

SCIENTIFIC REPORTS



OPEN

Structural and developmental expression of *Ss-riok-2*, an RIO protein kinase encoding gene of *Strongyloides stercoralis*

Wei-Qiang Lei¹, James B. Lok², Wang Yuan¹, Yue-Zhou Zhang¹, Jonathan D. Stoltzfus^{2,4}, Robin B. Gasser^{1,3}, Si-Yuan He¹, Huan Zhou¹, Rui Zhou¹, Jun-Long Zhao¹ & Min Hu¹

RIO kinases are essential atypical protein kinases in diverse prokaryotic and eukaryotic organisms, playing significant roles in yeast and humans. However, little is known about their functions in parasitic nematodes. In the present study, we have isolated and characterized the full-length cDNA, gDNA and a putative promoter of a RIOK-2 protein kinase (*Ss*-RIOK-2) encoding gene (*Ss-riok-2*) from *Strongyloides stercoralis*, a medically important parasitic nematode (Order Rhabditida). A three-dimensional structure (3D) model of *Ss*-RIOK-2 was generated using the *Chaetomium thermophilum* RIOK-2 protein kinase (*Ct*-RIOK-2) crystal structure 4GYG as a template. A docking study revealed some critical sites for ATP binding and metal binding. The putative promoter of *Ss-riok-2* contains a number of conserved elements. RNAseq analysis revealed the highest levels of the *Ss-riok-2* transcript in free-living females and parasitic females. To identify anatomical patterns of *Ss-riok-2* expression in *S. stercoralis*, we observed expression patterns of a transgene construct encoding green fluorescent protein under the *Ss-riok-2* promoter in post free-living *S. stercoralis*. Expression driven by this promoter predominated in intestinal cells. This study demonstrates significant advancement in molecular and cellular biological study of *S. stercoralis* and of parasitic nematodes generally, and provides a foundation for further functional genomic studies.

The atypical serine protein kinase RIOK-2, one of four members of the RIOK family (RIOK-1, RIOK-2, RIOK-3 and RIOK-B), named for the right open reading frame 1 domain in the prototype member of this family, RIOK-1 from yeast¹. RIOK-2 is essential for its regulatory functions in ribosome biogenesis and other cell cycle events²⁻⁶. Among the four RIOK protein kinases, RIOK-2 has been observed in eubacteria, archaea and many eukaryotes including humans, yeast and *Plasmodium* spp.^{2,6-9}. Crystal structures of *Afu*-RIOK-2 from *Archaeoglobus fuldigi* (Archaea) and *Cth*-RIOK-2 from *Chaetomium thermophilum* (a thermophilic fungus) reveal that RIOK-2 consists of a conserved N-terminal winged helix (wHTH) domain and a RIO domain structurally homologous to atypical protein kinases. The N-terminal domain of RIOK-2 is similar to a nucleic acid-binding motif common to many DNA-binding proteins that contain a winged helix fold, which mediates protein-protein interactions and protein-RNA interactions¹⁰⁻¹². Another structural feature found exclusively in the C-terminal domains of eukaryotic RIOK-2s is a conserved extension sequence⁴.

In yeast, RIOK-2 occurs in both the nucleus and cytoplasm⁴. In human cells, RIOK-2 (hRIOK-2) localizes to the cytoplasm of cells at steady state but rapidly accumulates in the nucleus when the exportin CRM1 (chromosome region maintenance 1) protein is inhibited by leptomycin B⁶. There is evidence that RIOK-2 is released to the cytoplasm after binding CRM1 at a functional leucine-rich nuclear export signal in hRIOK-2's C-terminal domain via RanGTP^{13,14}. Although the function of RIOK-2 in nuclear steps of biogenesis is presently unclear,

¹State Key Laboratory of Agricultural Microbiology, Key Laboratory of Development of Veterinary Diagnostic Products, Ministry of Agriculture, College of Veterinary Medicine, Huazhong Agricultural University, 1 Shizishan Street, Wuhan, 430070, China. ²Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, 3800 Spruce Street, Philadelphia, PA, 19104, USA. ³Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Corner of Flemington Road and Park Drive, Parkville, Victoria, 3010, Australia. ⁴Present address: Department of Biology, Millersville University, Millersville, Pennsylvania, USA. Correspondence and requests for materials should be addressed to M.H. (email: mhu@mail.hzau.edu.cn)

RIOK-2 also has an important function in ribosome biogenesis after nuclear export. In both yeast and human cells, RIOK-2 is a non-ribosomal factor necessary for processing the 18 S precursor ribosomal RNA (pre-rRNA) and for cytoplasmic maturation of the 40 S ribosomal subunit. RIOK-2 is also a trans-acting factor involved in late steps of 40 S biogenesis^{2, 4, 5}. In human cells, RIOK-2 is required for the processing of 18 S pre-rRNA, and for the release of ENP-1 (essential nuclear protein 1, a 40S-associated trans-acting factor) from cytoplasmic 40S precursors, and for the recycling of DIM-2 (defective entry into mitosis 2), LTV-1 (low temperature viability Protein 1) and NOB-1 (nin one binding protein 1), which are additional 40S-associated trans-acting factors⁶.

In *Chaetomium thermophilum*, kinase activity of RIOK-2 is identified as the enzyme is released from the late pre-40S particle via its catalytic activity and thus influences the progression and recycling of other late pre-40S processing factors. Moreover, loss of kinase activity results in significant growth defects and cold sensitivity in *C. thermophilum*¹⁵. In human HeLa cells, phosphorylation of RIOK-2 by PLK-1 (Polo-like Kinase 1, encoding an essential gene in the maintenance of the mitotic progression) at Ser-335, Ser-380 and Ser-548 regulated the programmed transition of the cell cycle from metaphase to anaphase. RIOK-2 over-expression causes prolonged cessation of mitosis whereas knockdown of RIOK-2 can accelerate mitotic progression¹⁶. However, in glioblastoma cells, over-expression of RIOK-2 can also promote tumorigenesis by acting upstream of Akt signaling to stimulate Tor-complex-2 (TORC2) signaling¹⁷. This suggests that RIOK-2 may have crucial but different functions in complex signaling events in diverse cells to regulate the entire cell-cycle.

In *Plasmodium* spp, the RIOK-2 kinase is also conserved among the 65 protein kinases in this parasite^{8, 18}. The peptide binding sites and protein kinase inhibitor binding domain of *Plasmodium falciparum* RIOK-2 kinase highlights this molecule as a potential druggable target^{19, 20}. The genome of the free-living nematode *C. elegans* encodes a RIOK-2, and silencing of *Ce-riok-2* by RNAi leads to embryonic lethality and sterility (http://www.wormbase.org/species/c_elegans/gene/WBGene00013688). This suggests that RIOK-2 is essential for the development and reproduction of this worm. The necessity of RIOK-2 for larval development in *C. elegans* opens the question of its potential as an essential regulator of development in parasitic nematodes. Consequently, we are characterizing a *riok-2* homolog in the intestinal parasitic nematode *Strongyloides stercoralis*. *S. stercoralis*, a parasite of dogs and humans, causes chronic debilitation in uncomplicated infections, and, by virtue of its ability to undertake cycles of autoinfection, can precipitate a potentially fatal disseminated hyperinfection in patients co-infected with HTLV-1 or undergoing corticosteroid therapy²¹. Human strongyloidiasis is grouped among the “neglected tropical diseases”²². *Strongyloides* spp. also offer significant advantages as subjects for functional genomic study, having published genome sequences²³ and being amenable to transgenesis using techniques adapted from *C. elegans* biology²⁴.

In the present paper, we report the discovery of *Ss-riok-2* in *S. stercoralis*. As first steps in characterizing the structure and function of this gene, we have isolated and sequenced cDNA corresponding to *Ss-riok-2* homolog, the gDNA and the putative promoter for this gene. We also explored the temporal and spatial patterns of *Ss-riok-2* expression. The primary objective of this study was to uncover clues to the function of *Ss-riok-2* with an eye to deducing its role in the biology of *Strongyloides* spp.

Results

Characterisation of *Ss-riok-2* cDNA and the evolutionary relationship of *Ss*-RIOK-2 to its homologues from other species. The full-length coding region of *Ss-riok-2* is 1572 bp in length, encoding 523 aa. The coding sequence of *Ss-riok-2* has an A + T content of 69.21%, and comprises two exons. Comparison of *Ss*-RIOK-2 with its homologues showed that *Ss*-RIOK-2 has significant similarities to RIOK-2s from a range of organisms, including three nematodes (*L. loa*; *C. elegans*; *H. contortus*) and six other organisms (*A. florea*; *H. sapiens*; *D. rerio*; *C. familiaris*; *A. fulgidus*; *S. cerevisiae*). The highest amino acid similarity recorded was to the *Hc*-RIOK-2 from *H. contortus*. Pairwise comparisons of amino acid sequence between *Ss*-RIOK-2 and selected sequences revealed sequence identities ranging from 36.99% to 45.94%.

The alignment of the amino acid sequences of *Ss*-RIOK-2 with selected RIOK-2 kinases (Fig. 1) showed conserved regions in the ATP-binding motif (sub-domain I), the hinge region (subdomain V), the active site (sub-domain VIb), the metal binding loop (subdomain VII), the flexible loop inserted between the third β strand and the α helix C, and other subdomains, including II, III, IX. This level of conservation suggests that *Ss*-RIOK-2 is functionally similar to other RIOK-2 kinases. In addition, *Ss*-RIOK-2 contains the specific phosphate-binding loops of subfamily RIOK-2 domains (with the sequence GxGKES) in the ATP binding motif and a highly conserved N-terminal winged-helix (wHWH) domain, the signature characteristic of the RIOK-2 kinase family. The alignment also showed that the amino acid sequences in regions external to these subdomains are more divergent (Fig. 1) than the sequences in the subdomains.

The full-length amino acid sequences of *Ss*-RIOK-2 and 14 other RIOK-2 homologues representing a range of different species were aligned and subjected to phylogenetic analyses (Fig. 2). There was concordance in topology between the MP, ML and NJ trees, which showed that *Ss*-RIOK-2 was grouped with RIOK-2s from other nematode species with absolute bootstrap support (100%) to the exclusion of molecules from organisms from other phyla. RIOK-2s from amphibians and mammals each formed separate clades, also supported by strong bootstrap values (99% and 100%, respectively, Fig. 2).

Three-dimensional structural modelling of the *Ss*-RIOK-2 protein kinase. To obtain the structural details and arrangement of these domains within *Ss*-RIOK-2 protein kinase, we modeled the structure of the *Ss*-RIOK-2 protein kinase using crystal structure of *C. thermophilum* RIOK-2 (*Ct*-RIOK-2) kinase as a template (PDB accession number 4GYG). Structural features present in *Ss*-RIOK-2 protein kinase are comparable to the structure of *Ct*-RIOK-2 with similar arrangements of structural elements (Fig. 3). The N-terminal wHWH domain is followed by an ATP binding site and a β -sheet as well as a long α helix, known as α C, followed by a long Hinge region and a C-terminal domain with catalytic and metal-binding loops, followed by a long α -helix

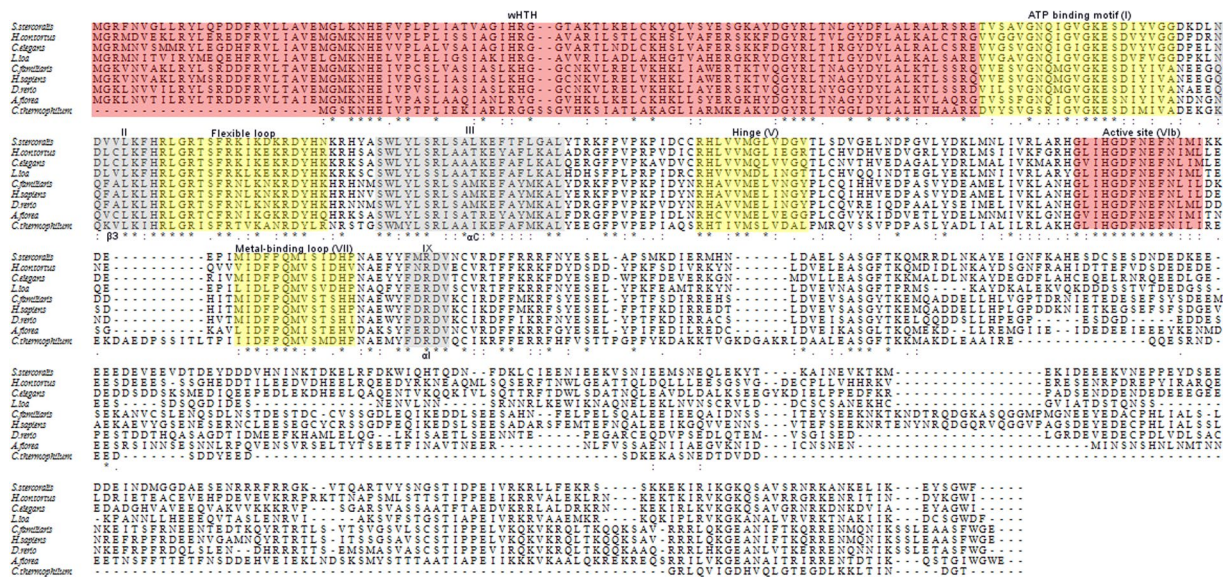


Figure 1. Multiple sequence alignment among predicted RIOK-2 proteins from eight organisms (*Strongyloides stercoralis*, *Loa loa*, *Haemonchus contortus*, *Caenorhabditis elegans*, *Apis florea*, *Danio rerio*, *Canis familiaris*, *Homo sapiens*, *Chaetomium thermophilum*). The N-terminal domain contains the conserved wHTH domain (red). The RIO kinase domain contains the ATP-binding motif (yellow), the flexible loop (yellow), the hinge region (yellow), and the catalytic (red) and metal-binding loops (yellow) as determined from the structure of the *Archaeoglobus fulgidus* RIOK-2 protein. The predicted subdomains I-IX are marked above the alignment. Asterisks indicate identical residues. Dashes indicate gaps in the sequence, included for alignment purposes.

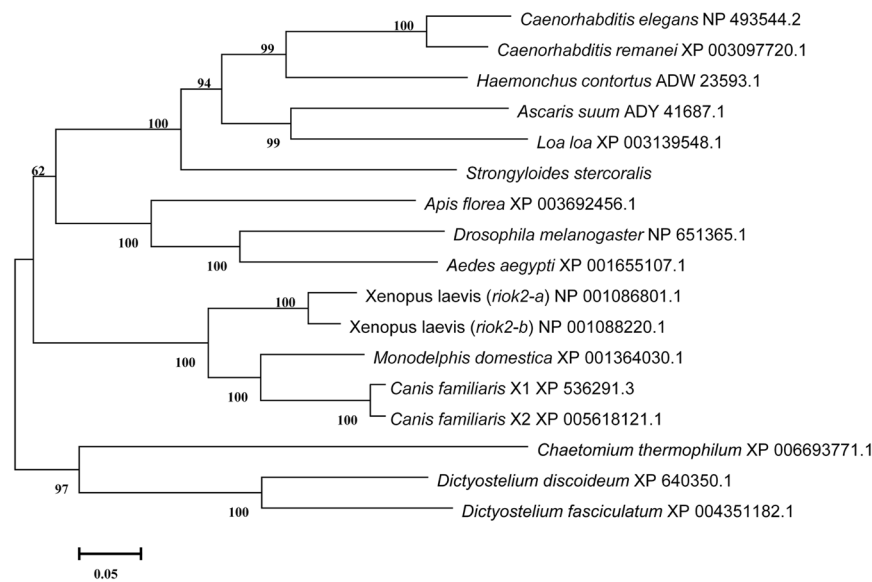


Figure 2. The neighbour-joining tree of RIOK-2 amino acid sequences from a range of organisms. These species are *Ascaris suum* (ADY41687.1), *Caenorhabditis elegans* (NP_493544.2), *Caenorhabditis remanei* (XP_003097720.1), *Haemonchus contortus* (ADW23593.1), *Loa loa* (XP_003139548.1), *Aedes aegypti* (XP_001655107.1), *Apis florea* (XP_003692456.1), *Canis familiaris* (XP_536291.3, XP_005618121.1), *Drosophila melanogaster* (NP_651365.1), *Dictyostelium discoideum* (XP_640350.1), *Dictyostelium fasciculatum* (XP_004351182.1), *Chaetomium thermophilum* (XP_006693771.1), *Monodelphis domestica* (XP_001364030.1), *Xenopus laevis* (NP_001086801.1, NP_001088220.1). Accession numbers for the various sequences in the NCBI Protein Database are given next to each species name. Numbers represent bootstrap values (after 1000 iterations) and the scale bar represents branch length as a fraction of the total tree length. The bootstrap values of >95% were displayed in the tree.

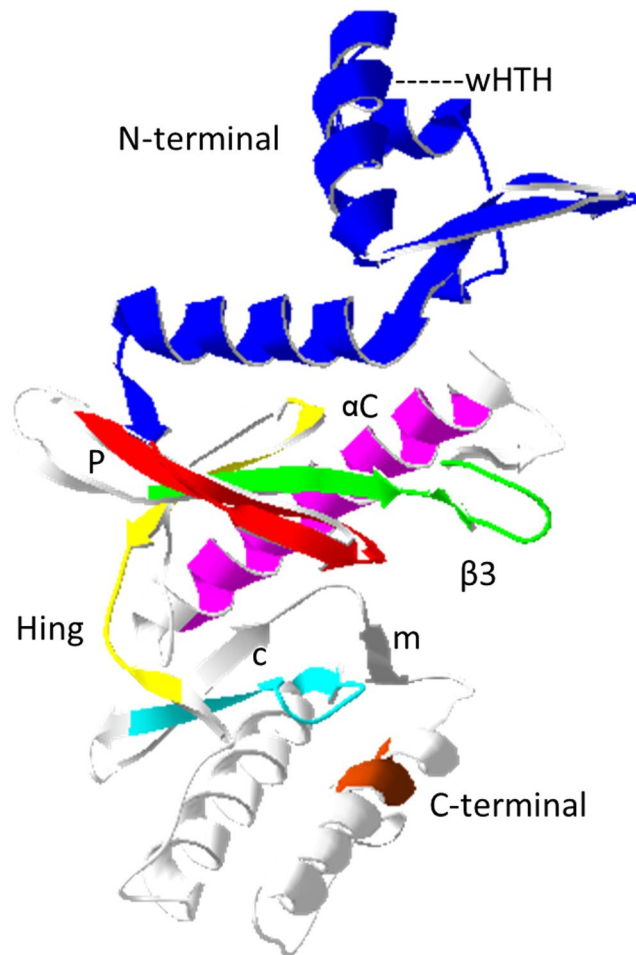


Figure 3. Homology model of the RIOK-2 protein kinase from *Strongyloides stercoralis*. The modeled structure of the *Ss*-RIOK-2 protein kinase shows different structural elements, N-terminal helix winged domain (blue), ATP binding motif (red), $\beta 3$ (green), αC (pink), Hing (yellow), active site (cyan), metal-binding loop (grey) and αI (Amaranth).

(named αI), which is wedged between the N and C lobes close to the active site in eukaryotic *C. thermophilum*. Most of the residues were present within the allowed region of the Ramachandran plot (Supplementary Fig. S2B). ERRAT PLOT analysis of the final model shows that most of the residues are below 95% with a model structural quality factor of 90.612. Overall, no short contact was observed in the final model (Supplementary Fig. S2C). In a Verify 3D plot, 82.46% residues of the model are above 0.2 in the 3D/1D profile (Supplementary Fig. S2D).

The docking simulation demonstrated that *Ss*-RIOK-2 has a conserved ATP binding site (Fig. 4A). Hydrophobic amino acid residues Met188, Gly189, Leu190, Val191 and Met235 form adenine subpocket within this site. The 6-NH₂ of adenine is predicted to hydrogen bond (2.8 Å) to the main chain carboxyl group of Gly189. A weak hydrogen bond (3.2 Å) was also suggested between 1-N atom of adenine and main chain NH of Val191. The ribose moiety of ATP was exposed to solvent and sandwiched by amino acid Ile109, Ile245 and Phe232. The 3'-hydroxyl of ribose oriented to the main chain carboxyl of Phe232 and hydrogen bonded (3.1 Å). Phosphate was in a highly polar and charged subpocket that included Lys105, Glu106, Lys123, Asn233, Asp246 and magnesium ions (Fig. 4B). The magnesium ion (purple sphere in Fig. 4B) was chelated by the negatively charged α and γ phosphates of ATP, Asn233 and Asp246 residues. The ionic interaction (3.2 Å) was predicted between the negatively charged β phosphate and positively charged Lys105 residue.

Genomic organization of *S. stercoralis riok-2* and characterization of its putative promoter.

The structures of the genes *riok-2* from *S. stercoralis* and its orthologues from *C. thermophilum*, *C. elegans* and *H. contortus* (Fig. 5) were inferred through separate alignments of genomic DNA sequences with cDNA sequences. These alignments revealed that *Ss-riok-2* is the shortest of the *riok-2* genes in these nematodes, comprising two exons and one intron and that, similar to *Ss-riok-2*, the *C. elegans* homologue (code Y105E8B.3) has one ORF (4953 nucleotides) comprising three exons and two introns. The spliced transcript of the *C. elegans* gene *riok-2* (1587 nucleotides) encoded a protein comprising 529 aa. By contrast, the alignment revealed that the *riok-2* of *H. contortus* (*Hc-riok-2*) is the longest and the most complex *riok-2* gene among the four species, comprising 6835 bp

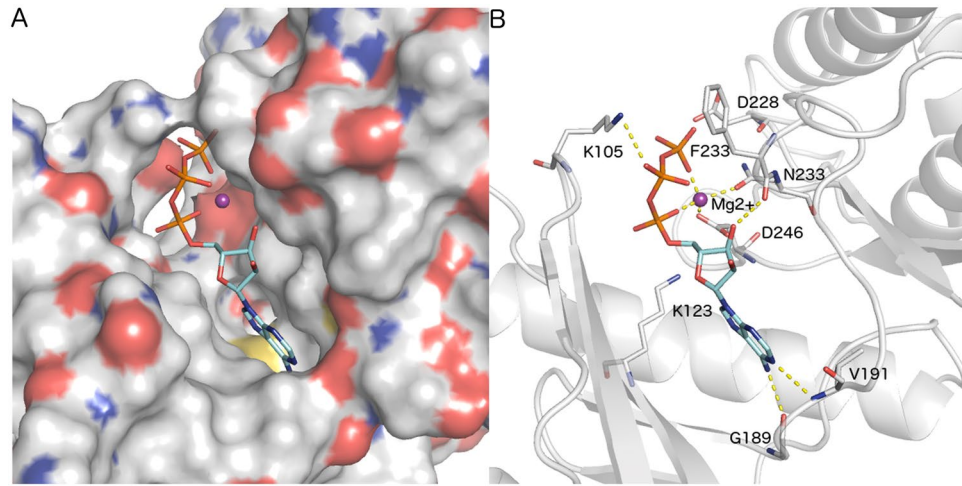


Figure 4. Surface structure of the *Ss*-RIOK-2 ATP binding pocket and interaction of ATP within modeled structure of *Ss*-RIOK-2-ATP. (A) Surface structure of the ATP binding pocket of *Ss*-RIOK-2-ATP. Blue is positive, red is negative, Mg^{2+} is shown as a purple sphere, ATP and catalytic residues are shown in stick representation. (B) Interaction of ATP with key residues present within the binding pocket of *Ss*-RIOK-2-ATP.

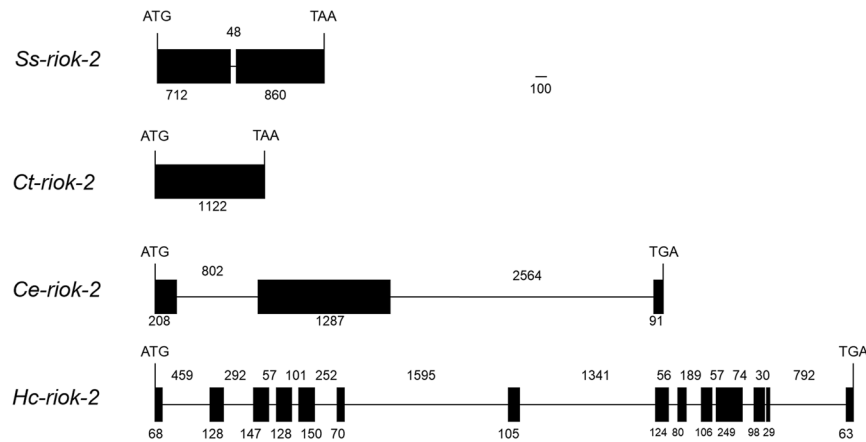


Figure 5. Diagrammatic representation of the genomic organizations of the *riok-2* from *Strongyloides stercoralis* (*Ss-riok-2*) and orthologues from *Chaetomium thermophilum* (*Ct-riok-2*), *Caenorhabditis elegans* (*Ce-riok-2*) and *Haemonchus contortus* (*Hc-riok-2*). The organization of each gene was determined by aligning the cDNA and genomic DNA sequences, with intron-exon boundaries being defined using the GT-AG rule. Black boxes represent exons, whilst horizontal lines represent introns. Numbers below the boxes indicate the sizes of exons (in nucleotides), whereas numbers above the lines indicate the intron sizes.

apportioned between 14 exons and 13 introns and encoding a transcript (code HQ198855.1) of 1587 nucleotides and a predicted protein 529 aa in length.

The predicted promoter of *Ss-riok-2* is 1308 bp in length, and the full length *Ce-riok-2* promoter sequence is 1269 bp in length. A sequence comparison revealed that the *Ss-riok-2* promoter is 51% identical to that of *Ce-riok-2* (Supplementary Fig. S3). The putative promoters of both genes are A + T rich, with the A + T content of 79.85% for *Ss-riok-2* and 62.49% for *Ce-riok-2*. A number of promoter elements, including eight TATA boxes, and one GATA (WGATAR) box, four inverse GATA (TTATC) boxes, three E-boxes (CANNTG), one CAAT (CCAAT) and one inverse CAAT (ATTGG) boxes are predicted within the *Ss-riok-2* promoter. In contrast, *Ce-riok-2* promoter region has one TATA box, one inverse GATA (TTATC) box, one E-box (CANNTG), two CAAT (CCAAT) and two GC boxes. Most of these motifs are dispersed across the promoters of the two genes, with no apparent pattern to their distribution.

Transcriptional and expression profiles of *Ss-riok-2*. RNAseq revealed that *Ss-riok-2* is transcribed in all developmental stages of *S. stercoralis* with the highest transcript abundance in parasitic and free-living females and the lowest transcript abundance in post-free-living first-stage larvae (Fig. 6). Abundance of these transcripts increased significantly during the transitions from PFL L1 to iL3 and PP L1 to PP L3. In general, transcript abundance was significantly higher in post-parasitic stages, most of which develop heterogonically to free-living males

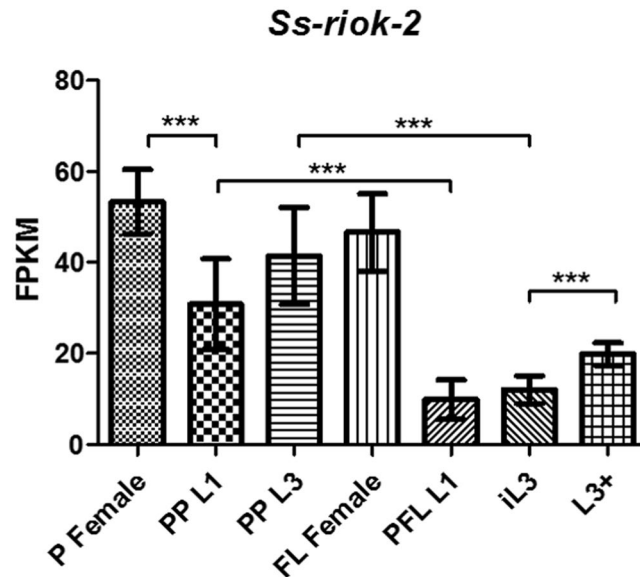


Figure 6. Transcriptional profile of *Ss-riok-2* in seven *Strongyloides stercoralis* life stages. Transcript abundances were compared in biological triplicate. Stages examined were: parasitic females (P Female), post-parasitic first-stage larvae (PP L1), post-parasitic third-stage larvae (PP L3), free-living females (FL Female), post-free-living first-stage larvae (PFL L1), infectious third-stage larvae (iL3), and *in vivo* activated third-stage larvae (L3+). Transcript abundances are expressed as fragments per kilobase of coding exon per million mapped reads (FPKM). Error bars represent 95% confidence intervals.

and females in the UPD strain, compared with post-free living larval stages, which develop exclusively via the homogenic to iL3 in all *S. stercoralis*.

To investigate the anatomical expression pattern of *Ss-riok-2*, we transformed parental free-living *S. stercoralis* females with construct pPR2 (Supplementary Fig. S1). Transformed females were paired with wild type *S. stercoralis* males and their post-free-living L1 progeny were screened for GFP fluorescence. Some of the immature eggs had GFP expression, even when they were still in the uteri of the free-living females (data not shown). GFP expression was most frequently observed in the intestine of transgenic L1 (Fig. 7) despite a varying intensity of GFP expression among different individuals.

Autophosphorylation of *Ss*-RIOK-2. As the natural substrate of *Ss*-RIOK-2 is unknown, we assessed its autophosphorylation activity. Recombinant His-tagged enzyme (designated His-*Ss*-RIOK-2) was expressed in *E. coli* (Fig. 8A). Purified recombinant His-*Ss*-RIOK-2 incubated with [γ ³²P] ATP produced radioactive signals, whereas purified recombinant His-GFP incubated with [γ ³²P] ATP didn't produce radioactive signal, indicating that the His-*Ss*-RIOK-2 is capable of auto-phosphorylation (Fig. 8B).

Discussion

Atypical serine/threonine RIOK-2 kinases are essential for viability of yeast, human cells and for *C. elegans*^{2,9,25–27}. In the present study, we determined the structure and discovered some functionalities of the *Ss-riok-2* gene in *S. stercoralis*, a parasitic nematode infecting human beings and dogs.

The catalytic cores of the majority of kinases are made up of 11 conserved subdomains similar to *Ct*-RIOK-2 (Fig. 3). The kinase domain in *Ss*-RIOK-2 is also shorter than the typical protein kinases present in eukaryotes. It lacks a critical activation loop (subdomain VIII), which was thought to be necessary for peptide recognition and binding¹, and the last two helices of the C-terminal lobe (subdomain X and XI). The ATP binding pocket of *Ss*-RIOK-2 is well configured to dock with this molecule. The interaction between *Ss*-RIOK-2 and ATP features several strong interactions with atoms in the substrate (Fig. 4B). The conserved RIOK-2 signature sequence “GxGKES” in the ATP binding motif is slightly different from the signature sequences of RIOK-1 domains (with the sequence STGKEA) and RIOK-3 domains (with the sequence STGKES). The crystal structure of the fungal *Ct*-RIOK-2¹⁵ allows the study of the back pocket of the ATP-binding region and the design of specific compounds to bind these hydrophobic residues. In *P. falciparum*, a number of protein kinase inhibitors were shown within the ATP binding domain in an in-silico docking simulation using *Pf*-RIOK-2 kinase as target enzyme^{19,20}. Therefore, this *Ss*-RIOK-2-ATP model could similarly be used to screen and identify some potent inhibitors by molecular docking. The 18 conserved residues of the flexible loop between β 3 and α C of *Ss*-RIOK-2 are greatly disordered, making the sequence of this flexible loop distinct from those of the other RIO kinases. Thus, the function of this region may be different for each RIO kinase in an organism. Deletion of the flexible loop caused a slow-growth phenotype in yeast, particularly at lower temperatures¹⁵. Furthermore, the flexible loop is speculated to be a recognition or binding feature that orients the substrate and maintains its position during the catalytic phosphoryl transfer. *Ss*-RIOK-2 is capable of auto-phosphorylation (Fig. 8), a property of the RIOK-2 already described from

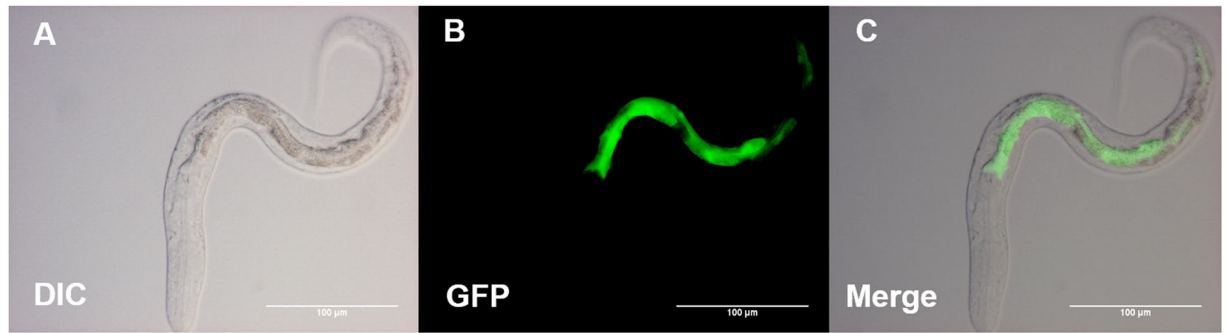


Figure 7. Representative expression pattern of GFP under the putative *Ss-riok-2* promoter in a transgenic second-stage larva of *Strongyloides stercoralis*. Differential interference contrast (DIC) and fluorescence (GFP) images showing the expression of construct *S. stercoralis Ss-riok-2p::gfp* (pRP2; Fig. S1) in a post-free-living second-stage larva. GFP reporter expression predominated in intestine. Scale bars = 10 μ m.

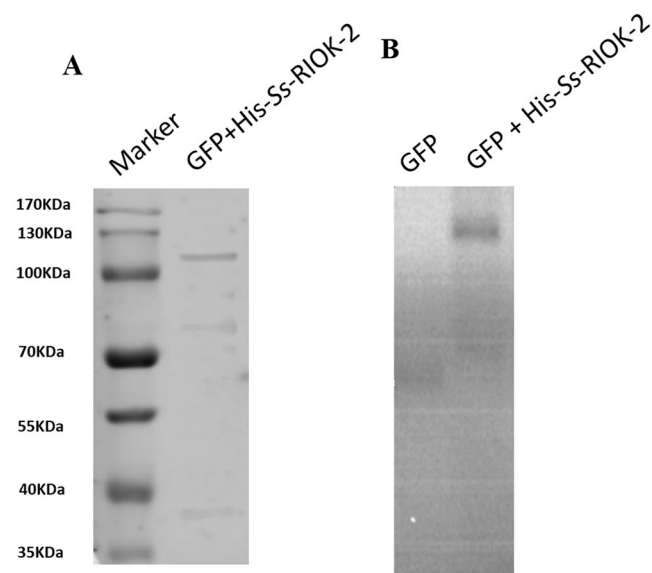


Figure 8. Autophosphorylation of recombinant His-Ss-RIOK-2. (A) Western-blot of His-Ss-RIOK-2 immunoprecipitated with His antibody. (B) Kinase assay showing autophosphorylation activities of recombinant His-Ss-RIOK-2.

A. fulgidus, *S. cerevisiae* and humans^{2,6,7}. By analogy, *Ss-RIOK-2* may also exclusively phosphorylate serines. The auto-phosphorylation site identified by phosphopeptide mapping and sequencing analysis was determined to be invariant at Ser 128 in *Af-RIOK-2*²⁸. This residue is located on the flexible loop and within the sequence RLGRTSFRKIK in *Ss-RIOK-2* and *Ce-RIOK-2*. The conserved catalytic regions in ePKs are thought to play key roles in transferring phosphate from ATP to substrate proteins. In *S. cerevisiae* *RIOK-2*, active site mutations indicate that *RIOK-2* kinase activity is required for its dissociation from the late pre-40S particle and that it influences the progression and recycling of other late pre-40S processing factors. All these mutants supported cell growth but yielded a cold-sensitive phenotype in yeast¹⁵. The *Ss-RIOK-2* catalytic loop was predicted by the model to interact between the conserved catalytic residue Asp-228 (Asp-229 in *S. cerevisiae*, is the proposed catalytic base for deprotonation of the substrate serine hydroxyl group in *S. cerevisiae* *RIOK-2*) and the metal-binding loop containing the conserved Asp-246 residue (Asp253 in *Sc-RIOK-2* is responsible for binding to Mg^{2+}), indicating that phosphoryl transfer from ATP to Asp 246 in *Ss-RIOK-2* (Fig. 4B) is similar to that in *S. cerevisiae* *RIOK-2*.

Besides the active site, the conserved N-terminal domain (wHTH) of *RIOK-2*s also participates in some biological processes. In multiple sequence alignment, we identified a wHTH domain (1–98) in *Ss-RIOK-2* (Fig. 1), the characteristic structure present in *RIOK-2* kinase family. The wHTH domain is significant for its capacity as a DNA-binding protein, and also mediates protein-protein and protein-RNA interactions^{10–12}. Such winged helices bind DNA through site-specific interactions of its major grooves with residues on -helix 3 of transcription factor HNF-3²⁹ and specific interactions of W1, β 2 and β 3 with the major groove of the DNA in hRFX1¹⁰. Homologous conserved residues from the wHTH domain of *Ss-RIOK-2* (Lys68, Arg91) suggest similar DNA binding properties.

In addition to the active sites and the N-terminal region, the C-terminal region of RIO domain consists of a kinase-inhibiting activity domain. This domain's interaction with the ribosomal complex suggests this region as a target for the design of specific compounds that can inhibit RIOK activation or interactions. These findings suggest that Ss-RIOK-2 is an active protein kinase with functions that are similar to those of its homologues in yeast and humans, and it follows from this that Ss-RIOK-2 may represent a promising drug target.

Ss-riok-2 has only one intron as compared to its homologues from *C. elegans* and *H. contortus* (Fig. 5). *Ss-riok-2* gene is the shortest and has the fewest exons of the nematode *riok-2s* studied to date. Realizing that gene structure is not directly related to essential function, we nevertheless hypothesize that *Ss-riok-2* is required for transcriptional regulation by a complex mechanism. The 5'-UTR of *Ss-riok-2* is similar in length to that of *C. elegans riok-2*. The 5'-UTR of *Ss-riok-2* has 51% identity to that of the *Ce-riok-2* gene and also shares some promoter elements with the *C. elegans* homolog, though the elements' similarity is limited (Fig. S3). Common elements in the promoters of *Ss-riok-2* and *Ce-riok-2* included TATA boxes, GATA and E-box. The TATA box is recognized by transcription factor IID and is necessary for regulation of transcriptional start point and the frequency with which the gene is transcribed^{30,31}. The GATA box is recognized by GATA transcription factors and is necessary for regulation of eukaryotic development and reproduction^{32–36}. E-boxes are recognized and bound by basic helix–loop–helix (bHLH) proteins which regulate a wide range of developmental processes in eukaryotic organisms^{37–41}. Conservation of these elements in *riok-2s* of *C. elegans* and *S. stercoralis*, as demonstrated here, suggests that their encoded kinases play an important role in gene transcription and expression in both free-living and parasitic nematodes.

To gain insight into the functions of *Ss-riok-2*, transcript abundance was measured in all developmental stages of *S. stercoralis*. *Ss-riok-2* transcripts occur throughout the life cycle (Fig. 6), suggesting that they function in the development of all stages of this parasite. In addition, the highest transcript abundance occurs in the parasitic and free-living females of *S. stercoralis*, suggesting that *Ss-riok-2* also plays important roles in the reproduction of this parasitic nematode. This is consistent with the transcriptional profiles of *riok-2* of *D. melanogaster*⁴² and of two parasitic nematodes *B. malayi*⁴³ and *H. contortus*^{44,45} as well as of the free-living nematode *C. elegans*. Silencing of *Ce-riok-2* by RNAi leads to embryonic lethality and sterility, indicating that *riok-2* is involved in nematode embryonic development and reproduction. Interestingly, after iL3 activation, abundance of *Ss-riok-2* transcripts is significantly higher in L3+ than in iL3, suggesting that RIOK-2 activity is necessary for *S. stercoralis* larvae to resume development in the host. Furthermore, an obvious trend seen in *S. stercoralis* transcript abundance profiles is that they are significantly higher in heterogonically developing larvae (those going to free-living adults) than in homogonically developing larvae (those going to L3i), suggesting that *Ss-riok-2* may have an important function in the reproductive development of *S. stercoralis*'s free-living adult stages, again supporting its role in the reproduction of this parasite.

Preliminary functional studies of *riok-2* genes in non-nematode species have revealed roles in ribosome maturation and cell cycle progression, indicating that *riok-2* genes are required for viability and fertility. In order to explore the function of *Ss-riok-2* in various tissues in *S. stercoralis*, the anatomical expression pattern was analyzed by expressing a transcriptional reporter based on *Ss-riok-2* in post free-living larvae of this parasite. Strong intestinal expression under the *Ss-riok-2* promoter was observed (Fig. 7). Intestinal expression is controlled by the ELT-2 transcription factor in *C. elegans*^{46,47}. A search of the 1.5 kb region upstream of the *Ss-riok-2* translational start site for ELT-2 recognition motifs revealed five characteristic TGATAA motifs, indicating that intestinal expression of *Ss-riok-2* may depend on an ELT-2-like transcription factor. This expression profile might not only identify spatio-temporal functionalities of the promoter, but also hint at specific functions of Ss-RIOK-2. In *C. elegans*, the intestine is a large organ that carries out diverse functions executed by multiple distinct organs in higher eukaryotes, including digestion of food, absorption of processed nutrients, synthesis and storage of macromolecules, initiation of an innate immune response to pathogens, and nurturing of germ cells by producing yolk^{48–51}. Nematode intestinal cells are very large and contain large nuclei with prominent nucleoli, many mitochondria and extensive rough endoplasmic reticulum with many ribosomes. RIOK-2 enzymatic activity is crucial for cleavage of 20S pre-rRNA and ribosome biogenesis^{1,2}. In this aspect, expression of *Ss-riok-2* in the intestine suggests that Ss-RIOK-2 also has an important function in ribosome biogenesis for *S. stercoralis* and underscores the potential of this protein as a therapeutic target. In addition, the digestive and metabolic activities of the intestine are central to the growth and development of any organism. The *C. elegans* intestine is an important endocrine tissue⁵². The yolk proteins are synthesized by and secreted from the intestine into their destination in the gonad^{48,51,53}. A large-scale double-stranded RNA interference (RNAi) study of *C. elegans* showed that the silencing of *Ce-riok-2* leads to sterility and embryonic lethality. Therefore, Ss-RIOK-2 may play an important role in development of eggs by interacting with yolk proteins in *S. stercoralis*.

In summary, this study isolated and characterised the cDNA, genomic DNA and promoter of *Ss-riok-2*, which encode the RIOK2 protein kinase in the parasitic nematode *S. stercoralis*. Ss-RIOK-2 contains a RIOK-2 signature motif (wHTH) and has high similarity with a range of homologues from different species. The predicted formation of a stable Ss-RIOK-2-ATP complex supports the feasibility of exploiting this novel allosteric site of Ss-RIOK-2 as a potential anti-parasite drug target. Recombinant Ss-RIOK-2 has kinase activity. We also ascertained its transcriptional profile in seven key developmental stages, finding the highest transcript abundance in parasitic and free-living females and confirming its localization in the intestine of this parasite. The characterization and temporal and spatial expression patterns of the RIOK-2 in *S. stercoralis* lay the foundation for exploring the function of RIOK-2 in the biological processes of parasitic nematodes.

Materials and Methods

Ethics statement. No human subjects were used in these studies. The *S. stercoralis* (UPD strain) was maintained in prednisolone-treated Beagles in accordance with a protocol (Permit Number: SYXK-0029) approved by the Committee on the Ethics of Animal Experiments of Hubei Province. The care and maintenance of animals

were in strict accordance with the recommendations in the Guide for the Regulation for the Administration of Affairs Concerning Experimental Animals of the P. R. China.

***S. stercoralis* strains, and their maintenance.** The *S. stercoralis* UPD strain was maintained in immuno-suppressed dogs and cultured as previously described⁵⁴. UPD strain iL3 for nucleic acid extraction and experimental infections in dogs were isolated via the Baermann funnel technique from seven-day-old charcoal coprocultures. The iL3s were washed several times with a sterile buffered saline called BU buffer⁵⁴. Free-living females for microinjection were transferred from the Baermann apparatus (along with a minimal fluid volume) to Nematode Growth Medium (NGM agar) plates seeded with *Escherichia coli* OP50⁵⁴. All cultures of *S. stercoralis* were incubated at 22 °C unless otherwise noted.

Genomic DNA and RNA preparation. *S. stercoralis* genomic DNA was extracted from 10,000–20,000 iL3s by sodium proteinase K treatment⁵⁵, followed by purification using an EasyPure Genomic DNA Kit (TransGen Biotech, Beijing, China). DNA samples were stored at –20 °C until use. RNA was extracted from 20,000–30,000 iL3s using the Trizol reagent (Life Technologies, USA) according to the manufacturer's protocol. RNA integrity and yields were verified by electrophoresis and spectrophotometry (Nano Drop Technologies, Thermo Scientific, USA), respectively, and 1 µg of RNA was reverse transcribed to cDNA using a kit (Smart RACE Kit, Clontech, USA). Extracted total RNA was treated with RQ1-RNase-Free DNase (Promega, USA) and then stored at –80 °C until use.

Isolation of the full-length *Ss-riok-2* cDNA. The genomic sequence of *Ss-riok-2* was determined using two highly degenerate primers *riok-2-BamHI/F* and *riok-2-SalI/R* (S1 Table) with restriction sites to amplify a coding sequence with homologies to *H. contortus* (GenBank: HQ207527) and to *C. elegans* (GenBank: NM_061143) from cDNA synthesised from total RNA extracted from iL3s. PCR (50 µl) was performed in 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 4 mM MgCl₂, 250 µM each of dNTP, 100 pmol of each primer and 1 U Taq polymerase (TaKaRa, Japan) under the following conditions: initial denaturation at 94 °C for 5 min; then 30 cycles of denaturation at 94 °C for 30 s, annealing at 57 °C for 30 s and extension at 72 °C for 2 min, with a final extension at 72 °C for 10 min. This PCR product was electrophoresed and visualised on a 1% agarose gel from which a single, major band with little background was excised, cloned into the pMD19-T vector (Takara Biotechnology) and sequenced.

Isolation of the putative promoter of *Ss-riok-2*. Template genomic DNA extracted from *S. stercoralis* iL3s, and a pair of primers with restriction sites (*Ss-riok-2-PstI/F* and *Ss-riok-2-SmaI/R*, S1 Table) were designed to amplify the *Ss-riok-2* promoter sequence using the following PCR cycling conditions: initial 94 °C, 5 min; then 94 °C, 30 s, 65 °C, 2 min for 35 cycles; final extension at 65 °C for 10 min. The PCR product was electrophoresed and visualised on a 1% agarose gel from which a single, major band with little background was excised, cloned into pMD19-T vector (Takara Biotechnology) and sequenced.

Bioinformatic and phylogenetic analyses. The sequence of *Ss-riok-2* was compared by BLASTx³⁷ with sequences in non-redundant databases from NCBI (<http://www.ncbi.nlm.nih.gov/>) to confirm the identity of genes isolated. Conceptual translation of *Ss-riok-2* cDNA into predicted amino acid (aa) sequences was performed using the free software package Bioedit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html#downloads>). The protein motifs of *Ss-RIOK-2* were identified by scanning the databases PROSITE⁵⁶ (www.expasy.ch/tools/scnpsit1.html) and Pfam⁵⁷ (www.sanger.ac.uk/Software/Pfam/). The aa sequence alignments were carried out with the homologues from selected species using the program MAFFT 7.0⁵⁸ (<http://mafft.cbrc.jp/alignment/software/>), and adjusted by manual inspection. The functional domains and subdomains of *Ss-RIOK-2* were identified in the protein alignment. Promoter elements in the 5'-UTR were predicted using the transcription element search system Matrixcatch (<http://www.gene-regulation.com/cgi-bin/mcatch/MatrixCatch.pl>)⁵⁹.

The aa sequences inferred from *Ss-riok-2* and 14 selected sequences from homologues in other invertebrates and vertebrates were subjected to phylogenetic analysis by Clustal X⁶⁰ and manually adjusted. The species selected were five nematodes, including *Ascaris suum* (ADY41687.1), *Caenorhabditis elegans* (NP_493544.2), *Caenorhabditis remanei* (XP_003097720.1), *Chaetomium thermophilum* (XP_006693771.1), *Haemonchus contortus* (ADW23593.1), *Loa loa* (XP_003139548.1), and eight non-nematode species, including *Aedes aegypti* (XP_001655107.1), *Apis florea* (XP_003692456.1), *Canis familiaris* (XP_536291.3, XP_005618121.1), *Drosophila melanogaster* (NP_651365.1), *Dictyostelium discoideum* (XP_640350.1), *Dictyostelium fasciculatum* (XP_004351182.1), *Monodelphis domestica* (XP_001364030.1), *Xenopus laevis* (NP_001086801.1, NP_001088220.1). The phylogenetic analysis was conducted using the neighbour-joining (NJ), maximum parsimony (MP) and maximum likelihood (ML) methods based on the Jones-Taylor-Thornton (JTT) model⁶¹. Confidence limits were assessed by bootstrapping using 1,000 pseudo-replicates for NJ, MP and ML, and other settings were obtained using the default values in MEGA v.5.0⁶¹. A 50% cut-off value was implemented for the consensus tree.

Molecular modeling. The three-dimensional structure of the RIOK-2 kinase from *Chaetomium thermophilum* (*Ct-RIOK-2* kinase; PDB code 4GYG) was used to create a homology model of *Ss-RIOK-2* with the program DeepView Swiss-PdbViewer and the SWISS-MODEL server⁶². The crystal structure of RIOK-2 (native form) from *C. thermophilum* was suitable with 52.19% identity in the ~155 residues (25–321) in *S. stercoralis* (see ref. 15). Remaining residues of *Ss-RIOK-2* (approximately 368 residues) did not show significant similarity and did not contain the sites of ATP binding motif. Hence, this part was not taken for homology modeling. All models were energy minimized by Swiss PDB VIEWER 4.1.0, and structural quality of the refined models was again assessed

by PROCHECK⁶². The structural quality of final model was verified by Verify 3D, ERRAT plot, Procheck, and Ramachandran plot^{63,64}.

Given the importance of magnesium in RIOK-2 catalysis, one magnesium ion was artificially introduced into the empirical phosphate binding site of the model by comparison with the *C. thermophilum* Ct-RIOK-2 structure co-crystallized with ADP (PDB code: 4GYI). Meanwhile an ADP was “borrowed” from *C. thermophilum* kinase Ct-RIOK-2 and deployed in the homology model after structure superimposition.

Binding of ATP to the Ss-RIOK-2 protein kinase active site was evaluated using the docking suite Glide_XP⁶⁵. This software elaborates low-energy poses when searching the conformational space of the ligand while the protein remained rigid based on enthalpy interactions. The possible ionic species states of ATP at pHs ranging from 5.0 to 9.0 were generated by LigPre in Maestro 2015. The Schrödinger suite protein preparation wizard was used to assign atom type and side chain protonation states⁶⁶. A modelled Ss-RIOK-2 protein structure was prepared using the Protein Preparation Wizard panel⁶⁶ prior to use. Docking simulations were run using default parameters, and Glide receptor grids were generated by defining a 10 Å box localized at the centroid of ADP.

Transcript abundance based on RNA-seq analysis. The *S. stercoralis* PV001 line, derived from a single female worm of the UPD strain was used to derive transcriptomic data for the following stages of the parasite: parasitic females (P Female), post-parasitic first-stage larvae (PP L1), post-parasitic third-stage larvae (PP L3), free-living females (FL Female), post free-living first-stage larvae (PFL L1), infective third-stage larvae (iL3), *in vivo* activated third-stage larvae (L3+). These data are publicly available^{67,68}. Transcript abundances were quantified using RNAseq⁶⁷ and calculated using Cufflinks v2.0.2 (<http://cufflinks.cbcb.umd.edu/>) as fragments per kilobase of coding exon per million fragments mapped (FPKM), with paired-end reads counted as single sampling events⁶⁹. FPKM values for coding sequences (CDS), \pm 95% confidence intervals, were calculated for each gene using Cuffdiff v2.0.2^{70,71}. *P*-values \leq 0.05 were considered statistically significant, and *P*-values \leq 0.001 were considered statistically highly significant. FPKM and 95% confidence intervals were plotted in Prism version 5.01 (GraphPad Software, Inc., <http://www.graphpad.com/>).

Protein expression and purification. A full-length cDNA of *Ss-riok-2* was PCR amplified using primers *Ss-riok-2-NdeI* and *Ss-riok-2-XhoI* (S1 Table). The PCR product was then cloned into pMD19-T and sequenced, and further subcloned into the vector pET-42b-*gfp*. The insert of the recombinant plasmid pET-42b-*gfp-riok-2* was sequenced and its open-reading frame (ORF) encoding the fusion protein GFP-His-Ss-RIOK-2 was confirmed. This construct was used to transform *E. coli* (*Transetta*; Transgene) cells for protein expression. The bacterial cells were diluted 1:100 into new LB/Kna⁺ medium, after 3 h of growth at 37 °C. The expression cultures were induced with IPTG (1:1000), grown at 28 °C and 150 rpm/min agitation overnight and then harvested by centrifugation at 10000 rpm for 2 min. The bacteria were re-suspended in 50 mM Tris-Cl with 0.1 M NaCl, passed through a 0.45 μ m filter and loaded onto 1 mL His Trap FF affinity columns (GE Healthcare). The bound Ss-RIOK-2 was eluted with 50 mM Tris-HCl and 40 mM reduced glutathione, pH 8.0. The eluate was concentrated using Ultra-15 50 KD centrifugal filter devices (Millipore). The final concentration was 1 mg/mL.

Kinase assays. The assay for autophosphorylation was carried out in a reaction buffer containing 25 mM Tris pH 7.5, 50 mM NaCl, 2 mM MgCl₂ and 1 μ Ci [γ -³²P] ATP⁷². 10 μ g purified GFP-His-Ss-RIOK-2 were added into the autophosphorylation reaction. Control reactions consisted of all assay components but replacing GFP-His-Ss-RIOK-2 with GFP. The reaction mixtures were incubated at 37 °C for 1 h 30 min. Reactions were terminated by the addition of gel loading buffer and run on SDS polyacrylamide gel for 1 h at 120 volts. Radiolabeled proteins in the gel were detected by autoradiography.

Preparation of transgene constructs. Plasmid vectors for transformation of *S. stercoralis* were prepared by first amplifying the *Ss-riok-2* promoter sequence from *S. stercoralis* genomic DNA with forward and reverse primers including restriction site sequences for *PstI* and *SmaI*, respectively (Thermo). The amplicon was gel-purified using the Tiangen Gel purification kit (Tiangen Biotech). The purified product was then subcloned into the promoter-less vector pAJ01 upstream of the *gfp* coding sequence⁷³ (Addgene, Cambridge, Massachusetts, USA, <http://www.addgene.org>) to create the plasmid pRP2 (Fig. S1). The plasmid was sequenced and transformed into *E. coli* (DH5 α ; Transgene) and amplified in the bacteria cells at 37 °C. After 8 h of growth, the construct was extracted by TIAN pure Midi Plasmid Kit (Tiangen Biotech) and diluted to 25 ng/ μ L and 50 ng/ μ L. The diluted construct was stored at -20 °C.

DNA transformation. Standard methods for transforming *S. stercoralis* by microinjecting plasmid constructs into gonads of free-living females were carried out as described⁷⁴. In brief, free-living *S. stercoralis* females were immobilized on dry agar pads overlain with halocarbon oil, and a solution of 25 ng/ μ L of plasmid *Ss-riok-2p::gfp::Ss-era-1t* (pRP2) in injection buffer was injected into the gonad between the distal tip cell and the bend using finely drawn glass micropipets pressurized with air at 20 psi. Microinjected female worms were transferred singly, along with one or two free-living males, to NGM agar plates with lawns of *E. coli* OP50, and plates were sealed with parafilm and incubated at 22 °C. Microinjected *S. stercoralis* and their progeny were observed daily. Initial screening for *gfp* expression in F1 progeny was done at intervals of 24 h and 48 h following injection using a stereomicroscope outfitted with a coaxial epifluorescence unit with a GFP LP filter. Larvae expressing GFP were anesthetized using 20–50 mM levamisole (Sigma-Aldrich), and transferred to 2% agar pads (Sigma-Aldrich) on standard microscope slides⁶⁸. Worms with GFP expression were examined in detail using a compound epifluorescence microscope equipped with Differential Interference Contrast (DIC) optics and a digital camera (Olympus BX51, Japan).

References

- LaRonde-LeBlanc, N. & Wlodawer, A. The RIO kinases: an atypical protein kinase family required for ribosome biogenesis and cell cycle progression. *Biochimica et biophysica acta* **1754**, 14–24, doi:10.1016/j.bbapap.2005.07.037 (2005a).
- Geerlings, T. H., Faber, A. W., Bister, M. D., Vos, J. C. & Raue, H. A. Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20 S Pre-rRNA in *Saccharomyces cerevisiae*. *The Journal of biological chemistry* **278**, 22537–22545, doi:10.1074/jbc.M300759200 (2003).
- LaRonde-LeBlanc, N. & Wlodawer, A. A family portrait of the RIO kinases. *The Journal of biological chemistry* **280**, 37297–37300, doi:10.1074/jbc.R500013200 (2005b).
- Vanrobays, E., Gelugne, J. P., Gleizes, P. E. & Caizergues-Ferrer, M. Late cytoplasmic maturation of the small ribosomal subunit requires RIO proteins in *Saccharomyces cerevisiae*. *Molecular and cellular biology* **23**, 2083–2095 (2003).
- Rouquette, J., Choismel, V. & Gleizes, P. E. Nuclear export and cytoplasmic processing of precursors to the 40S ribosomal subunits in mammalian cells. *The EMBO journal* **24**, 2862–2872, doi:10.1038/sj.emboj.7600752 (2005).
- Zemp, I. *et al.* Distinct cytoplasmic maturation steps of 40S ribosomal subunit precursors require hRio2. *The Journal of cell biology* **185**, 1167–1180, doi:10.1083/jcb.200904048 (2009).
- LaRonde-LeBlanc, N. & Wlodawer, A. Crystal structure of *A. fulgidus* Rio2 defines a new family of serine protein kinases. *Structure (London, England: 1993)* **12**, 1585–1594, doi:10.1016/j.str.2004.06.016 (2004).
- Chouhan, D. K., Sharon, A. & Bal, C. Molecular and structural insight into plasmodium falciparum RIO2 kinase. *Journal of molecular modeling* **19**, 485–496, doi:10.1007/s00894-012-1572-3 (2013).
- Esser, D. & Siebers, B. Atypical protein kinases of the RIO family in archaea. *Biochemical Society transactions* **41**, 399–404, doi:10.1042/bst20120317 (2013).
- Gajiwala, K. S. *et al.* Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**, 916–921, doi:10.1038/35002634 (2000).
- Dong, G., Chakshumathi, G., Wolin, S. L. & Reinisch, K. M. Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch. *The EMBO journal* **23**, 1000–1007, doi:10.1038/sj.emboj.7600115 (2004).
- Wolberger, C. & Campbell, R. New perch for the winged helix. *Nature structural biology* **7**, 261–262, doi:10.1038/74004 (2000).
- Fischer, U. *et al.* A non-canonical mechanism for Crm1-export cargo complex assembly. *eLife* **4**, 10.7554/eLife.05745 (2015).
- Fung, H. Y., Fu, S. C., Brautigam, C. A. & Chook, Y. M. Structural determinants of nuclear export signal orientation in binding to exportin CRM1. *eLife* **4**, doi:10.7554/eLife.10034 (2015).
- Ferreira-Cerca, S. *et al.* ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nature structural & molecular biology* **19**, 1316–1323, doi:10.1038/nsmb.2403 (2012).
- Liu, T. *et al.* Phosphorylation of right open reading frame 2 (Rio2) protein kinase by polo-like kinase 1 regulates mitotic progression. *The Journal of biological chemistry* **286**, 36352–36360, doi:10.1074/jbc.M111.250175 (2011).
- Read, R. D. *et al.* A kinome-wide RNAi screen in *Drosophila* Glia reveals that the RIO kinases mediate cell proliferation and survival through TORC2-Akt signaling in glioblastoma. *PLoS genetics* **9**, e1003253, doi:10.1371/journal.pgen.1003253 (2013).
- Varin, T. *et al.* Discovery of selective RIO2 kinase small molecule ligand. *Biochimica et biophysica acta* **1854**, 1630–1636, doi:10.1016/j.bbapap.2015.04.006 (2015).
- Trivedi, V. & Nag, S. In silico characterization of atypical kinase PFD0975w from *Plasmodium* kinome: a suitable target for drug discovery. *Chemical biology & drug design* **79**, 600–609, doi:10.1111/j.1747-0285.2012.01321.x (2012).
- Nag, S., Prasad, K., Bhowmick, A., Deshmukh, R. & Trivedi, V. PFRIO-2 kinase is a potential therapeutic target of antimalarial protein kinase inhibitors. *Current drug discovery technologies* **10**, 85–91 (2013).
- Marcos, L. A., Terashima, A., Dupont, H. L. & Gotuzzo, E. Strongyloides hyperinfection syndrome: an emerging global infectious disease. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **102**, 314–318, doi:10.1016/j.trstmh.2008.01.020 (2008).
- Olsen, A. *et al.* Strongyloidiasis—the most neglected of the neglected tropical diseases? *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103**, 967–972, doi:10.1016/j.trstmh.2009.02.013 (2009).
- Hunt, V. L. *et al.* The genomic basis of parasitism in the *Strongyloides* clade of nematodes. **48**, 299–307, doi:10.1038/ng.3495 (2016).
- Lok, J. B., Shao, H., Massey, H. C. & Li, X. Transgenesis in *Strongyloides* and related parasitic nematodes: historical perspectives, current functional genomic applications and progress towards gene disruption and editing. *Parasitology* **1–16**, 10.1017/s0031182016000391 (2016).
- Granneman, S., Petfalski, E., Swiatkowska, A. & Tollervey, D. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *The EMBO journal* **29**, 2026–2036, doi:10.1038/emboj.2010.86 (2010).
- Simpson, K. J. *et al.* Identification of genes that regulate epithelial cell migration using an siRNA screening approach. *Nature cell biology* **10**, 1027–1038, doi:10.1038/ncb1762 (2008).
- Ceron, J. *et al.* Large-scale RNAi screens identify novel genes that interact with the *C. elegans* retinoblastoma pathway as well as splicing-related components with synMuv B activity. *BMC developmental biology* **7**, 30, doi:10.1186/1471-213x-7-30 (2007).
- LaRonde-LeBlanc, N., Guszczynski, T., Copeland, T. & Wlodawer, A. Autophosphorylation of *Archaeoglobus fulgidus* Rio2 and crystal structures of its nucleotide-metal ion complexes. *The FEBS journal* **272**, 2800–2810, doi:10.1111/j.1742-4658.2005.04702.x (2005).
- Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420, doi:10.1038/364412a0 (1993).
- Wang, J. C. & Van Dyke, M. W. Initiator sequences direct downstream promoter binding by human transcription factor IID. *Biochimica et biophysica acta* **1216**, 73–80 (1993).
- Tokusumi, Y., Ma, Y., Song, X., Jacobson, R. H. & Takada, S. The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters. *Molecular and cellular biology* **27**, 1844–1858, doi:10.1128/mcb.01363-06 (2007).
- Lowry, J. A. & Atchley, W. R. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *Journal of molecular evolution* **50**, 103–115 (2000).
- LaVoie, H. A. The role of GATA in mammalian reproduction. *Experimental biology and medicine (Maywood, N.J.)* **228**, 1282–1290 (2003).
- Pikkarainen, S., Tokola, H., Kerkela, R. & Ruskoaho, H. GATA transcription factors in the developing and adult heart. *Cardiovascular research* **63**, 196–207, doi:10.1016/j.cardiores.2004.03.025 (2004).
- Bresnick, E. H., Martowicz, M. L., Pal, S. & Johnson, K. D. Developmental control via GATA factor interplay at chromatin domains. *Journal of cellular physiology* **205**, 1–9, doi:10.1002/jcp.20393 (2005).
- Murakami, R., Okumura, T. & Uchiyama, H. GATA factors as key regulatory molecules in the development of *Drosophila* endoderm. *Development, growth & differentiation* **47**, 581–589, doi:10.1111/j.1440-169X.2005.00836.x (2005).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389–3402 (1997).
- Massari, M. E. & Murre, C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and cellular biology* **20**, 429–440 (2000).
- Kageyama, R., Ohtsuka, T., Hatakeyama, J. & Ohsawa, R. Roles of bHLH genes in neural stem cell differentiation. *Experimental cell research* **306**, 343–348, doi:10.1016/j.yexcr.2005.03.015 (2005).

40. McMiller, T. L. & Johnson, C. M. Molecular characterization of HLH-17, a *C. elegans* bHLH protein required for normal larval development. *Gene* **356**, 1–10, doi:10.1016/j.gene.2005.05.003 (2005).
41. Zhao, J., Wang, P. & Corsi, A. K. The *C. elegans* Twist target gene, *arg-1*, is regulated by distinct E box promoter elements. *Mechanisms of development* **124**, 377–389, doi:10.1016/j.mod.2007.01.005 (2007).
42. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479, doi:10.1038/nature09715 (2011).
43. Ghedin, E. *et al.* Draft genome of the filarial nematode parasite *Brugia malayi*. *Science (New York, N.Y.)* **317**, 1756–1760, doi:10.1126/science.1145406 (2007).
44. Schwarz, E. M. *et al.* The genome and developmental transcriptome of the strongyloid nematode *Haemonchus contortus*. *Genome biology* **14**, R89, doi:10.1186/gb-2013-14-8-r89 (2013).
45. Laing, R. *et al.* The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. *Genome biology* **14**, R88, doi:10.1186/gb-2013-14-8-r88 (2013).
46. McGhee, J. D. *et al.* ELT-2 is the predominant transcription factor controlling differentiation and function of the *C. elegans* intestine, from embryo to adult. *Developmental biology* **327**, 551–565, doi:10.1016/j.ydbio.2008.11.034 (2009).
47. McGhee, J. D. *et al.* The ELT-2 GATA-factor and the global regulation of transcription in the *C. elegans* intestine. *Developmental biology* **302**, 627–645, doi:10.1016/j.ydbio.2006.10.024 (2007).
48. Kimble, J. & Sharrock, W. J. Tissue-specific synthesis of yolk proteins in *Caenorhabditis elegans*. *Developmental biology* **96**, 189–196 (1983).
49. Schulenburg, H., Kurz, C. L. & Ewbank, J. J. Evolution of the innate immune system: the worm perspective. *Immunological reviews* **198**, 36–58 (2004).
50. Pauli, F., Liu, Y., Kim, Y. A., Chen, P. J. & Kim, S. K. Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in *C. elegans*. *Development (Cambridge, England)* **133**, 287–295, doi:10.1242/dev.02185 (2006).
51. McGhee, J. D. The *C. elegans* intestine. *WormBook: the online review of C. elegans biology* 1–36, doi:10.1895/wormbook.1.133.1 (2007).
52. Lin, K., Hsin, H., Libina, N. & Kenyon, C. Regulation of the *Caenorhabditis elegans* longevity protein DAF-16 by insulin/IGF-1 and germline signaling. *Nature genetics* **28**, 139–145, doi:10.1038/88850 (2001).
53. Bossinger, O. & Schierenberg, E. The use of fluorescent marker dyes for studying intercellular communication in nematode embryos. *The International journal of developmental biology* **40**, 431–439 (1996).
54. Lok, J. B. *Strongyloides stercoralis*: a model for translational research on parasitic nematode biology. *WormBook: the online review of C. elegans biology* 1–18, doi:10.1895/wormbook.1.134.1 (2007).
55. Gasser, R. B., Chilton, N. B., Hoste, H. & Beveridge, I. Rapid sequencing of rDNA from single worms and eggs of parasitic helminths. *Nucleic acids research* **21**, 2525–2526 (1993).
56. Bairoch, A. The PROSITE dictionary of sites and patterns in proteins, its current status. *Nucleic acids research* **21**, 3097–3103 (1993).
57. Bateman, A. *et al.* The Pfam protein families database. *Nucleic acids research* **28**, 263–266 (2000).
58. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772–780, doi:10.1093/molbev/mst010 (2013).
59. Deyneko, I. V. *et al.* MatrixCatch—a novel tool for the recognition of composite regulatory elements in promoters. *BMC bioinformatics* **14**, 241, doi:10.1186/1471-2105-14-241 (2013).
60. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)* **23**, 2947–2948, doi:10.1093/bioinformatics/btm404 (2007).
61. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**, 2731–2739, doi:10.1093/molbev/msr121 (2011).
62. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723, doi:10.1002/elps.1150181505 (1997).
63. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science: a publication of the Protein Society* **2**, 1511–1519, doi:10.1002/pro.5560020916 (1993).
64. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology* **7**, 95–99 (1963).
65. Friesner, R. A. *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of medicinal chemistry* **49**, 6177–6196, doi:10.1021/jm051256o (2006).
66. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* **27**, 221–234, doi:10.1007/s10822-013-9644-8 (2013).
67. Stoltzfus, J. D., Minot, S., Berriman, M., Nolan, T. J. & Lok, J. B. RNAseq analysis of the parasitic nematode *Strongyloides stercoralis* reveals divergent regulation of canonical dauer pathways. *PLoS neglected tropical diseases* **6**, e1854, doi:10.1371/journal.pntd.0001854 (2012b).
68. Stoltzfus, J. D., Massey, H. C. Jr., Nolan, T. J., Griffith, S. D. & Lok, J. B. *Strongyloides stercoralis* age-1: a potential regulator of infective larval development in a parasitic nematode. *PloS one* **7**, e38587, doi:10.1371/journal.pone.0038587 (2012a).
69. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**, 621–628, doi:10.1038/nmeth.1226 (2008).
70. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* **31**, 46–53, doi:10.1038/nbt.2450 (2013).
71. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**, 562–578, doi:10.1038/nprot.2012.016 (2012).
72. Laronde-Leblanc, N., Guszczynski, T., Copeland, T. & Wlodawer, A. Structure and activity of the atypical serine kinase Rio1. *The FEBS journal* **272**, 3698–3713, doi:10.1111/j.1742-4658.2005.04796.x (2005).
73. Junio, A. B. *et al.* *Strongyloides stercoralis*: Cell- and tissue-specific transgene expression and co-transformation with vector constructs incorporating a common multifunctional 3' UTR. *Experimental parasitology* **118**, 253–265, doi:10.1016/j.exppara.2007.08.018 (2008).
74. Lok, J. B. & Massey, H. C. Jr. Transgene expression in *Strongyloides stercoralis* following gonadal microinjection of DNA constructs. *Molecular and biochemical parasitology* **119**, 279–284 (2002).

Acknowledgements

This work was supported by “Fundamental Research Funds for the Central Universities” (Program No. 2662015PY180) and “the National Key Basic Research Program (973 program) of China” (Grant No. 2015CB150300) to MH, by a grant from The National Institute of Health (NIH), USA (AI-50688) to JBL and by funds from Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program No. 2015RC005) the National Health and Medical Research Council (NHMRC) and the Australian Research Council (ARC) of Australia to RBG.

Author Contributions

M.H. conceived and designed the experiments. W.Q.L., W.Y., Y.Z.Z., J.D.S. performed the experiments and analyzed the data. S.Y.H., H.Z., R.Z., J.L.Z. coordinated the research. W.Q.L., J.B.L., M.H. and R.B.G. wrote the paper.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-07991-2](https://doi.org/10.1038/s41598-017-07991-2)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017