negl. Trop. Dis.* **7**, e1988 (2013).
- Hanks, S. K. Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biol.* **4**, 111 (2003).
- Cohen, P. Protein kinases - the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* **1**, 309–315 (2002).
- Eglen, R. M. & Reisine, T. The current status of drug discovery against the human kinome. *Assay Drug Dev. Technol.* **7**, 22–43 (2009).
- Cohen, P. The regulation of protein function by multisite phosphorylation - a 25 year update. *Trends Biochem. Sci.* **25**, 596–601 (2000).
- Ubersax, J. A. & Ferrell, J. E., Jr. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8**, 530–541 (2007).
- Hanks, S. K. & Hunter, T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9**, 576–596 (1995).
- Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C. & Manning, G. Structural and functional diversity of the microbial kinome. *PLoS Biol.* **5**, e17 (2007).
- Dissous, C. & Grevelding, C. G. Piggy-backing the concept of cancer drugs for schistosomiasis treatment: a tangible perspective? *Trends Parasitol.* **27**, 59–66 (2011).
- Dissous, C. *et al.* Receptor tyrosine kinase signaling and drug targeting in schistosomes in *Protein Phosphorylation in Parasites* (eds Doerig, C., Spaeth, G. & Wiese, M.) 337–356 (Wiley-Blackwell, 2013).
- Morel, M., Vanderstraete, M., Hahnel, S., Grevelding, C. G. & Dissous, C. Receptor tyrosine kinases and schistosome reproduction: new targets for chemotherapy. *Front. Genet.* **5**, 238 (2014).
- Kapp, K. *et al.* The *Schistosoma mansoni* Src kinase TK3 is expressed in the gonads and likely involved in cytoskeletal organization. *Mol. Biochem. Parasitol.* **138**, 171–182 (2004).
- Beckmann, S., Buro, C., Dissous, C., Hirzmann, J. & Grevelding, C. G. The Syk kinase SmTK4 of *Schistosoma mansoni* is involved in the regulation of spermatogenesis and oogenesis. *PLoS Pathog.* **6**, e1000769 (2010).
- Andrade, L. F. *et al.* Regulation of *Schistosoma mansoni* development and reproduction by the mitogen-activated protein kinase signaling pathway. *PLoS Negl. Trop. Dis.* **8**, e2949 (2014).
- Swierczewski, B. E. & Davies, S. J. A schistosome cAMP-dependent protein kinase catalytic subunit is essential for parasite viability. *PLoS Negl. Trop. Dis.* **3**, e505 (2009).
- Vanderstraete, M. *et al.* Dual targeting of insulin and venus kinase receptors of *Schistosoma mansoni* for novel anti-schistosome therapy. *PLoS Negl. Trop. Dis.* **7**, e2226 (2013).
- Andrade, L. F. *et al.* Eukaryotic protein kinases (ePKs) of the helminth parasite *Schistosoma mansoni*. *BMC Genomics* **12**, 215 (2011).
- Brindley, P. J. & Hotez, P. J. Break out: urogenital schistosomiasis and *Schistosoma haematobium* infection in the post-genomic era. *PLoS Negl. Trop. Dis.* **7**, e1961 (2013).
- Rollinson, D. A wake up call for urinary schistosomiasis: reconciling research effort with public health importance. *Parasitology* **136**, 1593–1610 (2009).
- Takei, Y. Structural and functional evolution of the natriuretic peptide system in vertebrates. *Int. Rev. Cytol.* **194**, 1–66 (2000).
- Hahnel, S. *et al.* Gonad RNA-specific qRT-PCR analyses identify genes with potential functions in schistosome reproduction such as SmFz1 and SmFGFRs. *Front. Genet.* **5**, 170 (2014).
- You, H. *et al.* Cloning and characterisation of *Schistosoma japonicum* insulin receptors. *PLoS One* **5**, e9868 (2010).
- Kapp, K., Schussler, P., Kunz, W. & Grevelding, C. G. Identification, isolation and characterization of a Fyn-like tyrosine kinase from *Schistosoma mansoni*. *Parasitology* **122**, 317–327 (2001).
- Morel, M. *et al.* Compound library screening identified Akt/PKB kinase pathway inhibitors as potential key molecules for the development of new chemotherapeutics against schistosomiasis. *Int. J. Parasitol. Drugs Drug Resist.* **4**, 256–266 (2014).
- Caffrey, C. R. *et al.* A comparative Caffrey genomics strategy to predict potential drug targets in the metazoan pathogen, *Schistosoma mansoni*. *PLoS One* **4**, e4413 (2009).
- Imatinib monograph*. Available at <http://www.drugs.com/monograph/imatinib-mesylate.html> (Accessed: 2nd April 2015).

43. *Dasatinib monograph*. Available at <http://www.drugs.com/monograph/dasatinib.html> (Accessed: 2nd April 2015).
44. *Bosutinib monograph*. Available at <http://www.drugs.com/monograph/bosutinib.html> (Accessed: 2nd April 2015).
45. *Ponatinib monograph*. Available at <http://www.drugs.com/monograph/ponatinib.html> (Accessed: 2nd April 2015).
46. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **28**, 405–420 (1997).
47. Martin, D. M., Miranda-Saavedra, D. & Barton, G. J. Kinomer v. 1.0: a database of systematically classified eukaryotic protein kinases. *Nucleic Acids Res.* **37**, D244–D250 (2009).
48. Marchler-Bauer, A. *et al.* CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**, D348–D352 (2013).
49. Goldberg, J. M. *et al.* Kinanote, a computer program to identify and classify members of the eukaryotic protein kinase superfamily. *Bioinformatics* **29**, 2387–2394 (2013).
50. Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
51. Protasio, A. V. *et al.* A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **6**, e1455 (2012).
52. Hanks, S. K., Quinn, A. M. & Hunter, T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52 (1988).
53. Mallatt, J., Craig, C. W. & Yoder, M. J. Nearly complete rRNA genes from 371 Animalia: updated structure-based alignment and detailed phylogenetic analysis. *Mol. Phylogenet. Evol.* **64**, 603–617 (2012).
54. Wu, M. H. *et al.* Sequence and expression of testis-expressed gene 14 (*Tex14*): a gene encoding a protein kinase preferentially expressed during spermatogenesis. *Gene Expr. Patterns* **3**, 231–236 (2003).
55. Greenbaum, M. P. *et al.* TEX14 is essential for intercellular bridges and fertility in male mice. *Proc. Natl. Acad. Sci. USA* **103**, 4982–4987 (2006).
56. Mondal, G., Ohashi, A., Yang, L., Rowley, M. & Couch, F. J. Tex14, a Plk1-regulated protein, is required for kinetochore-microtubule attachment and regulation of the spindle assembly checkpoint. *Mol. Cell* **45**, 680–695 (2012).
57. Xu, C., Bailly-Maitre, B. & Reed, J. C. Endoplasmic reticulum stress: cell life and death decisions. *J. Clin. Invest.* **115**, 2656–2664 (2005).
58. Badiola, N. *et al.* Induction of ER stress in response to oxygen-glucose deprivation of cortical cultures involves the activation of the PERK and IRE-1 pathways and of caspase-12. *Cell Death Dis.* **2**, e149 (2011).
59. Osowski, C. M. & Urano, F. Measuring ER stress and the unfolded protein response using mammalian tissue culture system. *Methods Enzymol.* **490**, 71–92 (2011).
60. Cornford, E. M. & Fitzpatrick, A. M. The mechanism and rate of glucose transfer from male to female schistosomes. *Mol. Biochem. Parasitol.* **17**, 131–141 (1985).
61. Bahia, D. *et al.* The distribution of motor proteins in the muscles and flame cells of the *Schistosoma mansoni* miracidium and primary sporocyst. *Parasitology* **133**, 321–329 (2006).
62. Picchio, L., Plantie, E., Renaud, Y., Poovthumkadavil, P. & Jagla, K. Novel *Drosophila* model of myotonic dystrophy type 1: phenotypic characterization and genome-wide view of altered gene expression. *Hum. Mol. Genet.* **22**, 2795–2810 (2013).
63. Manning, G., Plowman, G. D., Hunter, T. & Sudarsanam, S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**, 514–520 (2002).
64. Breugelmans, B. *et al.* Flatworms have lost the right open reading frame kinase 3 gene during evolution. *Sci. Rep.* **5**, 9417 (2015).
65. Ressurreição, M. *et al.* Protein kinase C and extracellular signal-regulated kinase regulate movement, attachment, pairing and egg release in *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **8**, e2924 (2014).
66. Vicogne, J. *et al.* Conservation of epidermal growth factor receptor function in the human parasitic helminth *Schistosoma mansoni*. *J. Biol. Chem.* **279**, 37407–37414 (2004).
67. Lim, S., Jin, K. & Friedman, E. Mirk protein kinase is activated by MKK3 and functions as a transcriptional activator of HNF1 $\alpha$ . *J. Biol. Chem.* **277**, 25040–25046 (2002).
68. Ahier, A., Khayath, N., Vicogne, J. & Dissous, C. Insulin receptors and glucose uptake in the human parasite *Schistosoma mansoni*. *Parasite* **15**, 573–579 (2008).
69. LoVerde, P. T., Andrade, L. F. & Oliveira, G. Signal transduction regulates schistosome reproductive biology. *Curr. Opin. Microbiol.* **12**, 422–428 (2009).
70. Swierczewski, B. E. & Davies, S. J. Conservation of protein kinase A catalytic subunit sequences in the schistosome pathogens of humans. *Exp. Parasitol.* **125**, 156–160 (2010).
71. Beckmann, S. *et al.* Serum albumin and alpha-1 acid glycoprotein impede the killing of *Schistosoma mansoni* by the tyrosine kinase inhibitor Imatinib. *Int. J. Parasitol. Drugs Drug Resist.* **4**, 287–295 (2014).
72. Morphy, R. Selectively nonselective kinase inhibition: striking the right balance. *J. Med. Chem.* **53**, 1413–1437 (2010).
73. Anighoro, A., Bajorath, J. & Rastelli, G. Polypharmacology: challenges and opportunities in drug discovery. *J. Med. Chem.* **57**, 7874–7887 (2014).
74. Malumbres, M. & Barbacid, M. Cell cycle kinases in cancer. *Curr. Opin. Genet. Dev.* **17**, 60–65 (2007).
75. Goldberg, D. E., Zhang, M. & Nussenzweig, V. Plasmodium eIF2 $\alpha$  kinases in *Protein Phosphorylation in Parasites* (eds Doerig, C., Spaeth, G. & Wiese, M.) 123–130 (Wiley-Blackwell, 2013).
76. Vicogne, J. *et al.* An unusual receptor tyrosine kinase of *Schistosoma mansoni* contains a Venus Flytrap module. *Mol. Biochem. Parasitol.* **126**, 51–62 (2003).
77. Vanderstraete, M. *et al.* Venus kinase receptors control reproduction in the platyhelminth parasite *Schistosoma mansoni*. *PLoS Pathog.* **10**, e1004138 (2014).
78. Guidi, A. *et al.* Application of RNAi to genomic drug target validation in schistosomes. *PLoS Negl. Trop. Dis.* **9**, e0003801 (2015).
79. Buro, C. *et al.* Imatinib treatment causes substantial transcriptional changes in adult *Schistosoma mansoni* *in vitro* exhibiting pleiotropic effects. *PLoS Negl. Trop. Dis.* **8**, e2923 (2014).
80. Hong, Y. *et al.* Proteomics analysis of differentially expressed proteins in schistosomula and adult worms of *Schistosoma japonicum*. *Acta Trop.* **126**, 1–10 (2013).
81. Wang, Y. *et al.* Advances in metabolic profiling of experimental nematode and trematode infections. *Adv. Parasitol.* **73**, 373–404 (2010).
82. Rojo-Arreola, L. *et al.* Chemical and genetic validation of the statin drug target to treat the helminth disease, schistosomiasis. *PLoS One* **9**, e87594 (2014).
83. Panic, G., Vargas, M., Scandale, I. & Keiser, J. Activity profile of an FDA-approved compound library against *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **9**, e0003962 (2015).
84. Abdulla, M. H. *et al.* Drug discovery for schistosomiasis: hit and lead compounds identified in a library of known drugs by medium-throughput phenotypic screening. *PLoS Negl. Trop. Dis.* **3**, e478 (2009).
85. Katz, N., Couto, F. F. & Araujo, N. Imatinib activity on *Schistosoma mansoni*. *Mem. Inst. Oswaldo Cruz* **108**, 850–853 (2013).
86. Ramamoorthi, R., Graef, K. M. & Dent, J. Repurposing pharma assets: an accelerated mechanism for strengthening the schistosomiasis drug development pipeline. *Future Med. Chem.* **7**, 727–735 (2015).

87. Young, N. D. *et al.* Whole-genome sequence of *Schistosoma haematobium*. *Nat. Genet.* **44**, 221–225 (2012).
88. Berriman, M. *et al.* The genome of the blood fluke *Schistosoma mansoni*. *Nature* **460**, 352–358 (2009).
89. Kent, W. J. BLAT - the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
90. Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
91. *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* **460**, 345–351 (2009).
92. Huang, Y. *et al.* The carcinogenic liver fluke, *Clonorchis sinensis*: new assembly, reannotation and analysis of the genome and characterization of tissue transcriptomes. *PLoS One* **8**, e54732 (2013).
93. Young, N. D. *et al.* The *Opisthorchis viverrini* genome provides insights into life in the bile duct. *Nat. Commun.* **5**, 4378 (2014).
94. Young, N. D., Hall, R. S., Jex, A. R., Cantacessi, C. & Gasser, R. B. Elucidating the transcriptome of *Fasciola hepatica* - a key to fundamental and biotechnological discoveries for a neglected parasite. *Biotechnol. Adv.* **28**, 222–231 (2010).
95. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
96. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
97. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
98. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. & Bairoch, A. UniProtKB/Swiss-Prot. *Methods Mol. Biol.* **406**, 89–112 (2007).
99. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
100. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
101. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
102. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
103. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
104. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
105. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
106. Harris, T. W. *et al.* WormBase 2014: new views of curated biology. *Nucleic Acids Res.* **42**, D789–D793 (2014).
107. Drysdale, R. & FlyBase Consortium. FlyBase: a database for the *Drosophila* research community. *Methods Mol. Biol.* **420**, 45–59 (2008).
108. Eppig, J. T. *et al.* The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res.* **43**, D726–D736 (2015).
109. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).
110. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–D1097 (2014).
111. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
112. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).

## Acknowledgements

This project was funded by the National Health and Medical Research Council (NHMRC) of Australia and the Australian Research Council (ARC), and supported by a Victorian Life Sciences Computation Initiative (VLSCI) grant number VR0007 on its Peak Computing Facility at the University of Melbourne, an initiative of the Victorian Government. Other support from the Australian Academy of Science, the Australian-American Fulbright Commission, Alexander von Humboldt Foundation and Melbourne Water Corporation (R.B.G.) is gratefully acknowledged. A.J.S. is a recipient of a Melbourne International Research Scholarships (MIRS) and a Melbourne International Fee Remission Scholarship (MIFRS) from the University of Melbourne. N.D.Y. is an NHMRC Early Career Research (ECR) Fellow. We thank Stefano Mangiola for help with preliminary data analysis.

## Author Contributions

A.J.S., R.B.G. and N.D.Y. were involved in the experimental design. A.J.S., R.B.G. and N.D.Y. were responsible for writing and editing of the manuscript. A.J.S. and N.D.Y. conducted bioinformatic analyses. Other authors (A.R.J., P.W.S., P.T., P.R.B. and A.H.) contributed to the writing of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Stroehlein, A. J. *et al.* Defining the *Schistosoma haematobium* kinome enables the prediction of essential kinases as anti-schistosome drug targets. *Sci. Rep.* **5**, 17759; doi: 10.1038/srep17759 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>