The Transcriptome Analysis of *Strongyloides stercoralis* L3i Larvae Reveals Targets for Intervention in a Neglected Disease

Antonio Marcilla1,*, Gagan Garg2, Dolores Bernal3, Shoba Ranganathan2,4,*, Javier Forment5, Javier Ortiz6, Carla Muñoz-Antolí1, M. Victoria Domínguez7, Laia Pedrola8, Juan Martínez-Blanch8, Javier Sotillo1, Maria Treliš5, Rafael Toledo3, J. Guillermo Esteban1

1 Área de Parasitología, Departamento de Biología Celular y Parasitología, Universitat de València, Burjassot, Valencia, Spain, 2 Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales, Australia, 3 Departamento de Bioquímica y Biología Molecular, Universitat de València, Burjassot, Valencia, Spain, 4 Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, 5 Servicio de Bioinformática, Instituto de Biología Molecular y Celular de Plantas, Universitat Politècnica de València, Ingeniero Fausto Elio, Valencia, Spain, 6 Unidad de Bioinformática, SCSIE, Universitat de València, Burjassot, Valencia, Spain, 7 Servicio de Microbiología y Parasitología, Hospital General de Castellón, Castellón, Spain, 8 LifeSequencing S.L., Parc Científic Universitat de València, Paterna, Valencia, Spain

Abstract

**Background:** Strongyloidiasis is one of the most neglected diseases distributed worldwide with endemic areas in developed countries, where chronic infections are life threatening. Despite its impact, very little is known about the molecular biology of the parasite involved and its interplay with its hosts. Next generation sequencing technologies now provide unique opportunities to rapidly address these questions.

**Principal Findings:** Here we present the first transcriptome of the third larval stage of *S. stercoralis* using 454 sequencing coupled with semi-automated bioinformatic analyses. 253,266 raw sequence reads were assembled into 11,250 contiguous sequences, most of which were novel. 8037 putative proteins were characterized based on homology, gene ontology and/or biochemical pathways. Comparison of the transcriptome of *S. stromyloides* with those of other nematodes, including *S. ratti*, revealed similarities in transcription of molecules inferred to have key roles in parasite-host interactions. Enzymatic proteins, Like kinases and proteases, were abundant. 1213 putative excretory/secretory proteins were compiled using a new pipeline which included non-classical secretory proteins. Potential drug targets were also identified.

**Conclusions:** Overall, the present dataset should provide a solid foundation for future fundamental genomic, proteomic and metabolomic explorations of *S. stercoralis*, as well as a basis for applied outcomes, such as the development of novel methods of intervention against this neglected parasite.

Citation: Marcilla A, Garg G, Bernal D, Ranganathan S, Forment J, et al. (2012) The Transcriptome Analysis of *Strongyloides stercoralis* L3i Larvae Reveals Targets for Intervention in a Neglected Disease. PLoS Negl Trop Dis 6(2): e1513. doi:10.1371/journal.pntd.0001513

Editor: John Pius Dalton, McGill University, Canada

Received August 23, 2011; Accepted December 20, 2011; Published February 28, 2012

Copyright: © 2012 Marcilla et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by project PS09/02355 from the Fondo de Investigación Sanitaria (FIS), Spanish Ministry of Science and Innovation (Madrid, Spain) and FEDER and project PROMETEO/2009/081 from Conselleria d’Educaçion, Generalitat Valenciana (Valencia, Spain). GG would like to acknowledge Macquarie University for an Australian Post-graduate Award scholarship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: antonio.marcilla@uv.es (AM); shoba.ranganathan@mq.edu.au (SR)

Introduction

Strongyloidiasis caused by *Strongyloides stercoralis* is a soil-transmitted helminthiasis distributed worldwide, affecting more than 100 million people, with endemic areas in Southeast Asia, Latin America, sub-Saharan Africa, and parts of the southeastern United States [1,2]. Recently, it was classified as one of the most neglected tropical diseases (NTD) [3]. Chronic infections in endemic areas may be maintained asymptptomatically for decades through the autoinfective cycle with the filariform larvae L3 [1][4,5]. The diagnosis of these chronic infections requires more sensitive diagnostic methods, particularly in low-level infections and immunocompromised patients [1].

Epidemiological studies in developed countries have identified endemic areas where misdiagnosis, inadequate treatment and the facilitation of hyperinfection syndrome by immunosuppression (i.e. by the administration of steroids) are too frequent and can cause a high mortality rate ranging from 15 to 87% [5,6]. Among these areas, an endemic area with chronic patients have been described at the Valencian Mediterranean coastal region of Spain related to environmental conditions [7].

The diagnosis of strongyloidiasis is suspected when clinical signs and symptoms, or eosinophilia is observed [8], but current definitive diagnosis of strongyloidiasis is usually made on the basis of detection of larvae in agar plate coproculture and serological diagnosis by ELISA [9,10]. Those methods have the drawbacks of

www.plosntds.org 1 February 2012 | Volume 6 | Issue 2 | e1513
Author Summary

Strongyloides stercoralis (Nematoda) is an important parasite of humans, causing Strongyloidiasis, considered as one of the most neglected diseases, affecting more than 100 million people worldwide. Chronic infections in endemic areas can be maintained for decades through the autoinfective cycle with the L3 filariform larvae. In these areas, misdiagnosis, inadequate treatment and the facilitation of hyperinfection syndrome by immunosuppression are frequent and contribute to a high mortality rate. Among the affected areas, chronic patients have been described in the Valencian Mediterranean coastal region of Spain. Despite its serious impact, very little is known about this parasite and its relationship with its host at the molecular level, and more effective diagnostic tests and treatments are needed. Next generation sequencing technologies now provide unique opportunities to rapidly advance in these areas. In this study, we present the first transcriptome of S. stercoralis L3i using 454 sequencing followed by semi-automated bioinformatic analyses. Our study identifies 8037 putative proteins based on homology, gene ontology, and/or biochemical pathways, including putative excretory/secretory proteins as well as potential drug targets. The present dataset provides a useful resource and adds greatly to our understanding of a human parasite affecting both developed and developing countries.

being time consuming and requiring expertise in the first case, and of low specificity due to remaining antibodies from previous infection or cross-reactive antibodies [11]. A recent paper has described a promising copropoantigen ELISA based on a polyclonal rabbit antiserum raised against excretory/secretory (ES) antigens from the closely related Strongyloides ratti [12], but the identification of S. stercoralis specific ES proteins that could be new potential targets for diagnosis is still required.

Control of strongyloidiasis has relied mostly on the treatment of infected individuals with only three anthelmintic drugs: thiabendazole (no longer available), albendazole, and more recently ivermectin [3,13]. A recent study by Suputtamongkol et al. (2011) [14] has confirmed that both a single and double dose of oral ivermectin are more effective than a 7-day course of high dose albendazole for patients with chronic infection due to S. stercoralis. The risk of developing genetic resistance against the current drugs administered (if used excessively and at suboptimal dosages) exists and is based on the experience with drug resistance in parasitic nematodes of livestock [15]. Thus, the current focus is on the discovery of novel drugs against human parasites like S. stercoralis. Such a discovery effort could be strengthened with an integrated genomic and bioinformatics approach, using functional genomic and phenomic information available for the free-living nematode Caenorhabditis elegans (see WormBase; www.wormbase.org). This nematode, which is the best characterized metazoan organism [16,17], is considered to be related to nematodes of the order Strongyloida (to which Strongyloides belong) [18]. Recent studies have reported that nearly 60% of genes in strongyloides have orthologues/homologues in C. elegans, with a wide range of biological pathways being conserved between parasitic nematodes and C. elegans [19]. The comparison of molecular data sets between nematodes should therefore allow the identification of specific biological pathways as potential new targets for nematocidal drugs [20].

As pointed out recently by Cantacessi et al. (2011) [20], advances in genomic sequencing like Next Generation Sequencing (NGS) and annotation as well as the integrated use of ‘-omic’ technologies are now shedding light on our understanding of the systems biology of nematodes on an unprecedented scale, and is likely to provide unique opportunities for the development of entirely new strategies for the treatment and control of neglected parasitic diseases. New bioinformatic tools based on robust assembly protocol for NGS data, along with compilation of a dataset of experimentally determined ES proteins of parasitic helminths, and annotation software like KAAS [21], allow efficient and up-to-date homology-based predictions [22].

To date, there are few molecular and genomic studies on Strongyloides species, and only the transcriptome from S. ratti adults has become recently available (http://worm1.liv.ac.uk/file_summary.html) [23]. In fact, 39166 ESTs are currently available in the NCBI database of November 2011 (27366 from S. ratti and 11392 from S. stercoralis). Yoshida et al. (2011) have obtained 162 unique singletons and contigs from S. venezuelensis [24], and a recent study by Ramanathan et al. (2011) has described DNA microarray for S. stercoralis and used them to compare infective third-stage larvae (L3) with non-infective first stage larvae (L1), with 935 differentially expressed genes identified [25].

In the present study, we have explored and functionally annotated the transcriptome of L3i of S. stercoralis by 454 sequencing coupled to semi-automated bioinformatic analyses and predicted potential therapeutic targets for strongyloidiasis.

Materials and Methods

Accession numbers

The nucleotide sequence data obtained for this study are available in the GenBank database under accession number ERP000798.

The assembled data from this study can be requested from the corresponding author.

Parasite material and ethical issues

Fecal samples were obtained at the Hospital La Ribera, Alzira, Valencia (Spain) from an infected individual in compliance with Spanish ethical regulations [7], and approved by the Ethics Committee in human research from the Universitat de Valencia. Oral consent from the patient was obtained (she was happy to participate in the study but felt uncomfortable with signing a form), and documented as a tick on the case record form following the Hospital Reviewing Board protocols. Samples were cultured on Agar Petri dishes and L3i larvae were harvested and concentrated by centrifugation for 5 min at 1000 g, washed three times in 1 ml of phosphate buffered saline (PBS) pH 7.2 containing protease inhibitors (10 mM EDTA, and 1 mM PMSF) and samples were processed for RNA isolation.

RNA isolation, cDNA synthesis and 454 sequencing

Total RNA from around 500 larvae was prepared using Vantage™ Total RNA purification kit (Marligen Biosciences, Ijamsville, MD, USA) following the manufacturers’ instructions and treated with Ambion DNA-free™ DNase (Ambion/Applied Biosystems, Austin, TX). The integrity of the RNA was verified by gel electrophoresis and the yield determined using the nanoDrop ND-1000 UV-VIS spectrophotometer v.3.2.1 (NanoDrop Technologies, Wilmington, DE).

The cDNA library was constructed from 0.5 µg total RNA using MINT cDNA Synthesis Kit (Cat.#SK001, Evrogen). First strand cDNA synthesis starts from 3’-primer comprising oligo(dT) to enrich mRNA as template. Double strand cDNA synthesis was performed using 17 cycles of PCR amplification. Total cDNA was
digested with restriction enzyme *Gall* in order to remove Poly (A) tails. cDNA obtained was used to perform a library with the required sequencing adaptors and was then sequenced using the Genome Sequencer (GS) FLX instrument (Roche Diagnostics) [26].

**Bioinformatic analyses of sequence data**

The overall bioinformatic analysis strategy followed was as described originally by Nagaraj et al. [27,28], implemented in the analysis pipelines ESTExplorer [29] and EST2Secretome [30]. This workflow approach has been successfully used for the analysis of transcriptomic data from *Dictyocaulus viviparus* [30], Fasciola hepatica [31], Clonorchis sinensis [32] and Opisthorchis viverrini [32]. However, to better identify non-classically secreted proteins from helminth parasites [33,34], we have recently implemented a novel analysis strategy for short reads applied on Strongyloides ratti [22] (see Figure S1).

FASTA and associated quality files were extracted from the SFF file after removing the sequence adapters. These reads were preprocessed and their contigs were assembled using MIRA v.3.2 (http://chevreux.org/projects_mira.html) [35] with the following parameters:

- `job=denovo.est,accurate;454 -fasta`
- `-OUT:rrol=1:lor=1:org=0:os=0:osf=0:otf=0:osd=0`
- `-GE:asr=1`
- `-LR:mxl=1 -AS:sd=0 -as=0 -ard=1 -SK:mmhr=2`
- `-nn:mm=ye 454_SETTINGS`
- `-DP:are=0 -CO:mpg=10 -AS:bdg=40`
- `-CP:cvdc=0:mbc=1:mbcmg=30:mbcmeg=30:ode=0 -ED:ace=1`
- `-AL:zegp=no`
- `-AL:GN:bi=pa=20 -bmax=120 -mo=10`

Contigs generated from MIRA were aligned and reassembled into second order contigs using the Contig Assembly Program v.3 (CAP3) [36], employing a minimum sequence overlap length cutoff of 40 bases and an identity threshold of 90%. Following the assembly of *S. stercoralis* reads into second order contigs by CAP3 and contigs by MIRA, this contig dataset was matched using BLASTX with the NCBI non-redundant sequence database; http://www.ncbi.nlm.nih.gov, BLASTN with Nematode.net *S. stercoralis* ESTs (www.nematode.net/) and BLASTN with dbEST *Strongyloides* ESTs (www.ncbi.nlm.nih.gov/dbEST) using permissive (E-value: <1E-05), moderate (<1E-15) and/or stringent (<1E-30) search strategies.

*S. stercoralis* contigs were conceptually translated into putative proteins using the program ESTScan [37]. Putative protein sequences were subjected to secretome analysis using TMHMM (a membrane topology prediction program) [38] to predict transmembrane domains, SignalP 3.0 (signal peptide prediction program) [39], SecretomeP (a prediction programme used to identify non-classical secretory proteins in mammals [40], but used in the case of parasitic helminths as well [41]), and TargetP (mitochondrial protein prediction program) [42]. Briefly, excretory/secretory (ES) proteins were selected based on the presence of a signal peptide at the N-terminus using SignalP 3.0 (employing both the neural network and hidden Markov models) or predicted as secretory using SecretomeP, predicted as non-mitochondrial by TargetP and absence of transmembrane domains. In addition to computational prediction of ES proteins were identified and collated based on sequence homology (BLASTP, E-value<1E-15) to known ES proteins found in parasitic helminths secretome studies.

Putative proteins were classified functionally using InterProScan [43], employing the default search parameters. Based on their homology to conserved domains and protein families, predicted proteins were classified into Gene Ontology (GO) categories (http://www.geneontology.org/) based on molecular function, cellular component and biological process using interpro terms. Putative proteins were also subjected to pathway analysis, utilizing KEGG-Automatic Annotation Server (KAAS) [21], which maps the putative proteins to biochemical pathways in which they are involved and categories of Brite objects like enzymes, transcription factors and translation factors.

Putative proteins were subjected to BLAST2GO software to identify homologues from the most abundant ES transcripts [44]. BLASTP (Wormpep v 224) was used to identify *C. elegans* known proteins homologues present in *S. stercoralis* proteins using moderate search strategy (E-value: <1E-15). These proteins were also searched for sequence homology (BLASTP, E-value<1E-05) in human (host) proteins. All the proteins which were found homologous to *C. elegans* proteins and non-homologous to human proteins were mapped to *C. elegans* RNAi phenotypes and known drug targets present in the DrugBank database (http://drugbank.ca/), a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information [45].

**Results**

The transcriptome of *S. stercoralis* L3i larvae

Initially a total of 253266 short reads (82490223 bases) were generated with 325±132.4 bases (average length ± standard deviation), with a GC content of 31.84%. These short reads were pre-processed, which resulted in 237341 (93.7%) quality short reads (EBI Sequence Read Archive [SRA] accession ID ERP000798). High quality reads were assembled into 12333 contigs using MIRA as described in the pipeline (Figure S1). Using CAP3, we were able to achieve 507 second order contigs, leaving 10845 *MIRA* contigs not assembled further by CAP3. We considered 11250 (99.1%) contigs with a minimum length of 90 bases, discarding sequences yielding peptides <30 amino acids, for further secretory protein prediction and analysis. These contigs were conceptually translated into 8037 proteins by ESTScan (Table 1; sequences available from http://biolinux.uv.es/marcilla/).

Putative proteins were annotated based on protein families and domains using Interproscan and mapped to biochemical pathways using KAAS [21]. Of the 8037 putative proteins, we were able to annotate 4494 (55.91%) proteins with protein domains and

| Table 1. Expressed sequence tag (EST) data for the *S. stercoralis* L3i. |
|--------------------------------------------------------------|
| **No. of EST Clusters by MIRA** | 12,333 |
| **Average length (± standard deviation)** | 538.7 (± 260.7) |
| **Re clustered unigenes after CAP3 assembly** | 11,352 |
| **Containing an Open Reading Frame** | 8,037 |
| **Returning InterProScan results** | 4,494 |
| **Gene Ontology** | 3,534 |
| **Biological process** | 1,905 |
| **Cellular component** | 1,068 |
| **Molecular function** | 3,083 |
| **Prediction of biological pathways (KAAS)** | 1,559 |

doi:10.1371/journal.pntd.0001513.t001
families (Table 1). The most represented Interpro terms are shown in Table S1.

A total of 3534 proteins were annotated with GO terms (3083 (Molecular Function), 1068 (Cellular Component) and 1905 (Biological Process)) based on Interpro term annotations (Tables 1 and S1). We established pathway associations for 1559 (19.39%) putative proteins (Table 1).

All the contigs generated by using MIRA+CAP3 were checked for homologous proteins against the non-redundant nucleotide database (NR-NCB) including Strongyloides expressed sequence tags (ESTs) present in dbEST, S. stercoralis ESTs available from dbEST and nematode.net, S. ratti cDNA sequencing data from the University of Liverpool (available at http://worm1.liv.ac.uk/file_summary.html), and also for homologous proteins in C. elegans and human data (Figure S1). Similarity searches were done using using BlastX and BlastP algorithms at different E values (Table 2). A total of 3412 (42.45%) S. stercoralis putative proteins had homologues in the free-living nematode, Caenorhabditis elegans using stringent match conditions (E value: <1E-15). The recent availability of S. ratti transcriptome data prompted us to compare these with our data and 3055 similar putative proteins (47.96%) were found. As S. stercoralis infects humans, we checked the similarity of S. stercoralis proteins with known human proteins using BlastP at different E values. Our results showed that 3759 putative proteins were similar to human ones using a permissive search strategy (E-value: <1E-05), discarding them as potential targets for treatment (Table 2).

Predicted proteins were also categorized according to their inferred molecular function, cellular localization and association with biological pathways. Mapping to KEGG BRITE objects [46] is shown in Table 3. Enzymes were by far the most abundant category, with 720 putative proteins, followed by chromosome, spliceosome and ribosome components (with 90, 89 and 73 putative proteins, respectively). 73 putative protein kinases and 72 peptidases were also identified by BRITE (Table 3). These 72 peptidases corresponded to 60 different enzymes from 9 groups, including calpains, cathepsins, different proteasome components and aminopeptidases, and other “nematode common” proteases such as astacin, legumain, and insulinB (Table S2).

All the putative proteins were grouped according to KEGG pathways [46] into five categories, with metabolic proteins being the most abundant, followed by genetic information processing, environmental processing and cellular processes (Table 4). In the first group, the most abundant putative proteins were related to carbohydrate metabolism (201 proteins, 2.5%), amino acid metabolism (174; 2.16%) and lipid metabolism (104; 1.29%). Also 25 putative proteins were related to drug metabolism (Table 4). In the second group, the most abundant proteins were related to translation (195; 2.42%), meanwhile 144 putative proteins (1.79%) related to signal transduction were the most abundant in the group of cellular processes (Table 4).

### Prediction of ES proteins

We next analyzed ES proteins, which are key molecules to understand host-parasite interactions [47]. Molecules from the secretome contribute to important processes like parasite feeding, tissue penetration or larval migration, and could participate in blocking and/or evading host immune responses [48]. ES prediction was carried out in Phase III of the pipeline (Fig. S1). Firstly, 247 (3.07%) proteins were predicted as classical secreted proteins using SignalP [39]. The remaining 7785 (96.93%) proteins, which were predicted as non-secretory by SignalP were processed by SecretomeP [40] for prediction of non-classical secretory proteins, with 252 (3.14%) proteins identified here. The classical and non-classical secretory proteins (499, 6.21%) from these two programs were analyzed by TargetP [42] for mitochondrial proteins. Only 7 proteins were predicted as mitochondrial proteins using TargetP at 95% specificity. These seven proteins were removed from the set of 499 secreted proteins, with 492 secretory proteins passed to TMHMM [38] for the prediction of transmembrane proteins. 161 (2%) proteins, were predicted as transmembrane proteins, having one or more transmembrane helices, and removed from the secretory protein dataset. A total of 331 (4.12%) proteins were finally predicted as ES proteins from the computational prediction pipeline. Proteins that were considered non-secretory by SecretomeP and SignalP were matched to our in-house dataset of 1080 non-redundant experimentally determined parasitic helminth proteins [22] using the BLASTP [49] similarity search. We found an additional 882 (10.97%) putative proteins similar to known ES proteins by this homology search approach (E value: <1E-15) (Table S3). From those proteins, 50 have been recently described in the ES from infective larvae of the related species S. ratti [50] (data not shown).

Among the most abundant transcripts encoding ES proteins appeared a major antigen; cytoskeletal proteins like myosin heavy chain, troponin, tropomyosin, actin; galectins; enzymes like trehalase, PEPCK, GAPDH, enolase, as well as phosphatases and kinases; proteases like Metalloproteinase, Calpain-1 and Cathepsin L; stress proteins like HSPs; calcium binding proteins; detoxifying enzymes along with elongation factors, histones, ubiquitins and signaling molecules (Table S4). Thus, for annotation and analyses in Phase III, we compiled a total of 1213 ES proteins, which is 15.09% of our putative proteins.

### S. stercoralis proteins as drug targets

We found 4234 (52.68%) S. stercoralis putative proteins which had no homologues present in humans (Table 2) and therefore are

---

**Table 2.** Sequence homology inferred between *S. stercoralis* current dataset and other datasets.

| Dataset                        | Hits (1e-05) | Hits (1e-15) | Hits (1e-30) |
|--------------------------------|-------------|-------------|-------------|
| *Strongyloides* EST from dbEST (BLASTN) | 4267        | 3935        | 3646        |
| *S. stercoralis* EST from dbEST (BLASTN) | 2903        | 2682        | 2519        |
| *S. stercoralis* EST from nematode.net (BLASTN) | 2225        | 2020        | 1904        |
| *S. ratti* putative proteins (BLASTP) | 3855        | 3052        | 2081        |
| *C. elegans* proteins (BLASTP) | 4475        | 3,412       | 2163        |
| Human proteins (BLASTP) | 3759        | 2583        | 1489        |
| NR database (BLASTX) | 7633        | 5948        | 4023        |

doi:10.1371/journal.pntd.0001513.t002
Table 3. Functions of putative proteins inferred from the transcriptome of the *S. stercoralis* L3i.

| BRITE object name                     | Number of putative proteins mapped |
|----------------------------------------|------------------------------------|
| Enzymes                                | 720                                |
| Chromosome                             | 90                                 |
| Spliceosome                            | 89                                 |
| Ribosome biogenesis                    | 73                                 |
| Protein kinases                        | 73                                 |
| Peptidases                             | 72                                 |
| Ubiquitin system                       | 72                                 |
| Ribosome                               | 72                                 |
| Chaperones and folding catalysts       | 61                                 |
| Cytoskeleton proteins                  | 53                                 |
| Proteasome                             | 41                                 |
| Ion Channels                           | 38                                 |
| Transcription factors                  | 36                                 |
| Translation factors                    | 34                                 |
| DNA repair and recombination proteins  | 32                                 |
| GTP-binding proteins                   | 25                                 |
| DNA replication proteins               | 20                                 |
| Glycosyltransferases                   | 15                                 |
| Lipid biosynthesis proteins            | 14                                 |
| Cellular antigens                      | 11                                 |
| Transporters                           | 11                                 |
| Secretion system proteins              | 10                                 |
| Glycan binding proteins                | 9                                  |
| SNAREs                                 | 7                                  |
| Nuclear receptors                      | 6                                  |
| Prenyltransferases                     | 6                                  |
| G protein coupled receptors            | 4                                  |
| CAM ligands                            | 4                                  |
| Proteoglycans                          | 4                                  |
| Enzyme linked receptors                | 2                                  |
| Cytokine receptors                     | 2                                  |
| Cell adhesion molecules (CAMs)         | 2                                  |
| Bacterial toxins                       | 1                                  |

| doi:10.1371/journal.pntd.0001513.t003 |

preferred targets for parasite intervention strategies. These human dissimilar proteins of *S. stercoralis* were checked for known drug targets, which have lethal RNAi phenotypes present in *C. elegans*, not present in human and similar to known drug targets, data available from DrugBank 3.0 database [44], a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e., chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e., sequence, structure, and pathway) information. The database (available at http://drugbank.ca/) contains 6707 drug entries (as of November 2011).

We found 14 contigs and singletons corresponding to four different proteins. These could represent potential therapeutic targets for strongyloidiasis as shown in Table S5. Sequence comparison demonstrate that these proteins are homologous to 2,3-bisphosphoglycerate independent phosphoglycerate mutase from *Ascaris suum* (with 1 contig and 1 singleton), hypothetical protein CBG01975 from *Caenorhabditis briggsae* similar to glutamate synthase [NADPH] from *Ascaris suum* (1 contig and 3 singletons), isocitrate lyase from *S. stercoralis* (5 singletons), and alcohol dehydrogenase I from the fungus *Candida albicans* WO-1 (2 singletons) (Table 5).

With a comparative analysis searching protein domain mapping or sequence similarity with other drug targets, we found seven additional potential targets for treatment, including well known drug targets as tubulin β, γ-amino butyric acid A (GABA) receptor, glutamate-gated chloride channel or GST (Table S5). Only one of those proteins, homologous to *Ancylostoma caninum* metalloprotease precursor, which is also predicted to be secretory, was not found similar either to *C. elegans* or human proteins (Table S5).

Discussion

*Strongyloides stercoralis* can replicate within the host (autoinfection) allowing the infection to remain undiagnosed and untreated for years, resulting in perpetuating parasite dispersal, increasing the risk of infection and eventually the appearance of resistances [3]. Uncontrolled multiplication of the parasite (hyperinfection) can be life-threatening in immunocompromised individuals. We also face serious endemic recurring infections in the future if this infection is not controlled in transition economies like China, India, Southeast Asia and Latin America [3] where the use of immunosuppressive therapy is becoming common. As pointed out by Ohren et al. (2009) [3], there is an urgent need to employ modern molecular methods to improve and simplify diagnosis, differentiate species and strains to facilitate epidemiological studies of *S. stercoralis*.

The present study provides the first detailed analysis of the transcriptome of the human pathogenic *S. stercoralis* L3i larvae and has identified specific molecules predicted to play key biological functions in this parasite. A total of 12,333 contigs were inferred from the present EST dataset, thus increasing the number of predicted proteins currently available (for this stage/species) in public databases by approximately 141-fold [we obtained 8037 conceptually translated proteins, and there are currently 57 proteins in Genbank as of November 2011]. This quantity of contigs is similar to the numbers obtained with other nematodes like *Trichinella spiralis* [19], *N. americanus* and *Ancylostoma caninum* [51], *Haemonchus contortus* [52], *Dictyocaulus viviparous* [20], and *Teladorsagia circumcincta* [53]. The subset (55.91%) of *S. stercoralis* sequences with orthologues/homologues in public databases was slightly higher to that reported in similar transcriptomic studies of other animal-parasitic helminths such as *Necator americanus* [51,54]. It is noteworthy to mention that 44.09% of the putative proteins of *S. stercoralis* L3i transcriptome remain unannotated, warranting further genomic and functional characterization studies.

With the exception of three metabolic proteins (citrate synthase, arginine kinase and ATP:guanido phosphotransferase) all proteins identified in a previous proteomic study with *S. stercoralis* L3i [55] were included in the transcriptome described here. In addition, 41 antigenic proteins including SiR and tropomysin were present in the transcriptome (as searched in Table S1), confirming its value as a tool for searching targets for immunodiagnosis.

It is well characterized that upon infection, infective larvae (L3i) must penetrate skin as quickly as possible and then migrate within the host. In this context, proteases play an essential role. Among the proteins identified in our study, 60 different putative proteases were annotated in nine groups. These include nine metalloproteases and three aspartic proteases, some of them assumed to play a major role in skin penetration in *Strongyloides stercoralis* [56,57], and in other *Strongyloides* species like *S. venezuelensis* [58] or...
S. ratti [22, 23]. In S. venezuelensis, Yoshida et al. (2011) [24] have recently identified an astacin-like metalloproteinase as being specific of L3i in a transcriptomic study. Another abundant group was the cysteine proteases, including cathepsin B, legumain and calpain, proteins characterized as immunomodulators of host response and promising vaccine and drug targets [59–61]. Similar results have been reported for Ascaris suum, where 456 peptidases have been identified in its draft genome [62].

Kinases are also an important group of proteins considered to be good druggable targets from the medical and chemical viewpoints, since they play essential functions in the parasite, in mediating signal transduction [63–65]. In S. stercoralis L3i transcriptome analysis 73 putative kinases including 11 putative tyrosine kinases were identified (Table 3 and Table S1).

In S. stercoralis L3i, the most abundant transcripts encoding ES proteins include cytoskeletal proteins (i.e. myosin heavy chain, actin, tropomyosin, tubulin or paramyosin), metabolic enzymes (i.e. Trehalase, PEPCK, PGK, PGM, GAPDH, enolase), proteases, stress-response proteins, detoxifying enzymes, proteaseome components, most of them identified previously in S. stercoralis by proteomic studies [55]. These ES proteins play a major role in infection since they are present at the host-parasite interface and regulate host immune system [66]. ESPs also are among the target choice of new therapeutic solutions for helminth infections [67], as confirmed in the case of ivermectin (the currently the drug of choice for treating strongyloidosis) which has been shown to act reducing the secretion of ESPs from the ES apparatus in Brugia malayi microfilariae [68].

Recent studies using microarrays have identified highly expressed molecules in S. stercoralis L3i in comparison to L1 required by penetration and migration in the host. Supporting this notion, Soblik et al. (in press) have recently described the presence of 586 ES proteins in all the stages of S. ratti by proteomic analysis, 196 of which are also found in L3i [50]. When comparing larval S. ratti ES proteins with our predicted S. stercoralis L3i ES proteins, we find that 50 out of the 196 proteins identified from S. ratti were also detected in S. stercoralis L3i, supporting the value of the prediction.

Table 4. KEGG pathways of putative proteins inferred from the transcriptome of S. stercoralis L3i.

| Parent KEGG pathway                  | No of putative proteins (%) | Top KEGG pathway in the category                          |
|--------------------------------------|-----------------------------|-----------------------------------------------------------|
| Metabolism                           |                             |                                                           |
| Carbohydrate metabolism              | 201 (2.50)                  | Glycolysis/Gluconeogenesis                                 |
| Energy metabolism                    | 79 (0.98)                   | Oxidative phosphorylation                                 |
| Lipid metabolism                     | 104 (1.29)                  | Fatty acid metabolism                                      |
| Nucleotide metabolism                | 100 (1.24)                  | Purine metabolism                                          |
| Amino acid metabolism                | 174 (2.16)                  | Valine, leucine and isoleucine degradation                 |
| Metabolism of other amino acids      | 48 (0.60)                   | Glutathione metabolism                                    |
| Glycan biosynthesis and metabolism   | 30 (0.37)                   | N-Glycan biosynthesis                                      |
| Metabolism of Cofactors and Vitamins | 46 (0.57)                   | One carbon pool by folate                                 |
| Metabolism of Terpenoids and Polyketides | 5                       | Terpenoid backbone biosynthesis                            |
| Biosynthesis of other secondary metabolites | 1                       | Caffeine metabolism                                       |
| Xenobiotic biodegradation and metabolism | 23 (0.29)              | Drug metabolism – other enzymes                            |
| Genetic Information processing       |                             |                                                           |
| Transcription                        | 83 (1.03)                   | Spliceosome                                               |
| Translation                          | 195 (2.42)                  | Ribosome                                                  |
| Folding, sorting and degradation     | 178 (2.21)                  | Protein processing in endoplasmic reticulum               |
| Replication and repair               | 34 (0.42)                   | Nucleotide excision repair                                 |
| Environmental information processing |                             |                                                           |
| Membrane transport                   | 3                           | ABC transporters                                          |
| Signal transduction                  | 144 (1.79)                  | MAPK signaling pathway                                    |
| Signaling, molecules and interaction | 7                           | Neuroactive ligand-receptor interaction                   |
| Cellular processes                   |                             |                                                           |
| Transport and catabolism             | 132 (1.64)                  | Endocytosis                                               |
| Organism systems                     |                             |                                                           |
| Immune system                        | 7                           | Natural killer cell mediated cytotoxicity                 |
| Endocrine system                     | 11 (0.13)                   | Progesterone-mediated oocyte maturation                    |
| Development                          | 3                           | Dorsal-ventral axis formation                             |
| Environmental adaptation             | 6                           | Circadian rhythm – mammal                                 |

doi:10.1371/journal.pntd.0001513.t004
larvae, including cytochrome bcl1, Hsp-90 and FAR-1, which potentially constitute new targets for intervention [25], all of which were present in our transcriptome data, but did not appear as druggable targets following our pipeline, possibly as these are not present in DrugBank, where only lethal RNAi phenotype hits are included. Other important targets that interfered with, would still lead to expulsion of live worms form a host, like motility genes. In agreement with this, Garg and Ranganathan (2011) [22] have recently identified 19 contigs as putative drug targets in the S. stercoralis transcriptome in our study) did not appear as druggable molecules in our study, due to the presence of these proteins in host cells as well. In this context, efficient drugs as antihelmintics like benzimidazoles (they inhibit tubulin β resulting in impaired microtubule formation during cell division) have much more affinity for tubulin in helminth cells than the tubulin found in the cells of mammals [69]. We found 11 potential targets for treatment against L3i larvae. As already mentioned, these are the first evolutive phase of S. stercoralis in the host, and constitute a good target for treatment. From these target molecules, four, with no homologues in the host, suggesting parasite specificity, are: 2,3-bisphosphoglycerate independent phosphoglycerate mutase, glutamate synthase, isocitrate lyase and alcohol dehydrogenase I. Only the first one was predicted as present in ES. Further studies are required to confirm whether these molecules are good drug targets for strongyloidiasis. Next-generation sequencing technologies are improving genomic and transcriptomic studies, and complemented by proteomic investigations, should allow the characterization of differential gene expression and essential pathways in all the developmental stages of S. stercoralis. The transcriptomic dataset described here constitutes the basis for future investigations enlightening the search for control measures for one of the most neglected diseases.

Supporting Information

Figure S1  Bioinformatics workflow used for transcriptomic data analysis. Bioinformatics workflow comprising Phase I (pre-processing and assembly), II (Nucleotide level annotation), III (prediction of excretory/secretory (ES) proteins) and IV (Protein-level annotation). (DOC)

Table S1 Strongyloides stercoralis L3i gene ontology mapping using Interproscan. (XLS)

Table S2 Peptidases found in the S. stercoralis L3i transcriptome. (XLS)

Table S3 BLAST results of proteins found homologous to experimentally verified secretory proteins of parasitic helminths at 1e-15. (XLS)
Table S4  Top 50 abundant transcripts across Strongyloides stercoralis L3i excretory/secretory proteins using Blast2Go [44].

| Rank | Transcript | Fold Change |
|------|------------|-------------|
| 1    | L3i Transcript | 1000        |
| 2    | L3i Transcript | 950         |
| 3    | L3i Transcript | 900         |
| 4    | L3i Transcript | 850         |
| 5    | L3i Transcript | 800         |

Table S5 S. stercoralis L3i putative proteins found similar to known therapeutic targets in parasitic nematodes either by protein domains mapping or sequence similarity.

| Protein | Similarity | Target | References |
|---------|------------|--------|------------|
| L3i putative protein | Strongyloides stercoralis | 1 | 2 |
| L3i putative protein | Strongyloides stercoralis | 2 | 3 |
| L3i putative protein | Strongyloides stercoralis | 3 | 4 |
| L3i putative protein | Strongyloides stercoralis | 4 | 5 |
| L3i putative protein | Strongyloides stercoralis | 5 | 6 |

Acknowledgments

The Bioinformatics Core Service of the BMBCP (UPV-CSIC) is acknowledged for its help in bioinformatics analyses.

Author Contributions

Conceived and designed the experiments: AM DB JGE. Performed the experiments: JS CMA MT DB. Analyzed the data: GG JF JO AB SR. Contributed reagents/materials/analysis tools: MVD LP JMB JF JO GG SR. Wrote the paper: AM GG DB SR.

References

1. Siddiqi AA, Berk SL (2001) Diagnosis of Strongyloides stercoralis infection. Clin Infect Dis 33(7): 1040-1047.
2. Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, et al. (2006) Soil-transmitted helminth infections: Ascaris, trichuriasis, and hookworm. Lancet 367(9521): 1531-1532.
3. Olsen A, van Lieshout L, Marti H, Polderman T, Polman K, et al. (2009) Strongyloidiasis-the most neglected of the neglected tropical diseases? Trans R Soc Trop Med Hyg 103(10): 967-972.
4. Keiser PB, Nutman TB (2004) Strongyloides stercoralis in the immunocompromised population. Clin Microbiol Rev 17(1): 208-217.
5. Vadlamudi RS, Chi DS, Krishnasamy G (2006) Intestinal strongyloidiasis and hyperinfection syndrome. Clin Mol Allergy 4: 4.
6. Marcos LA, Terrabina A, Canales M, Gottuzzo E (2011) Update on strongyloidiasis in the immunocompromised host. Currr Infect Dis Rep 13(1): 35-46.
7. Oltra-Alcaraz C, Igual-Adell R, Sanchez P, Blasco M, Sanchez O, et al. (2004) Characteristics and geographical profile of strongyloidiasis in healthcare area 11 of the valencian community (spain). J Infect 49(2): 152-158.
8. Seyboth L, Christiansen D, Barnett E (2006) Diagnostic evaluation of newly arrived asymptomatic refugees with eosinophilia. Clin Infect Dis 42(3): 563-567.
9. Dreyer G, Fernandes-Silva E, Alves S, Rocha A, Albuquerque R, et al. (1996) Patterns of detection of Strongyloides stercoralis in stool specimens: Implications for diagnosis and clinical trials. J Clin Microbiol 34(4): 2569-2571.
10. Sato Y, Otsuru M, Takara M, Shiroma Y (1986) Intradermal reactions in strongyloidiasis. Int J Parasitol 16(1): 87-91.
11. Hirata T, Uchima N, Kishimoto K, Zaha O, Kinjo N, et al. (2006) Impairment of host immune response against Strongyloides stercoralis by human T cell lymphotropic virus type I infection. Am J Trop Med Hyg 74(2): 246-249.
12. Sykes AM, McCarthy JS (2011) A coproantigenic diagnostic test for Strongyloides infection. PLoS Negl Trop Dis 5(2): e555.
13. Igual-Depedro R, Oltra-Alcaraz C, Soler-Company E, Sanchez-Sanchez P, Matgo-Oyana J, et al. (2004) Efficacy and safety of ivermectin and thiabendazole in the treatment of strongyloidiasis. Expert Opin Pharmacother 5(12): 2615-2619.
14. Supattapongse Y, Premasartthan B, Bhumimunna K, Wayda D, Nilaganuwong S, et al. (2011) Efficacy and safety of single and double doses of ivermectin versus 7-day high dose albendazole for chronic strongyloidiasis. PLoS Negl Trop Dis 5(3): e1014.
15. Woltersholm A, Fairweather I, Richard R, von Samson-Himmelstjerna G, Santner N (2004) Drug resistance in veterinary helminths. Parasit Trends 20(10): 469-476.
16. Riddle HC, Blumenthal T, Meyer BJ, Price RS (ed) (1997) C. elegans II. Cold Spring Harbor Laboratory Press, New York, USA.
17. Sugimoto A (2004) High-throughput RNAs in Caenorhabditis elegans. Genome-wide screens and functional genomics. Differentiation 72(2-3): 81-91.
18. Blaxter ML, de Ley P, Carey JR, Lux LU, Schledermann P, et al. (1998) A molecular evolutionary framework for the phylum nematoda. Nature 392(6671): 71-75.
19. Cantacessi C, Mitreva M, Campbell BE, Hall RS, Young ND, et al. (2010) First transcriptomic analysis of the economically important parasitic nematode, Strongyloides stercoralis. Virol J 7: 71.
20. Beekmann P, Dumas S, Berlino IS, Grandjean P, Morin R, et al. (2007) Quantitative PCR analysis of the developmental stages of the human hookworm, Necator americanus. Int J Parasitol 37(13): 1575-1582.
21. Chen W, Zhou H, Hu Y, Feng Y, Zhang J, et al. (2010) Transcriptome analysis of the economically important parasitic nematode, Strongyloides stercoralis. PLoS Negl Trop Dis 4(6): e719.
22. Cantacessi C, Gasper RB, Ranganathan S (2009) An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen Fasciola hepatica. Proteomics: Associated with invasion and infection of the mammalian host. Mol Cell Proteomics 8(6): 1091-1097.
23. Chevreux B, Pisterer T, Drescher B, Driesel A, Muller WE, et al. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14(6): 1147-1159.
24. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9(2): 86-97.
25. Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol. pp 138-146.
26. Krogh A, Larsson B, von Heijne G, Brunak S (2001) Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. J Mol Biol 305(3): 567-580.
27. Benndorf JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides. SignalP 3.0. J Mol Biol 340(4): 783-795.
28. Benndorf JD, Jensen LJ, Blom N, von Heijne G, Brunak S (2004) Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel 17(4): 349-356.
29. Benhura S, Semanani R, Meng Z, Ribeiro JM, Venenuta TD, et al. (2009) Drug-induced excreted/secreted proteins at the host/parasite interface: Stage- and gender-specific proteomic profiling. PLoS Negl Trop Dis 3(4): e1016.
30. Emsenhaus O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300(4): 1005-1016.
31. Zdobnov EM, Apweiler R (2001) InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17(9): 847-848.
32. Emanuelsson O, Nielsen H, von Heijne G, Brunak S (2000) Improved prediction of subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300(4): 1005-1016.
33. Knox C, Law V, Jewison T, Liu P, Ly S, et al. (2011) DrugBank 3.0: A comprehensive resource for ‘omics’ research on drugs. Nucleic Acids Res 39(Database issue): D1035-41.
34. Bennet J, Carpena I, Espert AM, De la Rubia JE, Esteban JG, et al. (2006) Identification of proteins in excretory/secretory extracts of Echinostoma fetidum (trematoda) from chronic and acute infections. Proteomics 6(9): 2033-2043.
48. Maizels RM, Yazdanbakhsh M (2003) Immune regulation by helminth parasites: Cellular and molecular mechanisms. Nat Rev Immunol 3(9): 733–744.

49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3): 403–410.

50. Sobhik H, Younis AR, Mitreva M, Renard BY, Kirchner M, et al. (2011) Life cycle stage-resolved proteomic analysis of the excretome/secretome from Strongyloides ratti: Identification of stage-specific protease. Mol Cell Proteomics (in press).

51. Cantacessi C, Mitreva M, Jex AR, Young ND, Campbell BE, et al. (2010) Massively parallel sequencing and analysis of the Necator americanus transcriptome. PLoS Negl Trop Dis 4(5): e684.

52. Cantacessi C, Campbell BE, Young ND, Jex AR, Hall RS, et al. (2010) Differences in transcription between free-living and CO2-activated third-stage larvae of Haemonchus contortus. BMC Genomics 11: 266.

53. Dicker AJ, Nath M, Yaga R, Nisbet AJ, Lainson FA, et al. (2011) Teladorsagia circumcincta: The transcriptomic response of a multi-drug-resistant isolate to ivermectin exposure in vitro. Exp Parasitol 127(2): 351–356.

54. Rabelo EM, Hall RS, Loukas A, Cooper L, Hu M, et al. (2009) Improved insights into the transcriptomes of the human hookworm Necator americanus–fundamental and biotechnological implications. Biotechnol Adv 27(2): 122–132.

55. Marcilla A, Sotillo J, Perez-Garcia A, Igual-Adell R, Valero ML, et al. (2010) Proteomic analysis of Strongyloides stercoralis L3 larvae. Parasitology 137(10): 1577–1583.

56. Brindley PJ, Gam AA, McKerrow JH, Neva FA (1995) Ss40: The zinc endopeptidase secreted by infective larvae of Strongyloides stercoralis. Exp Parasitol 80(1): 1–7.

57. Gomez Gallego S, Loukas A, Slade RW, Neva FA, Varaharajalu R, et al. (2005) Identification of an astacin-like metallo-proteinase transcript from the infective larvae of Strongyloides stercoralis. Parasitol Int 54(2): 123–133.

58. Maruyama H, Nishimaki A, Takuma Y, Kurimoto M, Suzuki T, et al. (2006) Successive changes in tissue migration capacity of developing larvae of an intestinal nematode, Strongyloides venezuelensis. Parasitology 132(8): 411–418.

59. Smooker PM, Jayaraj R, Pike RN, Spithill TW (2010) Cathepsin B proteases of flukes: The key to facilitating parasite control? Trends Parasitol 26(10): 506–514.

60. Dalton JP, Brindley PJ, Donnelly S, Robinson MW (2009) The enigmatic asparaginyl endopeptidase of helminth parasites. Trends Parasitol 25(2): 59–61.

61. Ohia N, Kumagai T, Maruyama H, Yoshida A, He Y, et al. (2008) Research on calpain of Schistosoma japonicum as a vaccine candidate. Parasitol Int 57(2): 175–181.

62. Jex AR, Liu S, Li B, Young ND, Hall RS, et al. (2011) Ascaris suum draft genome. Nature (in press). doi: 10.1038/nature10533.

63. Campbell BE, Boag PR, Hofmann A, Cantacessi C, Wang CK, et al. (2011) Atypical (RIO) protein kinases from Haemonchus contortus–promise as new targets for nematocidal drugs. Biotechnol Adv 29(3): 338–350.

64. Campbell BE, Boag PR, Hofmann A, Cantacessi C, Wang CK, et al. (2011) Atypical (RIO) protein kinases from Haemonchus contortus–promise as new targets for nematocidal drugs. Biotechnol Adv 29(3): 338–350.

65. Campbell BE, Hofmann A, McCluskey A, Gasser RB (2011) Serine/threonine phosphatases in socioeconomically important parasitic nematodes–prospects as novel drug targets? Biotechnol Adv 29(1): 28–39.

66. Hewitson JP, Grainger JR, Maizels RM (2009) Helminth immunoregulation: The role of parasite secreted proteins in modulating host immunity. Mol Biochem Parasitol 167(1): 1–11.

67. Bungiro R, Cappello M (2011) Twenty-first century progress toward the global control of human hookworm infection. Curr Infect Dis Rep 13(3): 210–217.

68. Moreno Y, Nabhan JF, Solomon J, Mackenzie CD, Geary TG (2010) Ivermectin disrupts the function of the excretory-secretory apparatus in microfilariae of Brugia malayi. Proc Natl Acad Sci U S A 107(46): 20120–20125.

69. Lacey E (1990) Mode of action of benzimidazoles. Parasitol Today 6(4): 112–115.