

RESEARCH ARTICLE

A comparative 'omics' approach for prediction of candidate *Strongyloides stercoralis* diagnostic coproantigens

Tegwen Marlais^{1*}, Jack Bickford-Smith¹, Carlos Talavera-López², Hai Le¹, Fatima Chowdhury¹, Michael A. Miles¹

1 Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Institute of Computational Biology, Computational Health Centre, Helmholtz Munich, Neuherberg, Germany

* Tegwen.marlais@lshtm.ac.uk



OPEN ACCESS

Citation: Marlais T, Bickford-Smith J, Talavera-López C, Le H, Chowdhury F, Miles MA (2023) A comparative 'omics' approach for prediction of candidate *Strongyloides stercoralis* diagnostic coproantigens. PLoS Negl Trop Dis 17(4): e0010777. <https://doi.org/10.1371/journal.pntd.0010777>

Editor: Krystyna Cwiklinski, University of Liverpool, UNITED KINGDOM

Received: August 31, 2022

Accepted: April 3, 2023

Published: April 17, 2023

Copyright: © 2023 Marlais et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files. Data sourced from existing, open-access, databases are fully referenced and accession numbers are provided in the manuscript and supporting files. Accession numbers for the datasets used are as follows: RNA (transcriptomic data), accessed via NCBI: PRJEB3116, ERR146941-ERR146961. Genomic data, accessed via WormBase ParaSite v8: PRJEB528, PRJEB125, PRJEB530, PRJEB525. Protein coding sequences

Abstract

Human infection with the intestinal nematode *Strongyloides stercoralis* is persistent unless effectively treated, and potentially fatal in immunosuppressed individuals. Epidemiological data are lacking, partially due to inadequate diagnosis. A rapid antigen detection test is a priority for population surveillance, validating cure after treatment, and for screening prior to immunosuppression. We used a targeted analysis of open access 'omics' data sets and used online predictors to identify *S. stercoralis* proteins that are predicted to be present in infected stool, *Strongyloides*-specific, and antigenic. Transcriptomic data from gut and non-gut dwelling life cycle stages of *S. stercoralis* revealed 328 proteins that are differentially expressed. *Strongyloides ratti* proteomic data for excreted and secreted (E/S) proteins were matched to *S. stercoralis*, giving 1,057 orthologues. Five parasitism-associated protein families (SCP/TAPS, prolyl oligopeptidase, transthyretin-like, aspartic peptidase, acetylcholinesterase) were compared phylogenetically between *S. stercoralis* and outgroups, and proteins with least homology to the outgroups were selected. Proteins that overlapped between the transcriptomic and proteomic datasets were analysed by multiple sequence alignment, epitope prediction and 3D structure modelling to reveal *S. stercoralis* candidate peptide/protein coproantigens. We describe 22 candidates from seven genes, across all five protein families for further investigation as potential *S. stercoralis* diagnostic coproantigens, identified using open access data and freely-available protein analysis tools. This powerful approach can be applied to many parasitic infections with 'omic' data to accelerate development of specific diagnostic assays for laboratory or point-of-care field application.

Author summary

The worm *Strongyloides stercoralis* causes infectious disease in people throughout tropical and sub-tropical regions, leading to an extensive reduction in quality of life and even death. Millions of people are at risk of infection with this parasite and improved diagnostic and control methods and technologies are urgently required. Currently, most diagnosis

(genome-derived proteomes), accessed via WormBase ParaSite v8: PRJNA72581, PRJEB4950, PRJNA13758, PRJDA72781, PRJEB503, PRJNA72135, PRJEB513, PRJEB515, PRJEB525, PRJEB530, PRJEB125, PRJEB528, PRJEB524, PRJNA170813, PRJEB535. Human genome pep file, via Ensembl: GRCh38. Proteomic data, accessed from Supplementary Table 19 in Hunt et al., 2016, doi: [10.1038/ng.3495](https://doi.org/10.1038/ng.3495).

Funding: TM was funded by the Sir Halley Stewart Trust (<http://www.sirhalleystewart.org.uk/>). The views expressed within this report are those of the authors and not necessarily those of the Trust. TM was additionally supported by the John Henry Memorial Fund (UK Charity number: 1118007). 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.

is carried out through methods involving visual inspection of patient's faeces, which has a number of drawbacks, particularly its poor sensitivity. This paper presents a new method to develop improved diagnostic tests for *S. stercoralis*, by computational analysis of publicly available gene and protein sequences to predict proteins that may be detectable in faeces. This would enable the development of rapid diagnostic tests in the form of lateral flows or dipsticks, with better predictive ability and fewer drawbacks than current diagnostic methods. A number of potential proteins, predicted to have all the desired characteristics for use in such tests were found through the new method and have been presented in this paper. With validation, new diagnostic tests for *S. stercoralis* could be developed from these results and the computational approach could be used to target other parasitic diseases.

Introduction

The intestinal nematode *Strongyloides stercoralis* is a soil transmitted helminth (STH) prevalent in faecally-contaminated, humid soils in tropical and sub-tropical regions. Strongyloidiasis is estimated to affect up to 40% of people in many endemic regions [1,2]. Infection occurs when infective third stage (L3) larvae penetrate the skin. The parasitic adult female resides in the epithelium of the duodenum where it feeds on host tissue. Although clinical signs may be mild or non-specific, long term infection by *Strongyloides* can have significant impact on quality of life and child development and progress to severe and fatal disease [3–5].

Strongyloides stercoralis is unusual among human parasitic nematodes in that it can complete its life cycle within the host, thus sustaining infection for decades if untreated [6–8]. During reduced immune competence due to immunosuppression, such as corticosteroid treatment of co-morbidities, or HTLV-1 co-infection, very large numbers of larvae may undergo this autoinfective cycle, causing hyperinfection or disseminated strongyloidiasis, both with a high fatality rate [9–11]. This autoinfective lifecycle means that infections will fully re-establish if worms are not completely cleared from the host during treatment. Diagnosis of strongyloidiasis and validation of cure after treatment are therefore imperative.

Treatment with the first-line drug ivermectin has a reported efficacy of between 57% and 100%. However, accurate determination of cure depends on follow-up and the diagnostic method used [8,10,12]. Albendazole and mebendazole, used to treat infection with other STH are less effective or ineffective against *Strongyloides* [13–15]. Moxidectin has shown effectiveness against *S. stercoralis* that is equivalent to ivermectin, and trials continue to evaluate it further [16,17].

There is no single gold-standard diagnostic method for strongyloidiasis. Options include microscopy of cultured or incubated fresh stool samples, such as Koga agar plate culture [18] or Baermann funnel. qPCR on extracted stool DNA is used in research and highly resourced laboratories, and provides a reliable confirmation test but may not improve sensitivity over larval detection methods if used for screening [19]. Sensitivity of PCR may be affected by DNA extraction method and choice of PCR reagents [20], as well as irregular larval excretion [12]. Serology, detecting antibodies to either whole worm or recombinant antigens NIE and SsIR, has the highest sensitivity of the diagnostic options; ranging from 72 to 98%, and 90 to 99% specificity, when compared with microscopic detection as gold standard. Specific antigen assays achieve accuracy at the upper end of the ranges and continue to be improved upon [21–23]. While serology can indicate cure, this requires re-sampling several months after effective treatment [24–26]. Therefore, there is a need for a rapid diagnostic test (RDT) that can be used

for screening as well as for timely confirmation of cure. A coproantigen-based assay would fulfil this need, because it may achieve high sensitivity and specificity, and test for active infection. Pilots of such assays have shown proof of principle under research conditions using antibodies against *Strongyloides ratti* [27,28], *Strongyloides venezuelensis* [29,30] and *S. stercoralis* somatic, excretory/secretory (E/S) or faeces-derived antigens [31], as reviewed by Blachandra *et al.* (2021) [32]. However, identification of specific *Strongyloides* protein antigens would enable production of standardised diagnostic tests on a large scale.

The wealth of ‘omic’ data now available in the public domain, coupled with online protein analysis tools, enables a computational approach to antigen discovery. This concept was termed reverse vaccinology when used in vaccine candidate discovery [33]. The approach begins with genomic analysis, as opposed to biochemical or serological methods and it also has significant potential in diagnostic antigen discovery. It has the advantages of not requiring culture of the organism and of revealing therapeutically or diagnostically relevant antigens that may be less abundant or difficult to purify *in vitro*, such as those from host-dwelling life stages of pathogens. Incorporation of transcriptomic data can inform candidate gene/protein selection in parasites with multiple life stages, such as *S. stercoralis* [33].

Our approach was facilitated by the publication of the genomes of *S. stercoralis* and three related *Strongyloides* species [34]. Here, we have applied a series of computational analyses to open access transcriptomic, genomic and proteomic data from *Strongyloides* species and other helminths. We have used common bioinformatic tools, to identify *Strongyloides* protein antigens that may be diagnostic targets detectable in human stool, using a coproantigen capture RDT.

Methods

Data sources

This study used data obtained from public databases, including Wormbase ParaSite, a resource for parasitic worm genomics curated by the Sanger Institute and the EMBL European Bioinformatics Institute [35,36]. Full details of data sources are listed in Table 1. *Strongyloides stercoralis* transcript sequences identified by the prefix ‘SSTP’ can be obtained via UniProtKB (www.uniprot.org) or WormBase ParaSite (WBPS: www.parasite.wormbase.org).

Overview of method

Three criteria were applied for candidate antigen selection: predicted presence in infected stool; specificity to *Strongyloides* and/or *S. stercoralis*; predicted antigenicity, to facilitate raising sensitive antibodies. Datasets and computational analyses, all open-access, used to make this selection are detailed in Fig 1.

Evidence for the presence of candidate coproantigens in stool

Evidence for the presence of candidate coproantigens in stool was considered to be from proteins that are associated with parasitism and gut-dwelling life stages of *Strongyloides* and therefore with maximal chances of being released into the host gut lumen. For this we used open-access transcriptomic and proteomic data (Fig 1, boxes A and B).

Analysis of the *Strongyloides* genomes by Hunt *et al.* (2016) [38] was used to identify protein families associated with parasitism. Five parasitism-associated protein families (‘priority protein families’), of suitable size to be put through our pipeline, were our focus for identifying coproantigens: sperm-coating-proteins/Tpx-1/Ag5/PR-1/Sc7 (SCP/TAPS), transthyretin-like (TTL), acetylcholinesterase (AChE), prolyl oligopeptidase (POP) and aspartic peptidases. The large astacin, and small proteinase inhibitor, families were excluded from our analysis in this

Table 1. Data used in this study. Accession numbers, source databases, and original references where used. CDS, coding sequence; NCBI SRA, National Centre for Biotechnology Information Sequence Read Archive.

| Data type | Species | Accession numbers | Database/Source | Reference |
|---|---|------------------------|-----------------------------|--|
| RNA-seq | <i>Strongyloides stercoralis</i> | PRJEB3116 | NCBI BioProject | Stoltzfus <i>et al.</i> , 2012 [37] |
| | <i>Strongyloides stercoralis</i> (gut-dwelling life stages: Parasitic female, Post-parasitic L1, Post-parasitic L3) | ERR146953 to ERR146961 | NCBI SRA | |
| | <i>Strongyloides stercoralis</i> (non-gut-dwelling life stages: Free-living female, Post-free-living L1, infectious L3, tissue-migrating L3 larvae) | ERR146941 to ERR146952 | NCBI SRA | |
| Genome | <i>Strongyloides stercoralis</i> | PRJEB528 | WormBase ParaSite (WBPS) v8 | Hunt <i>et al.</i> , 2016 [34] |
| | <i>Strongyloides ratti</i> | PRJEB125 | | |
| | <i>Strongyloides venezuelensis</i> | PRJEB530 | | |
| | <i>Strongyloides papillosus</i> | PRJEB525 | | |
| Proteins (CDS files for helminths, pep file for <i>Homo sapiens</i>) | <i>Ancylostoma duodenale</i> | PRJNA72581 | WBPS v8 | Hunt <i>et al.</i> , 2016 [34]–(Strongyloides and Parastrongyloides species) |
| | <i>Ascaris lumbricoides</i> | PRJEB4950 | | |
| | <i>Caenorhabditis elegans</i> | PRJNA13758 | | |
| | <i>Clonorchis sinensis</i> | PRJDA72781 | | |
| | <i>Enterobius vermicularis</i> | PRJEB503 | | |
| | <i>Necator americanus</i> | PRJNA72135 | | |
| | <i>Onchocerca volvulus</i> | PRJEB513 | | |
| | <i>Parastrongyloides trichosuri</i> | PRJEB515 | | |
| | <i>Strongyloides papillosus</i> | PRJEB525 | | |
| | <i>Strongyloides venezuelensis</i> | PRJEB530 | | |
| | <i>Strongyloides ratti</i> | PRJEB125 | | |
| | <i>Strongyloides stercoralis</i> | PRJEB528 | | |
| | <i>Syphacia muris</i> | PRJEB524 | | |
| | <i>Taenia solium</i> | PRJNA170813 | | |
| | <i>Trichuris trichiura</i> | PRJEB535 | | |
| <i>Homo sapiens</i> | GRCh38 | Ensembl | | |
| E/S proteome | <i>Strongyloides ratti</i> | n/a | Supplementary Data Table 19 | Hunt <i>et al.</i> , 2016 [34] |

<https://doi.org/10.1371/journal.pntd.0010777.t001>

study, because they were not amenable to the manual sections of the pipeline, either by being too many or too few sequences for comparison.

Transcriptomics. To reveal candidate coproantigens, we used transcriptomic data from Stoltzfus *et al.* (2012) [37], which analysed transcriptomes of 7 life stages of *S. stercoralis* (Figs 2 and 1, box A). RNA data were downloaded from the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA), with triplicate reads from each stage. We grouped these data by life stage and by presence or absence within the host gut, thus representing stages excreting or secreting antigens into human stool (Fig 2).

We calculated relative abundance of transcripts using RSEM [39] and bowtie2 [40] and subsequently separated RNA data from the 4 non-gut-dwelling stages and 3 gut-dwelling stages into two groups. Differential gene expression between the two groups was analysed using edgeR in RSEM (Fig 1, box C). We selected genes that were differentially expressed with a 100% credibility interval (CI), in either direction, between gut-dwelling and non-gut-dwelling life stages.

Differentially expressed protein family identification. ClustalW [41] was used to perform multiple alignment of the differentially expressed (DE) genes and to produce a phylogenetic tree, which was annotated with iTOL [42]. The tree, labelled only with *S. stercoralis* gene accession numbers, facilitated grouping of the DE genes into protein/gene families but protein identities remained unknown at this stage (Fig 1, box D).

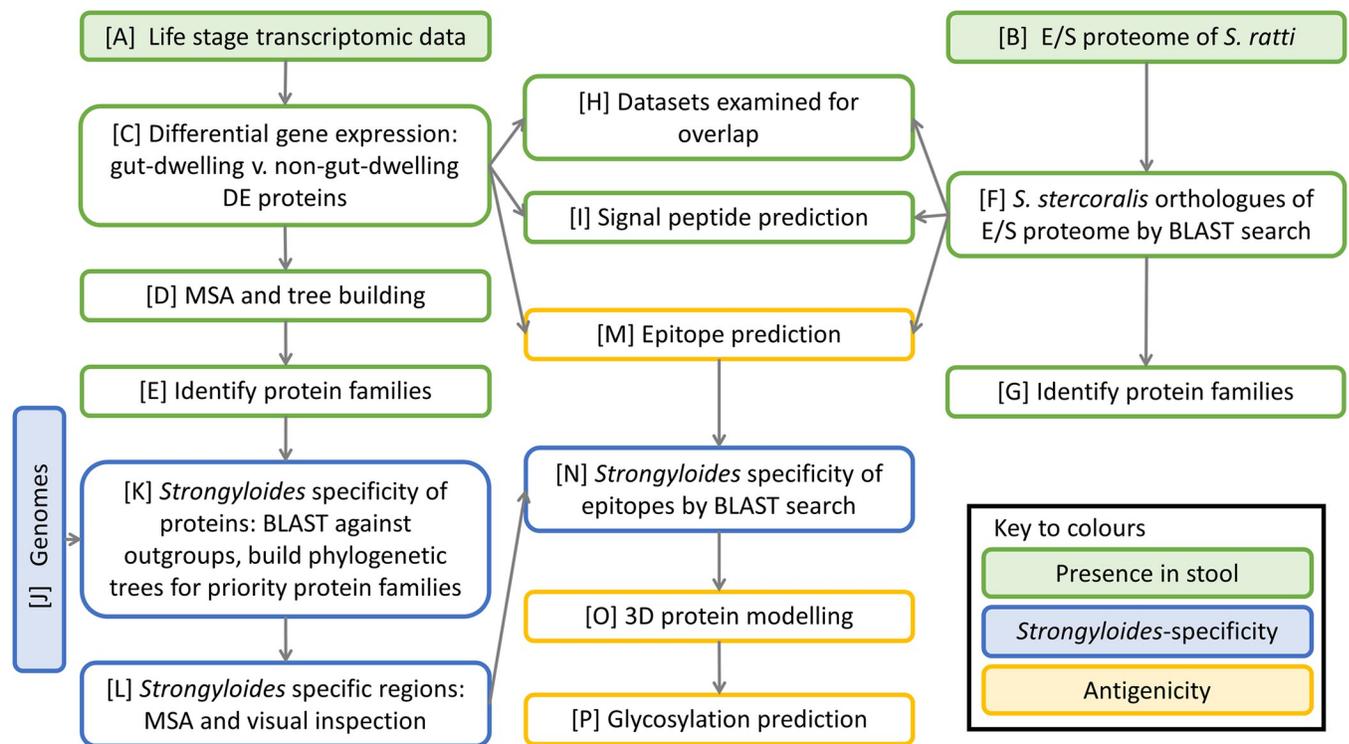


Fig 1. Schematic of the method for identifying coproantigens of *Strongyloides stercoralis* using 'omic' data and computational analyses. Filled shapes represent datasets, colours represent analyses against three characteristics of a candidate coproantigen. E/S, excreted and secreted; MSA, multiple sequence alignment; Ss, *Strongyloides stercoralis*; BLAST, basic local alignment search tool; DE, differentially expressed.

<https://doi.org/10.1371/journal.pntd.0010777.g001>

DE proteins were firstly identified by submitting amino acid sequences of selected clusters on the tree to three domain-finding tools: Delta BLAST [43], InterPro [44] and ExPASy Prosite [45,46] and a consensus of all three used to obtain probable protein identity. Secondly, DE protein sequences were submitted to BlastKOALA [47]. BlastKOALA protein identities were considered alongside previous consensus or used alone if there was no identity from three domain-finding tools (Fig 1, box E).

Excretory/Secretory proteomics

Source of proteomic data. E/S data was used in conjunction with gene expression data, to build evidence for the candidate coproantigens (Fig 1, box B). Soblik *et al.* (2011) [48] submitted excretory/secretory (E/S) material of *S. ratti* parasitic females to mass spectrometry and identified the constituent proteins. In their presentation of the *S. ratti* genome, Hunt *et al.* (2016) [34] re-analysed the spectral data and obtained protein identities from corresponding genomic data of *S. ratti*. We acquired the list of parasitic female E/S proteins, with *S. ratti* genome accession numbers and protein identities, from Supplementary Table 19 of Hunt *et al.* (2016) [34] and subsequently obtained the corresponding amino acid sequences from the *S. ratti* protein file (WBPS v8) using samtools [49].

S. stercoralis orthologues to the *S. ratti* excretory/secretory proteome

At the time of this study, there were no E/S proteomic data available for *S. stercoralis*. Therefore, we obtained *S. stercoralis* orthologues of the *S. ratti* E/S proteins by searching the *S. ratti* E/S proteins against a custom blast+ database consisting of the *S. stercoralis* protein file (WBPS v8), using blastp with word size 2 and e-value -50 (Fig 1, box F). *S. stercoralis* hits, in the form of accession

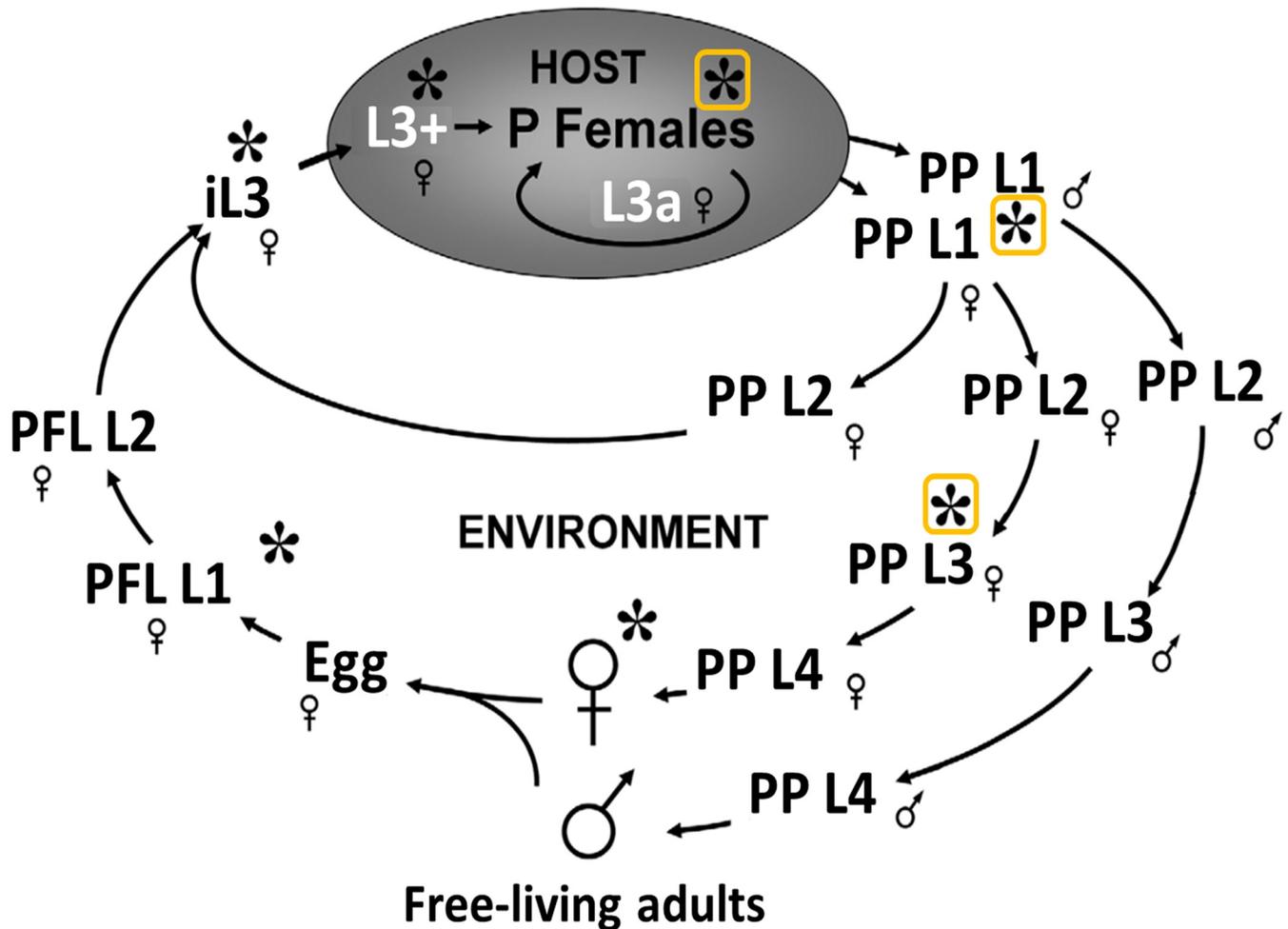


Fig 2. *Strongyloides stercoralis* life stages. Asterisks indicate life stages for which transcriptomic data were obtained by Stoltzfus *et al.* (2012). We grouped these transcriptomic data by presence in the host gut (boxed asterisks), and outside the host gut. P Females, parasitic females; PP, post parasitic; FL, Free-living; PFL, post free-living; L1, stage 1 larva; L3, stage 3 larva; iL3, infectious third stage larva; L3+, tissue-migrating larva; L3a, autoinfective L3. Figure modified from Stoltzfus *et al.* (2012) [37] under a CC BY license. Accession numbers are given for NCBI SRA for the triplicate reads in Table 1.

<https://doi.org/10.1371/journal.pntd.0010777.g002>

numbers, were extracted from the resulting table and duplicated hits removed. Corresponding *S. stercoralis* amino acid sequences were extracted from the *S. stercoralis* protein file using samtools. VENNY 2.1 [50] was used to reveal the *S. stercoralis* accession numbers that occurred in both the DE proteins and the E/S orthologues. All the E/S orthologues were submitted to BlastKOALA as before, to obtain protein family identities, as well as matching them with the protein identities reported for the original *S. ratti* E/S proteins, by Hunt *et al.* (2016) [34] (Fig 1, box G). Separately, differential gene expression data from analysis of the Stoltzfus *et al.* (2012) [37] dataset were extracted for the E/S orthologues that occurred in both datasets (Fig 1, box H).

Signal peptide prediction for evidence that a protein is secreted was performed on the *S. stercoralis* DE proteins and E/S orthologues using SignalP 4.1 [51] (Fig 1, box I).

Specificity of candidate coproantigens to *Strongyloides*

We used phylogenetic comparison to indicate *S. stercoralis* proteins with least homology to those of other relevant species, followed by multiple sequence alignment to identify exact regions of specificity (Fig 1, boxes J, K, L).

Table 2. Outgroup species used for phylogenetic comparison to *S. stercoralis*.

| Species | WormBase ParaSite accession number | Rationale for selection |
|---------------------------------------|------------------------------------|---|
| 1 <i>Ancylostoma duodenale</i> | PRJNA72581 | Hookworm, prevalent nematode of humans |
| 2 <i>Ascaris lumbricoides</i> | PRJEB4950 | Prevalent nematode of humans |
| 3 <i>Caenorhabditis elegans</i> | PRJNA13758 | Free-living nematode and model organism |
| 4 <i>Clonorchis sinensis</i> | PRJDA72781 | Trematode of humans |
| 5 <i>Enterobius vermicularis</i> | PRJEB503 | Prevalent nematode of humans |
| 6 <i>Homo sapiens</i> | GRCh38 (Ensembl) | Human, present in all samples |
| 7 <i>Necator americanus</i> | PRJNA72135 | Hookworm, prevalent nematode of humans |
| 8 <i>Onchocerca volvulus</i> | PRJEB513 | Non-gut nematode of humans |
| 9 <i>Parastrongyloides trichosuri</i> | PRJEB515 | Relative of <i>S. stercoralis</i> , facultative parasite of opossums |
| 10 <i>Strongyloides ratti</i> | PRJEB125 | Close relative of <i>S. stercoralis</i> , serological cross-reactivity with <i>S. stercoralis</i> [52] |
| 11 <i>Strongyloides venezuelensis</i> | PRJEB530 | Close relative of <i>S. stercoralis</i> |
| 12 <i>Strongyloides papillosus</i> | PRJEB525 | Close relative of <i>S. stercoralis</i> |
| 13 <i>Syphacia muris</i> | PRJEB524 | Nematode of mice and rats, serological cross-reactivity with <i>Strongyloides</i> in experimental animal infection [53] |
| 14 <i>Taenia solium</i> | PRJNA170813 | Cestode of humans |
| 15 <i>Trichuris trichiura</i> | PRJEB535 | Prevalent nematode of humans |

<https://doi.org/10.1371/journal.pntd.0010777.t002>

A custom blast+ database was created from the genome-derived proteomes of selected outgroup species (Table 2 and Fig 1, box J). The outgroups were selected to represent parasitic nematodes, including those that commonly co-infect with *S. stercoralis*, non-parasitic nematodes, as well as trematodes, cestodes and human, for a broad analysis of homology. Human protein/coding sequence (CDS) data were provided by the Human Genome Project at the Wellcome Trust Sanger Institute and obtained from Ensembl.

Specificity of differentially expressed proteins to *S. stercoralis*. Specificity of DE proteins to *Strongyloides* was assessed by their clustering in a phylogenetic tree after alignment with orthologues from the outgroups (Fig 1, box L). The *S. stercoralis* DE proteins that resulted from the transcriptomic comparison were searched as separate protein families, against the custom outgroup database with blast+ criteria: word size 2 and e-value of -5 or -10, as appropriate, to obtain about 100 to 1,000 hits (S1 File). In cases where there were very few DE proteins in a particular family, the DE protein(s) were also searched against a custom database consisting of only *S. stercoralis* genome-derived proteins (S1 File). This was intended to increase the number of *S. stercoralis* proteins to enable species-specific clusters to be revealed on phylogenetic trees.

Protein hits from each outgroup, and the additional *S. stercoralis* hits where appropriate, were aligned with their respective original blast queries by multiple sequence alignment (MSA) using ClustalW, and phylogenetic trees were constructed. Trees were annotated using iTOL [42] to show proteins from each outgroup species in a different colour. *S. stercoralis* proteins which formed a distinct cluster, or clusters, on each phylogenetic tree were viewed in MSA along with the most and least similar proteins from each of the outgroups on that tree. These ClustalW alignments were analysed by eye for *Strongyloides* and *S. stercoralis*-specific regions which were then submitted to BLASTP search against the NCBI non-redundant (nr), and Nematoda (taxid: 6231) databases, as relevant, to validate their specificity (Fig 1, box N).

Antigenic potential of candidate coproantigens

Epitope prediction. BepiPred 1.0 [54] and bcepred [55] were used to predict epitopes within the DE proteins and E/S proteome orthologues (Fig 1, box M). A BepiPred threshold of 1.3 (range -4 to 4) was selected for maximum specificity of 96%, with corresponding 13% sensitivity, of predicted epitopes in order to minimise the chance of false positive predictions. Minimum length was 9 amino acids with no maximum. In the differentially expressed proteins, longer sequences with an overall very high epitope score were allowed to contain small regions scoring below 1.3.

Bcepred criteria were based on the reported highest accuracy of 58.7% which was achieved using a threshold of 2.38 for the average score of four amino acid properties: hydrophobicity, flexibility, polarity and exposed surface. In addition to BepiPred 1.0 and Bcepred, we also used BepiPred version 2.0 [56] for certain candidate antigens that had already been identified by the pipeline. This version became available only after the majority of the analysis and offered improved prediction of conformational epitopes. BepiPred 2.0 was used with the same epitope length criteria and an epitope score threshold of 0.55 (range 0 to 1) which provided specificity of 81.7% and sensitivity of 29.2% on epitope predictions.

Outputs from the two prediction tools were compared, initially for proteins present in both outputs. The predicted epitope regions of these proteins were then examined for sequence overlap. Prior to selection as candidate antigens, predicted epitopes were assessed for their specificity to *Strongyloides* (Fig 1, box N). Sequences were searched using BLASTP against the NCBI nr database. The “expect threshold” in BLASTP was increased if no results were obtained with default parameters. BLASTP output was examined by eye for the sequence identity and biological relevance, i.e. likelihood of presence in a human stool sample.

3D modelling. Selected proteins of interest containing predicted epitopes, *Strongyloides*-specific regions, and in a priority protein family, were submitted to Phyre2 [57] for 3D structure modelling against known crystal structures, using the intensive mode (Fig 1, box O). UCSF Chimera [58] was used to visualise and annotate 3D models to highlight specific sequences of interest on the model.

Glycosylation prediction. N-linked glycosylation was predicted with NetNGlyc [59] to account for the potential of a glycan to obscure protein antigen regions, or conversely to contribute to antigenicity (Fig 1, box P). The prediction tool identified asparagine (N) residues with a high probability of being glycosylated via their amide nitrogen. Prediction was based on the motif N-X-S/T, where X is any residue except proline (P), and along with the presence of a signal peptide or trans-membrane domain on that protein, this indicates that potential glycosylation sites are likely to be glycosylated. Intracellular, intramembrane regions, or signal peptides of a protein are unlikely to be glycosylated. If present, a glycosylation site close to a candidate antigen region on the 3D protein could indicate that the protein is less likely to be accessible to antibodies in a capture assay and therefore a lower priority candidate, pending *in vitro* screening.

Results

Priority protein families

Hunt *et al.* (2016) [38] identified seven protein families associated with *Strongyloides* parasitism. For our coproantigen search, we focused on 5 of these that were amenable size to be put through our pipeline, namely: sperm coating protein/Tpx-1/Ag5/PR-1/Sc7 (SCP/TAPS), transthyretin-like (TTL), acetylcholinesterase (AChE), prolyl oligopeptidase (POP) and aspartic peptidases.

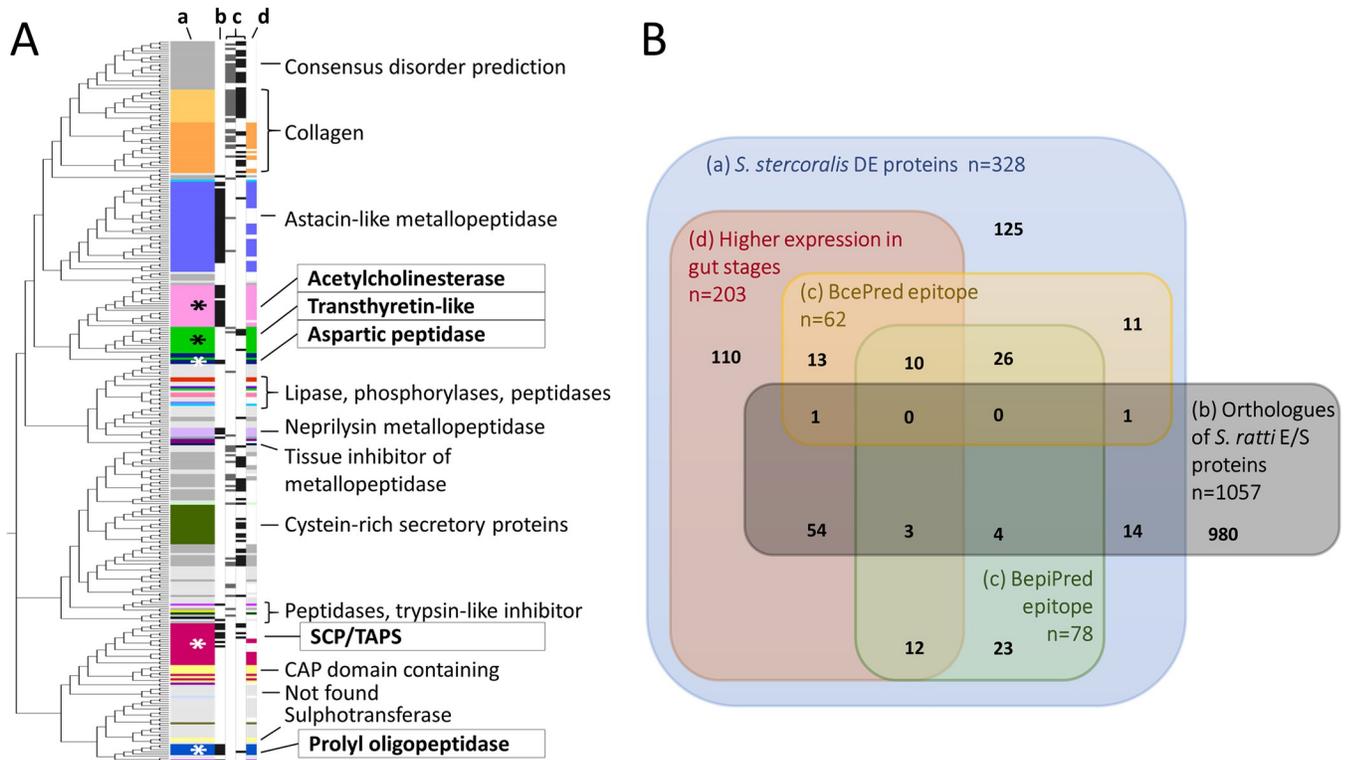


Fig 3. *Strongyloides stercoralis* proteins differentially expressed (DE) between gut-dwelling and non-gut-dwelling life stages and their overlap with other analyses. Panel A; DE proteins grouped by sequence similarity with protein families identified. Colours refer to features or families; the 5 priority protein families analysed in depth (SCP/TAPS; TTL; AChE; aspartic peptidases, and POP) are labelled with an asterisk and boxed text; pale grey indicates proteins not identified by any method, dark grey indicates disordered proteins. Panel B, overlap, in numbers of proteins, between the DE proteins and other analyses. In both panels, labels indicate: a, the 328 *S. stercoralis* DE proteins; b, proteins orthologous to *S. ratti* E/S proteins; c, proteins containing epitopes predicted by BcePred (left in Panel A) and BepiPred (right in Panel A); d, the 203 proteins that were DE in gut-dwelling life stages. Gene accession numbers and all protein family names for the DE proteins can be found in S2 and S4 Files for E/S orthologues. These results originate from methods in Fig 1, boxes C-H and M).

<https://doi.org/10.1371/journal.pntd.0010777.g003>

Differential gene expression in gut life stages

Of a total of 13,098 *S. stercoralis* genes identified in RNA-seq data by Stoltzfus et al. (2012) [37] we found 328 which were differentially expressed with a 100% CI between gut-dwelling and non-gut-dwelling life stages according to our groupings (Figs 3 and 1, box C). Of these, 198 (60.4%) contained a signal peptide (S2 File and Fig 1, box I). Of the 328 DE genes, 203 were more highly expressed in gut-dwelling life stages than non-gut and included proteins representing each of the 5 priority protein families analysed here (S2 File, columns B and C). These were therefore the focus of our coproantigen search (Fig 3 and S2 File). Twenty-eight protein families were identified among the differentially expressed (DE) proteins, accounting for 193 (58.8%) of the proteins, with the remainder either not identified (22%) or given a disorder prediction (19.2%), indicating that they do not have a fixed conformation and are difficult to assign to a particular function or family (Figs 3 and 1, boxes D,E).

Excretory/Secretory proteome

In the absence of a *S. stercoralis* excretory/secretory (E/S) proteome, we identified 1,057 *S. stercoralis* proteins that had high homology to the 584 proteins in the published E/S proteome of *S. ratti* [34], of which 325 (30.7%) contained signal peptides (Fig 1, boxes F and I). Original *S. ratti* E/S proteins were given as 582 accession numbers, however two were found to have

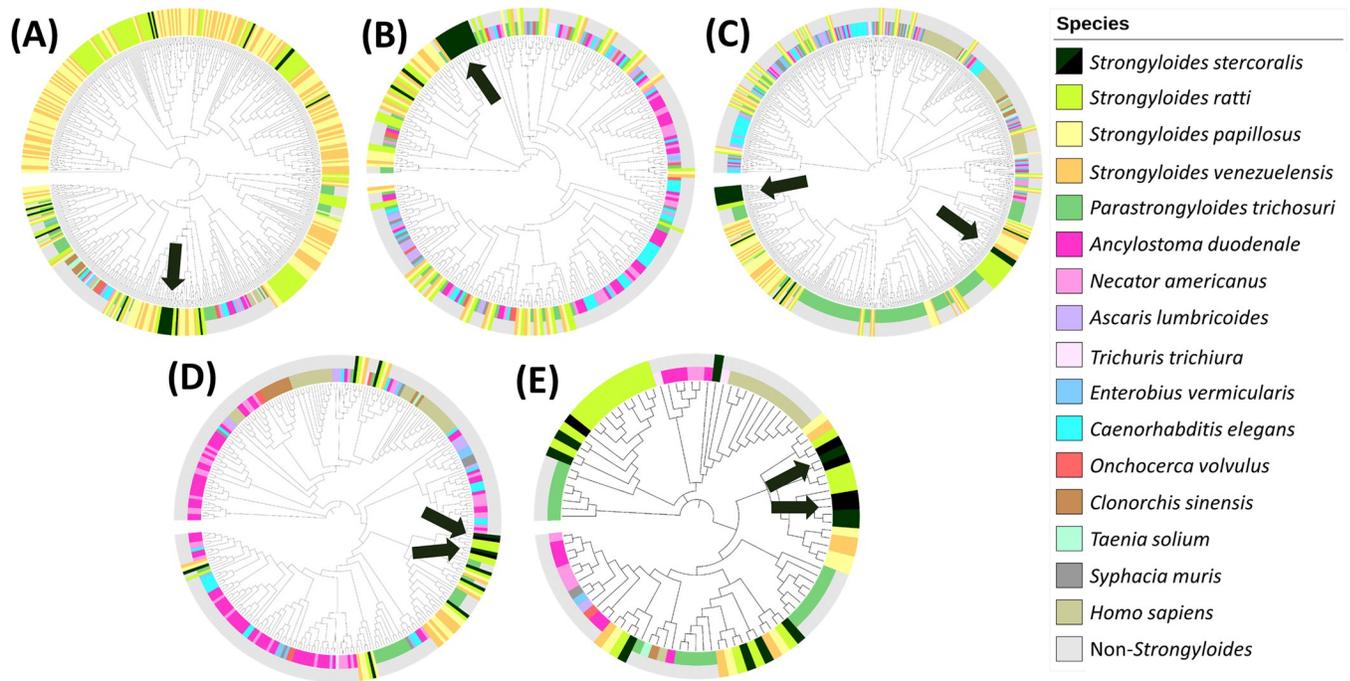


Fig 4. Phylogenetic alignment of *Strongyloides stercoralis* high priority family proteins with other *S. stercoralis* proteins and outgroups. Proteins of the five priority families of *S. stercoralis* (dark green), from the differentially expressed dataset, in phylogenetic alignment with BLAST hits from *S. stercoralis* itself (black) and outgroups (other colours). Colours reaching the outer edge represent *Strongyloides* species, whereas shorter colour bands are non-*Strongyloides* outgroups. (A) SCP/TAPS, (B) transthyretin-like (TTL), (C) acetylcholinesterase (AChE), (D) aspartic peptidase, (E) prolyl oligopeptidase (POP). Arrows indicate clusters of *S. stercoralis* proteins containing species-specific regions; derived from both gut- and non-gut-dwelling differential expression data, augmenting detection of lack of species specificity. These results originate from methods in Fig 1, box K.

<https://doi.org/10.1371/journal.pntd.0010777.g004>

alternative isoforms which are also included here. Multiple sequence alignment indicated that 550 (94.2%) of the 584 *S. ratti* E/S proteins had a *S. stercoralis* orthologue at the selected similarity level (e-value 1E-50) and that 284 (51.6%) of these had multiple homologues in *S. stercoralis* (S3 File).

To identify possible *S. stercoralis* E/S proteins among those differentially expressed in the host gut, we compared the 1,057 *S. stercoralis* orthologues with the 328 identified by transcriptomic data as DE in gut-dwelling life stages of *S. ratti*. Seventy seven (23.5%) proteins were shared between both data sets, of which 58 (28.6%) were shared between the 203 gut-stage DE proteins and the E/S orthologues (Fig 1, box H).

To investigate gene expression of the E/S orthologues, we extracted data for the 1,057 *stercoralis* proteins from RSEM analysis of the entire 13,098 active *S. stercoralis* genes (Fig 1, box H). Protein families were assigned by a combination of BlastKOALA, which identified 537 (50.8%) of the E/S orthologues, and the *S. ratti* protein identifications provided by Hunt *et al.* (2016) [34] (S4 File and Fig 1, box G). Relevant proteins were assigned to the five ‘priority protein families’ (Fig 3).

Predicted epitopes

Within the 328 confidently identified DE *S. strongyloides* proteins, BepiPred and bcepred jointly predicted epitopes in 104 proteins, 78 and 62 proteins respectively, with 36 proteins containing epitopes predicted by both tools (S2 File, Figs 3, and 1, box M). Within the 78 proteins, BepiPred predicted 125 epitopes, and within 62 proteins bcepred predicted 108 epitopes (S5 File). Fifty six epitopes contained overlap or identity between the two prediction tools (S5

Table 3. Candidate coproantigens of *S. stercoralis* based on upregulation in parasitism, amino acid sequence specificity to *Strongyloides* or *S. stercoralis*, and containing a predicted epitope. Gene accession numbers can be found in WormBase ParaSite or UniProtKB. Amino acid sequences are given in S6 File. PF: parasitic female, E/S: excretory/secretory.

| Protein family | Gene | Residues | Length | Peptide is specific to <i>Strongyloides</i> | Protein is upregulated in gut-dwelling life stages | Protein contains signal peptide | Protein occurs in E/S orthologues | Protein contains predicted epitope, surface exposed |
|-----------------------------|-----------------|----------|--------|---|--|---------------------------------|-----------------------------------|---|
| SCP/TAPS | SSTP_0000990000 | 363–743 | 381 | Yes | Yes- in PF | No | No | No |
| Transthyretin-like (TTL) | SSTP_0000700800 | 1–177 | 177 | Yes | Yes- in PF | No | No | Yes |
| | | 22–50 | 29 | Yes | | | | |
| | | 51–86 | 36 | Yes | | | | |
| | | 65–159 | 95 | Yes | | | | |
| | | 87–116 | 30 | Yes | | | | |
| | | 125–160 | 36 | Yes | | | | |
| Acetylcholinesterase (AChE) | SSTP_0000274700 | 17–43 | 27 | Yes | Yes- in PF | Yes | Yes | No |
| | | 63–104 | 42 | Yes | | | | |
| | | 187–207 | 21 | Yes | | | | |
| | | 243–266 | 24 | Yes | | | | |
| | | 268–306 | 39 | Yes | | | | |
| | | 308–328 | 21 | Yes | | | | |
| | | 372–451 | 80 | Yes | | | | |
| | | 551–607 | 57 | Yes | | | | |
| | SSTP_0000509400 | 85–103 | 19 | Yes | Yes- in PF | Yes | Yes | Yes |
| | | 390–411 | 22 | Yes | | | | |
| Aspartic peptidase | SSTP_0000164500 | 107–140 | 34 | Yes | Yes- in PF | Yes | Yes | No |
| Prolyl oligopeptidase (POP) | SSTP_0001108800 | 221–266 | 46 | Yes | Yes- in PF | No | Yes | No |
| | | 384–421 | 38 | Yes | | | | |
| | | 422–507 | 86 | Yes | | | | |
| | | 741–779 | 39 | Yes | | | | |

<https://doi.org/10.1371/journal.pntd.0010777.t003>

File). Predicted epitopes ranged from 9 residues to entire proteins of up to 651 amino acids (aa) with BepiPred, and 8 to 66 aa with bcepred. These regions were given greater scrutiny in the context of species specificity and antibody accessibility (Fig 1, box N).

S. stercoralis orthologues (n = 1,057) of the *S. ratti* E/S proteome were also submitted to BepiPred and bcepred which predicted 747 and 62 epitope regions respectively, in a total of 324 of the proteins. These ranged from 9 to 99 residues and contained 49 epitope sequences that overlapped, originating from 40 proteins (S5 File and Fig 1, box M).

***S. stercoralis*-specific candidates identified by phylogenetic comparisons**

The five *S. stercoralis* protein families linked to parasitism and among the DE proteins were analysed for *S. stercoralis* genus or species-specificity. Separate BLAST searches of each protein family against 15 outgroups revealed the *S. stercoralis* proteins most likely to contain specific regions (Figs 4 and 1, box K).

S. stercoralis proteins from the clusters identified in the phylogenetic trees were examined in alignment with outgroup representative homologues with the most and least similarity (Fig 1, box L). From the alignments, the *S. stercoralis* regions with least homology to outgroups were selected as candidate antigens. In some cases these included the entire protein. Results of this analysis are detailed below by protein family.

SCP/TAPS coproantigen candidates. Twenty one differentially expressed proteins were identified as definite (n = 19) or possible (n = 2) members of the SCP/TAPS protein family

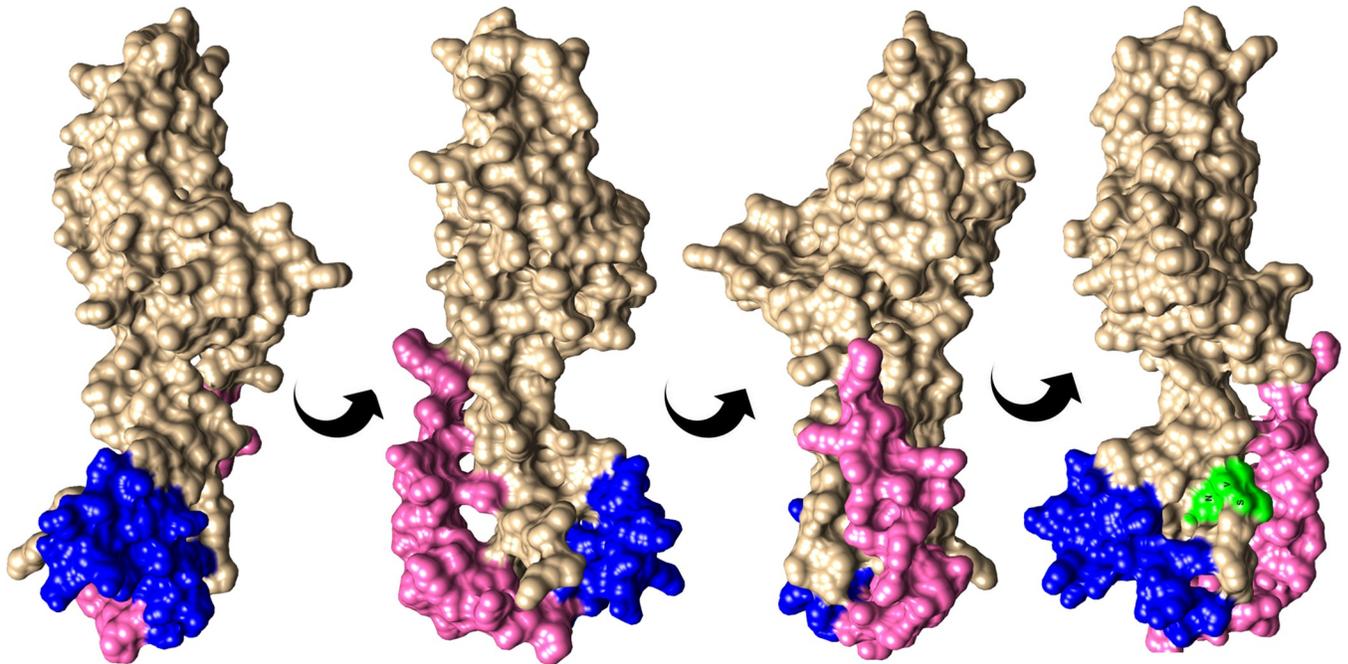


Fig 5. A tranthyretin-like (TTL) protein of *Strongyloides stercoralis* in rotated view showing surface-exposed epitope regions and N-linked glycosylation sites. Accession number SSTP_0000700800. Blue, BepiPred predicted epitope aa's 124–143; pink, bcepred predicted epitope aa's 99–123; green, predicted N-linked glycosylation site at N165 in the motif NVS. The bcepred epitope (pink) extends to residue 141 but has been shortened to show the BepiPred epitope (blue), which overlaps with it. These results originate from methods shown in Fig 1, boxes M–P.

<https://doi.org/10.1371/journal.pntd.0010777.g005>

(Fig 3 and S2 File). The two 'possible' SCP/TAPS proteins clustered separately among CAP-domain proteins (Fig 3) and were therefore not included as definite members of this protein family. In phylogenetic analysis following alignment against outgroup genomes, seven of the 19 SCP/TAPS proteins formed a cluster of higher *S. stercoralis* specificity (Fig 4A, arrowed) (SSTP_0001008500, 8600, 8700, 8900, 0000511800, 512000, 513400). Multiple sequence alignment (MSA) and BLAST searching showed that these SCP/TAPS proteins contained several regions of apparent *S. stercoralis* species specificity, however, most were more highly expressed in non-gut life stages, such as tissue-migrating larvae. In total 8 of the 19 were expressed more in gut-dwelling life stages. Therefore SCP/TAPS from outside the cluster (in Fig 4A) were also viewed in MSA. A species-specific region was identified in gut-stage DE protein, SSTP_0000990000, consisting of a large 381 aa sequence of this protein which is upregulated in the parasitic female life stage (Table 3).

Tranthyretin-like coproantigen candidates. Fourteen (14) TTL proteins were differentially expressed between gut and non-gut-dwelling life stages with 100% CI, all of which were expressed more in gut-dwelling life stages, almost exclusively in parasitic female worms (Fig 3 and S2 File). Twelve of these proteins grouped as a cluster while the other 2 were more similar to other protein families or features (Fig 3). When aligned against homologous proteins from the outgroup species, 11 TTL proteins clustered together (Fig 4B, arrowed). All 14 TTL proteins were inspected visually in sequence alignment with selected outgroup proteins and five (SSTP_0000700800, 700900, 1222000, 0485800, 1133200) showed greater specificity to *S. stercoralis* and low sequence similarity to any of the outgroup homologues. However, when many of the possible *Strongyloides*-specific regions were BLAST searched separately, they were not sufficiently specific to be used as coproantigens. In particular, the search revealed that the amino acid sequence VTCDGKPL in protein SSTP_0000485800 is conserved across many

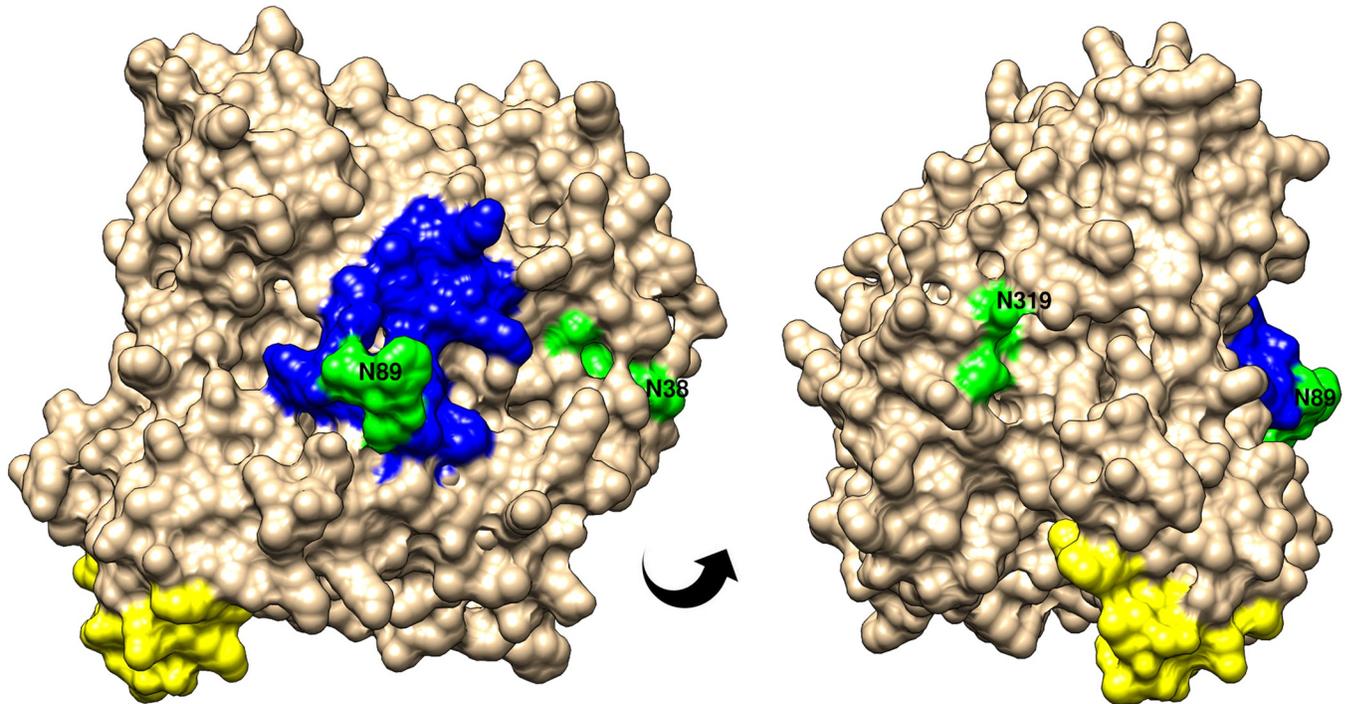


Fig 6. *Strongyloides stercoralis* acetylcholinesterase in rotated views showing surface-exposed epitope regions and N-linked glycosylation sites. Accession number SSTP_0000509400. Predicted features indicated in colour: blue, BepiPred 2.0 predicted epitope at residues 85–103; and yellow at 390–411; green, predicted N-linked glycosylation sites at N38 (NVT), N89 (NFS) and N319 (NLT). The site at N89 is within the blue epitope sequence. Both predicted epitope regions (blue and yellow) were selected as candidate coproantigens in this study. These results originate from methods shown in Fig 1, boxes M-P.

<https://doi.org/10.1371/journal.pntd.0010777.g006>

nematode genera and should therefore be avoided in any candidate antigen. Two of the TTL proteins (SSTP_0001222000 and SSTP_0000700800) did contain one or more regions of *Strongyloides* species or genus specificity, giving a total of 6 TTL candidate coproantigens (Table 3).

In addition to high *Strongyloides*-specificity across its whole sequence of 177aa, TTL protein SSTP_0000700800 also contained predicted epitopes. To investigate the position of these, the protein was 3D modelled to a template generated from TTR-52 protein of *C. elegans* (PDB accession: 3UAF) (Figs 5 and 1, box O). This model aligned to 89 residues (aa 2–90; 50% of the sequence) with 99.9% confidence. This protein contained a predicted glycosylation site at position N165 and, although not predicted to have a signal peptide, it was indicated as an ‘extracellular or secreted’ protein in UniProt [60] and is a known E/S protein family, therefore it is more likely to be glycosylated (method Fig 1, box P). The whole sequence of 177 amino acids was selected as a candidate coproantigen due to its *Strongyloides* specificity, high expression by parasitic females and the presence of several predicted epitopes (Table 3).

Acetylcholinesterase (AChE) coproantigen candidates. Nineteen AChE proteins were in the DE dataset, of which 18 were expressed more highly in gut-dwelling stages, particularly in parasitic female worms (Fig 3). The other AChE protein was highly expressed in infectious and tissue-migrating larvae (S2 File). All 19 proteins formed a cluster when compared with other DE protein families, indicating greater sequence similarity. The 19 proteins were then aligned with BLAST hits from outgroup species and most grouped into two distinct clusters in the corresponding phylogenetic tree, one cluster of 4 *S. stercoralis* proteins, the other of 10 (Fig 4C, arrowed), all of which were DE in gut-dwelling life stages. The cluster of 4 consisted of SSTP_0000274700, 671000, 638700 and 670800, all of which contained several regions of potential *S. stercoralis* specificity by visual inspection of the alignment. Equally, the 10 proteins

within the other cluster contained multiple regions of possible *Strongyloides* specificity which were then analysed by BLAST to confirm this specificity. For AChE, 10 candidate antigens originated from 2 proteins: SSTP_0000274700 and 509400 (Table 3).

From the larger cluster in Fig 4C, one of the sources of candidate antigens, AChE protein SSTP_0000509400 was modelled with 100% confidence to 6 templates that jointly covered aa's 17–551 (95%) with about 30% sequence identity. Peptides with potential *S. stercoralis* species-specific sequences were annotated on the model to view their surface exposure (Fig 6). BepiPred 1.0 and bcepred both failed to identify epitopes in this protein. However, BepiPred 2.0 did predict epitope regions, with moderate specificity and surface exposure. This AChE protein contained multiple potential glycosylation sites, three of which, at positions 38, 89 and 319, were predicted with high confidence on this known glycoprotein (Fig 6).

Aspartic peptidase coproantigen candidates. Only 4 aspartic peptidases appeared in the DE proteins, all of which were expressed more highly in gut-dwelling life stages. Two of these were less reliably identified in this protein family so the other two (SSTP_0000164500 and 164700), with very high expression levels in parasitic female worms and both orthologues of E/S proteins, were analysed for *S. stercoralis* specificity. An additional 9 homologous proteins from *S. stercoralis* itself were included and all 11 proteins were then compared with the outgroups (Fig 4D, arrowed). The two DE proteins had low sequence similarity with each other. However, BLAST searching indicated that most of the possible species-specific regions lacked *Strongyloides* specificity and only one suitable peptide was identified from SSTP_0000164500 (Table 3). 3D modelling of this protein could not be achieved with any reliability and there were no predicted epitopes in either of the DE aspartic peptidases. The only candidate coproantigen from this protein family was selected for its *Strongyloides* specificity, high expression and likely presence in E/S material.

Prolyl oligopeptidase (POP) coproantigen candidates. The 5 differentially expressed POP proteins, all very highly expressed in parasitic females, were combined with a further 10 *S. stercoralis* homologues. In the phylogenetic tree, all 5 clustered with other *Strongyloides* proteins and away from non-*Strongyloides* outgroups (Fig 4E). Four of the DE proteins (SSTP_0000289100, 1108800, 1108500, 1019400) had higher species-specificity, clustering with other *S. stercoralis* homologues (Fig 4E, arrowed). One of these 4 proteins, SSTP_0000289100 contained two BepiPred-predicted epitopes and was very highly expressed in parasitic females, as well as being in the E/S orthologues (Fig 3 and S2 File). However, the epitope regions had moderate sequence identity to other nematodes, a *Staphylococcus* species and to *Plasmodium vivax*, suggesting widespread conservation. Therefore, these peptides were not considered sufficiently specific to *Strongyloides*.

Analysis of the POP proteins in multiple sequence alignment and subsequent search for similar proteins using BLASTP against NCBI nr and Nematoda databases, revealed several regions of high specificity to *S. stercoralis* in one of the DE proteins mentioned above (SSTP_0001108800). Therefore, although it did not contain predicted epitopes, this protein was orthologous to E/S proteins and *Strongyloides*-specific peptides were selected as candidate coproantigens (Table 3). Amino acid motifs conserved across genera, and therefore not *Strongyloides* specific, were removed from the sequences originally selected from the MSA, these included: DKLEN, KTDSK, RNAH and DIFAFI.

Summary of candidate coproantigens

Table 3 presents selected candidate coproantigen protein regions that satisfy the criteria of being present in stool, specific to *Strongyloides* or *S. stercoralis*, and being antigenic. The full amino acid sequences of all candidate antigens are given in FASTA format in S6 File.

Discussion

The paucity of data on *S. stercoralis* infection prevalence and its low profile compared with the other STH species are largely due to inadequate diagnostics and lack of a single gold standard. While serology has the highest sensitivity for active disease, it is unsuitable for monitoring treatment outcome or defining cure in a timely manner [23,24]. Incomplete cure and reinfection post-treatment may occur [8,61,62]. Therefore, we aimed to identify specific diagnostic targets from the nematode that could be captured by a rapid antigen detection test on stool samples. Such coproantigen assays are commercially available for *Giardia* and *Cryptosporidium* and have been developed for a wide range of human and animal parasites including *Ascaris* [63], *Fasciola* [64], *Echinococcus* [65], *S. ratti* [28], *S. venezuelensis* [29,30], *Opisthorcis* [66], *Toxocara* [67] and *Entamoeba histolytica* [68], among others. These assays employ either somatic, E/S material, or known antigens as targets.

We used open access data sources, published literature and freely-accessible online protein analysis tools to shortlist candidate antigens, based on the three criteria: presence in infected stool; *Strongyloides* or *S. stercoralis* specificity, and antigenicity. A similar study by Culma (2021) investigated cellular location, antigenicity and allergenicity, among other features, in the *S. stercoralis* proteome to identify potential vaccine and diagnostic targets [69]. Recently, Dishnica et al (2023) [70] published the somatic proteome of *S. stercoralis* iL3 isolated from a clinical case and conducted a similar computational pipeline to discover possible novel serological antigens. While their main candidates do not overlap with those from the present study, three appear in the DE dataset (S2 File) as they were differentially expressed more highly in non-gut-dwelling life stages, predominantly iL3, as would be expected (Q9UA16/SSTP_0001008900, A0A0K0E2F4/SSTP_0000367400, A0A0K0DTP5/SSTP_0000060800), and three proteins were present in the E/S orthologues (A0A0K0E6J0/SSTP_0000511900, Q9UA16/SSTP_0001008900, A0A0K0DTP5/SSTP_0000060800). This suggests that there may be overlap between somatic and E/S proteomes which is worth investigating in future studies.

Here, we focused on proteins that were differentially expressed between gut-dwelling and non-gut-dwelling life stages of *S. stercoralis*, according to RNA-seq data [37]. Seven protein families have been identified by other studies as expanded in the genomes of parasitic nematodes, and upregulated in parasitic life stages [38]. Studies of the *Ancylostoma* hookworm E/S proteome and *S. venezuelensis* and *S. stercoralis* somatic larval proteomes detected some of the same families (SCP/TAPS, proteases and TTL) indicating possible presence in stool [70–72]. However, specific proteins within these families must still be selected as candidate antigens because every individual protein is not expressed simultaneously. To identify coproantigen diagnostic targets, we have performed a detailed analysis of 5 of the 7 parasitism-associated protein families. We also identified *S. stercoralis* orthologues of the *S. ratti* E/S proteome [34,48].

We found an overlap of 77 proteins (5.9% of the total) between DE proteins and E/S orthologues. This limited overlap may reflect post-transcriptional control of expression [34]. Thus, grouping together transcriptomic data of life stages with different gene expression profiles may have limited our resolution of stage-specific coproantigens. E/S proteomics are an alternative starting point for coproantigen discovery. However, when we viewed gene expression data for all the E/S orthologues, no single life stage accounted for all the parasitic female E/S proteome. Although the *S. stercoralis* E/S orthologues broadly represented the *S. ratti* E/S proteome dataset, there were 1,057 compared to the original *S. ratti* 550 dataset. Therefore, differential gene expression between species is also likely to be a complicating factor and ultimately, the E/S proteome of the species of interest would be most suitable for coproantigen discovery. However, shared epitopes between these closely related species still warrant the use of all available data [52].

Priority protein families and species specificity

Phylogenetic trees of *S. stercoralis* DE proteins and their homologues from a selection of out-groups focused our analysis on individual proteins from the priority protein families. These proteins were more likely to contain species- or genus-specific regions.

SCP/TAPS. The first of the priority protein families analysed here, SCP/TAPS, is among the CAP domain-containing proteins and is proposed to have a role in modulating the host immune response [38]. Here, we identified 21 SCP/TAPS proteins and 6 additional CAP-domain-containing proteins among the DE proteins. Many of the species-specific proteins or regions in this protein family, were highly expressed in tissue migrating larvae, rather than gut-dwelling life stages, therefore, only one protein expressed in adult female worms was the source of candidate antigens. This protein family has been studied in other parasitic nematodes, particularly the hookworms *Ancylostoma caninum* and *Necator americanus*, in which it has numerous members [73,74]. Multiple of these SCP/TAPS proteins, which include the NIE antigen, have been demonstrated to be antigenic and recognised by anti-E/S antisera, showing that they do possess at least two of the characteristics predicted by this pathway [73].

Transthyretin-like proteins. Transthyretin-like proteins have been identified in the E/S material of nematodes from 5 major clades, including parasitic and non-parasitic species [75]. Two TTL proteins were sources of candidate coproantigens in the present study, both of which were expressed highly in the parasitic female worm and very little in other life stages. Neither of these DE proteins were among the E/S orthologues, however other TTL proteins were, and were expressed highly across multiple life stages. In addition, a TTL protein has been identified as a potential vaccine or diagnostic candidate through a similar bioinformatic pipeline, adding to the evidence for this protein family as a source of diagnostic targets [69]. Various functions have been identified for TTL proteins including involvement in larval development, innate immunity of the nematode itself, apoptosis, and subduing the immune response to favour worm survival [76–78].

Acetylcholinesterase. Secreted AChE has a possible role in enabling certain parasitic helminths to evade host expulsion mechanisms from mucosal surfaces [38,79,80]. The transcriptomic and E/S proteomic data strongly supported this, with AChE family proteins in the E/S orthologues being expressed almost exclusively in the parasitic female life stage (S4 File). Secreted AChE differs from the neuromuscular protein in structure, gene family and substrate, being less specific to acetylcholine [79]. We found moderate homology between the predicted epitope regions of a *S. stercoralis* AChE and other nematode species. We identified candidate antigens in two AChE proteins. One of the candidate antigen proteins (SSTP_0000274700), contained 8 regions of *Strongyloides* specificity, when less specific regions were excluded, comprising over 50% of the full length 607 aa protein.

Aspartic peptidase. The aspartic peptidase family of enzymes is named for the creation of the active site from aspartic acid residues. Very few aspartic peptidases were DE, but had high expression in parasitic females and homologues to *S. ratti* E/S proteins. Aspartic peptidases play numerous roles in parasitic helminths including digestion, immune evasion and tissue invasion, among others linked to feeding within the host gut epithelium where the adult worm resides [81]. We identified a single candidate antigen from the few examined aspartic peptidases which had considerable homology to aspartic proteases from other *Strongyloides* species, but little to any other relevant species, whereas the full protein had high homology to multiple relevant nematode species including hookworm.

Prolyl oligopeptidase. Five POP family proteins were differentially expressed between gut and non-gut dwelling life stages. All 5 were very highly expressed in parasitic females where they may be involved in defending against the host immune and parasite expulsion

response. In the trematode *Schistosoma mansoni*, POP enzymes have been found to cleave peptide hormones and neuropeptides [82]. Inhibiting POP activity in *S. ratti* lead to immobility in a concentration-dependent manner, even in *in vitro* conditions, indicating that this protein family is also vital to worm survival [48]. Only one of the 5 POP proteins contained regions of sufficient *Strongyloides* specificity for consideration as a coproantigen.

In addition to the shorter, specific sequences listed here, there is broader potential to express whole recombinant proteins such as those indicating higher *Strongyloides* specificity, in order to raise polyclonal antibodies for screening. Alternatively, specific monoclonal antibodies could be developed against these whole proteins and again, screened for *Strongyloides* specificity, as conducted by Abduhaleem et al. (2019) with a monoclonal antibody raised against *S. ratti* somatic antigens [83].

Epitope prediction

We performed epitope prediction on the DE proteins and E/S proteome homologues using two open access online tools, which yielded many predicted epitope peptides. An alternative to this would be to scan the entire genome for epitopes, a method applied to vaccine candidate discovery [84]. The challenge faced by this approach is the complexity of conformational epitopes compared with linear peptide epitopes. Antibodies frequently bind to conformational epitopes formed by the 3D structure of the antigen, which therefore cannot easily be detected by sequence analysis alone [56].

The availability of 3D protein models can assist with selecting conformational epitopes by modelling a sequence onto the structure of a homologous protein and revealing adjacent amino acids on the surface of the protein [85]. Models do not necessarily have high sequence identity to the query sequence but this does not decrease the confidence in the model. Confidence in 3D models of >90% indicates that the protein adopts the overall folds of the model but may differ from the native protein in surface loops [57], thus this method is not guaranteed, but provides a good indication for selecting candidate antigens. *Ab initio*-modelled regions, where the sequence was not covered by the model, have very poor accuracy and should therefore be interpreted with caution and not used as the sole basis for selecting conformational epitopes. The field of 3D structure modelling has been progressed in leaps by the use of machine learning [86,87], with AlphaFold recently being applied to *Strongyloides* proteomic data [70].

We saw differences in predicted epitopes between the computational versus ‘by eye’ approach to selecting epitope regions. The DE protein dataset contained longer predicted epitopes due to the decision, where relevant, to extend predicted epitopes across a short region of lower epitope score whereas the computational selection worked only on the exact score threshold and would not join two adjacent high scoring regions.

Glycoprotein antigens were not considered in this study, apart from the presence of potential N-linked glycosylation sites on candidate protein antigens. Glycans form existing species-specific, highly antigenic diagnostic antigens, including CCA and CAA of *Schistosoma mansoni* and *Schistosoma* genus trematodes respectively [88], and LAM of *Mycobacterium tuberculosis* [89]. They have also been implicated in lysate seroantigen of *S. stercoralis* [90]. In helminths, glycan structures may not only be species-specific, but also life-stage specific [91]. Glycans may obscure some of the protein epitopes predicted here, particularly in the secreted AChE which is highly glycosylated. This is to be expected as glycans form many of the host-parasite interactions [91]. In addition, secreted candidate antigens may also contain O-linked glycans, via oxygen atoms of serine or threonine, which are not easily predicted. Although we have excluded potential glycan epitopes, they could be accounted for to some extent by

expressing antigens of interest, ideally in a closely-related system, potentially *C. elegans* or even *Strongyloides* itself [92]. As an alternative, glycans could be excluded altogether by synthesising peptides or expressing recombinant proteins in bacteria, thus focusing the antigen search purely on proteins, as we have done here.

Limitations and future work

We have described a methodological approach to discovery of diagnostic antigens. The process is predictive, based on computational analyses of ‘omic’ data. Predictions rely on the integrity of data sources, efficacy of software, and correct use of cut-off values at multiple stages of analysis. WormBase ParaSite, the source of most of the sequences used, is a curated database, regularly updated with high quality additional data and regarded as predominantly reliable. Thus, large-scale sequencing errors that might affect our use of software, such as antigen prediction or BLAST analysis, are rare. Single nucleotide and small-scale sequencing errors are unlikely to affect results of the predictive pathways. With the E value cut-off used to select matching proteins by BLAST analysis, some proteins returned multiple hits. This may be due to paralogs, or may reflect an E value that is too lenient. Even in the latter case, over-inclusion of proteins at this stage, pared down by downstream analysis, is preferable to rejection of candidates. Proof of candidate antigen validity requires follow-up experimental laboratory research, which can reveal whether constraints and cut-off values have been too strict or too lenient.

Another limitation is prediction of antigens that are present in the stool. In our pipeline, a protein being excreted or secreted by a gut-dwelling stage of the parasite has been used as a proxy for presence in the stool. This rests on the assumption that the protein passes through the gut, is present in the stool in sufficient quantity and with antigenic properties unchanged by digestion or denaturation. As *S. stercoralis* establishes and matures in the small intestine, E/S proteins will not be subject to the denaturation and enzymatic degradation in the stomach but may be affected by enzymes in the small intestine. However, post-parasitic larvae, which hatch from eggs laid in the intestine, migrate through the small and large intestine, and can be found in faeces, so any E/S proteins produced by these stages, especially lower down in the intestinal tract, are less likely to be affected by host processes. The validity of these assumptions can be investigated by analysis of infected stool and an appropriately sensitive diagnostic developed based on laboratory-confirmed antigens and validation in the field.

A further consideration is the focus of the approach on differentially expressed proteins and certain parasitism-associated protein families. There is rationale to targeting proteins with these characteristics, because they are proven to be expressed at higher levels in the relevant life cycle stages, are likely to be excreted or secreted and species or genus-specific due to adaptive radiation and protein family expansion. The focus on these groups limits the scope of an ‘omic’ approach, however, proteins that do not belong to one of these two groups might still have the characteristics of a candidate coproantigen. The approach presented here can therefore be applied to other helminth infections, but may be complimented by an expanded approach that uses whole genomic or proteomic data without focussing on parasitism-associated protein families or life stages, with the results of the two methods being used in conjunction.

Other protein families with potential to be sources of diagnostic coproantigens include: enolase, common to *Schistosoma japonicum* [93], *Echinostoma* and *Fasciola* [94], *Onchocerca* [95] and *Trichuris* [96], as well as among *S. stercoralis* E/S orthologous proteins reported here and with constitutive high expression across all life stages (S4 File); astacins and proteinase inhibitors, as other parasitism-associated protein families [38]; protein 14-3-3 from E/S and somatic extract of *Strongyloides* [70,97,98], *Ascaris* [99], *Schistosoma* and *Ancylostoma* [72]. In

addition, collagen, which forms some 80% of the outer cuticle of the nematode and was prominent among the DE proteins [100]. As an antigen, collagen would have to be carefully analysed for species specificity and the influence of glycosylation on its availability to capture antibodies [101].

Possible data sources for further antigen searches continue to be added to and include: an analysis of E/S proteomes by Tritten *et al.* (2021) [75]; analysis of *S. venezuelensis* somatic proteins by Fonseca *et al.* (2019) [71], and *S. venezuelensis* E/S proteins by Maeda *et al.* (2019) [102]; also *S. stercoralis* iL3 somatic proteins by Dishnica *et al.* (2023) [70]. Somatic protein datasets would include tegument antigens as might be present in stool if whole or parts of larvae were present, or if the proteins were secreted via non-classical pathways, as proposed for some *S. venezuelensis* proteins [102]. Vaccine candidates for *S. stercoralis*: sodium potassium ATPase (SsEAT), tropomyosin, and a galectin (LEC-5) [103,104] may also generate effective antibodies for antigen capture, if the relevant antigens are detectable in stool, as could serological antigens SsIR [105] and NIE [106], the latter of which has been detected in serum [107].

The most immediate work arising from this research is to validate the 'omic' pathway via wet laboratory analyses, specifically proteomic methods to detect and assess the antigenicity of proteins present in stool. Production of predicted candidates followed by antigenic testing could also be used to investigate the results of the 'omic' pathway as well as for development of lateral flow design for successful peptides. Results could be used to inform and refine results from the 'omic' pipeline, however, well-characterised stool sample collections are required to pursue diagnostic development.

Antigenic diversity must be considered, due to geographic differences between nematode strains [108]. This has impacted on diagnosis and vaccination for other parasitic infections [109,110]. Diversity can be readily investigated by amplifying and sequencing the genes of interest from a wide geographic range of samples, or investigating differences between actual protein sequences from clinical isolates [70] and their predicted counterpart from genomes. For *S. stercoralis*, this is especially required, because the reference genome strain PV001 originates from a dog infection with UPD (University of Pennsylvania dog) strain [111].

Conclusion

We have presented a detailed analysis of *S. stercoralis* proteins, leading to the selection of diagnostic coproantigens. We have identified multiple *S. stercoralis* candidate protein antigen sequences with evidence for their specificity to *Strongyloides* or *S. stercoralis* from phylogenetic and sequence comparison with relevant other species. Evidence supporting their presence in infected stool was assessed by belonging to parasitism-associated protein families, upregulation in gut-dwelling life stages, presence in E/S material of other helminths, and being among *S. stercoralis* orthologues of the *S. ratti* E/S proteome. Antigenicity was predicted using epitope prediction tools and 3D structure modelling. Peptides or whole proteins analysed and presented here form a selection of promising candidates for raising antibodies against and capturing *S. stercoralis* antigen in stool, with potential adaptation to prototype point-of-care rapid diagnostic tests. Access to coproantigen tests for *Strongyloides* will speed up the control efforts of this neglected public health problem.

Supporting information

S1 File. Number of outgroup protein hits for *S. stercoralis* protein families from the differentially expressed (DE) dataset.

(DOCX)

S2 File. *S. stercoralis* genes (n = 328) differentially expressed in gut-dwelling (dark blue) versus non-gut-dwelling (light blue) life stages of the nematode with those containing predicted epitopes indicated. Transcriptomic data from Stoltzfus *et al.* (2012), gut and non-gut groupings and differential analysis from the present study. Gene accession numbers (starting SSTP) are given for WormBase ParaSite or UniProtKB and life stage transcriptomic data accession numbers (starting ERR) for NCBI SRA. Expression levels are in normalized read counts where red is high and white is low.
(XLSX)

S3 File. BLAST results: *S. stercoralis* orthologues of the *S. ratti* E/S proteome. *S. ratti* E/S proteomic data from Hunt *et al.* (2016) was BLAST searched against the *S. stercoralis* genome-derived proteome using blast+ with e-value 1e-50
(XLSX)

S4 File. Protein identity, features and gene expression levels of *S. stercoralis* orthologues of *S. ratti* E/S proteomic data. *S. ratti* E/S proteins (from Hunt *et al.*, 2016) were searched against a custom blast+ database consisting of the *S. stercoralis* genome-derived proteome (from WBPS v8) to obtain homologues. E-value -50 and word size 2 were used to perform the search. Protein identities are given for the original *S. ratti* proteins (from Hunt *et al.*, 2016) and the BLAST KOALA results from this study along with corresponding KEGG orthology. Gene expression in normalised read count is from analysis of data from different life stages of *S. stercoralis* (obtained from Stoltzfus *et al.*, (2012)). Overlap is indicated with proteins which were also differentially expressed (DE) with 100% certainty between gut-dwelling and non-gut-dwelling life stages, as determined for this study. Proteins containing predicted epitopes according to two tools: BepiPred 1.0 and bcepred are indicated as well as proteins predicted by SignalP to contain a signal peptide. Accession numbers starting ERR refer to RNA-seq data in NCBI SRA and those starting SSTP are *S. stercoralis* gene transcripts which may be found in UniProt or WormBase ParaSite. Life-stage abbreviations are explained in the manuscript text.
(XLSX)

S5 File. Predicted epitopes in *S. stercoralis* proteins using BepiPred and bcepred. ES: Orthologues of *S. ratti* E/S proteome. DE: Differentially expressed proteins between gut-dwelling and non-gut-dwelling life stages.
(XLSX)

S6 File. Candidate coproantigen sequences in fasta format. FASTA headers are in the format: accession number, start aa, end aa.
(DOCX)

Acknowledgments

We thank Dr Andrew Falconar for useful discussions and advice during this project.

Author Contributions

Conceptualization: Michael A. Miles.

Formal analysis: Tegwen Marlais, Carlos Talavera-López.

Funding acquisition: Tegwen Marlais, Michael A. Miles.

Investigation: Tegwen Marlais, Jack Bickford-Smith, Carlos Talavera-López, Hai Le, Fatima Chowdhury.

Methodology: Tegwen Marlais, Carlos Talavera-López, Michael A. Miles.

Project administration: Tegwen Marlais, Michael A. Miles.

Resources: Michael A. Miles.

Supervision: Michael A. Miles.

Visualization: Tegwen Marlais.

Writing – original draft: Tegwen Marlais, Michael A. Miles.

Writing – review & editing: Tegwen Marlais, Jack Bickford-Smith, Michael A. Miles.

References

1. Schär F, Trostorf U, Giardina F, Khieu V, Muth S, Marti H, et al. *Strongyloides stercoralis*: global distribution and risk factors. PLoS Negl Trop Dis. 2013; 7(7):e2288. <https://doi.org/10.1371/journal.pntd.0002288> PMID: 23875033
2. Forrer A, Khieu V, Schar F, Vounatsou P, Chammartin F, Marti H, et al. *Strongyloides stercoralis* and hookworm co-infection: spatial distribution and determinants in Preah Vihear Province, Cambodia. Parasit Vectors. 2018; 11(1):33. <https://doi.org/10.1186/s13071-017-2604-8> PMID: 29329561
3. Buonfrate D, Baldissera M, Abrescia F, Bassetti M, Caramaschi G, Giobbia M, et al. Epidemiology of *Strongyloides stercoralis* in northern Italy: results of a multicentre case-control study, February 2013 to July 2014. Euro Surveill. 2016; 21(31). <https://doi.org/10.2807/1560-7917.ES.2016.21.31.30310> PMID: 27525375
4. Becker SL, Sieto B, Silue KD, Adjossan L, Kone S, Hatz C, et al. Diagnosis, clinical features, and self-reported morbidity of *Strongyloides stercoralis* and hookworm infection in a co-endemic setting. PLoS Negl Trop Dis. 2011; 5(8):e1292. <https://doi.org/10.1371/journal.pntd.0001292> PMID: 21886853
5. Minciullo PL, Cascio A, Gangemi S. Association between urticaria and nematode infections. Allergy Asthma Proc. 2018; 39(2):86–95. <https://doi.org/10.2500/aap.2018.38.4104> PMID: 29490766
6. Abrescia FF, Falda A, Caramaschi G, Scalzini A, Gobbi F, Angheben A, et al. Reemergence of strongyloidiasis, northern Italy. Emerg Infect Dis. 2009; 15(9):1531–3. <https://doi.org/10.3201/eid1509.090191> PMID: 19788836
7. Atkins NS, Lindo JF, Lee MG, Conway DJ, Bailey JW, Robinson RD, et al. Humoral responses in human strongyloidiasis: correlations with infection chronicity. Trans R Soc Trop Med Hyg. 1997; 91(5):609–13. [https://doi.org/10.1016/s0035-9203\(97\)90049-3](https://doi.org/10.1016/s0035-9203(97)90049-3) PMID: 9463683
8. Repetto SA, Ruybal P, Batalla E, López C, Fridman V, Sierra M, et al. Strongyloidiasis outside endemic areas: long-term parasitological and clinical follow-up after ivermectin treatment. Clin Infect Dis. 2018;cix1069–cix. <https://doi.org/10.1093/cid/cix1069> PMID: 29360939
9. Buonfrate D, Requena-Mendez A, Angheben A, Munoz J, Gobbi F, Van Den Ende J, et al. Severe strongyloidiasis: a systematic review of case reports. BMC Infect Dis. 2013; 13:78. <https://doi.org/10.1186/1471-2334-13-78> PMID: 23394259
10. Luvira V, Watthanakulpanich D, Pittisuttithum P. Management of *Strongyloides stercoralis*: a puzzling parasite. Int Health. 2014; 6(4):273–81. <https://doi.org/10.1093/inthealth/ihu058> PMID: 25173343
11. Tefé-Silva C R, Machado E, Faccioli LH, Ramos S. Hyperinfection syndrome in strongyloidiasis. Curr Top Trop Med 2012.
12. Schär F, Hattendorf J, Khieu V, Muth S, Char MC, Marti HP, et al. *Strongyloides stercoralis* larvae excretion patterns before and after treatment. Parasitology. 2014; 141(7):892–7. <https://doi.org/10.1017/s0031182013002345>
13. Henriquez-Camacho C, Gotuzzo E, Echevarria J, White AC Jr, Terashima A, Samalvides F, et al. Ivermectin versus albendazole or thiabendazole for *Strongyloides stercoralis* infection. Cochrane Database Syst Rev. 2016(1):Cd007745. <https://doi.org/10.1002/14651858.CD007745.pub3> PMID: 26778150
14. Marti H, Haji HJ, Savioli L, Chwaya HM, Mgeni AF, Ameir JS, et al. A comparative trial of a single-dose ivermectin versus three days of albendazole for treatment of *Strongyloides stercoralis* and other soil-transmitted helminth infections in children. Am J Trop Med Hyg. 1996; 55(5):477–81.
15. Palmeirim MS, Hurlimann E, Knopp S, Speich B, Belizario V Jr., Joseph SA, et al. Efficacy and safety of co-administered ivermectin plus albendazole for treating soil-transmitted helminths: a systematic review, meta-analysis and individual patient data analysis. PLoS Negl Trop Dis. 2018; 12(4):e0006458. <https://doi.org/10.1371/journal.pntd.0006458> PMID: 29702653

16. Barda B, Sayasone S, Phongluxa K, Xayavong S, Keoduangsy K, Odermatt P, et al. Efficacy of moxidectin versus ivermectin against *Strongyloides stercoralis* infections: a randomized controlled non-inferiority trial. *Clin Infect Dis*. 2017. <https://doi.org/10.1093/cid/cix278>
17. Hürlimann E, Hofmann D, Keiser J. Ivermectin and moxidectin against soil-transmitted helminth infections. *Trends Parasitol*. 2023. <https://doi.org/10.1016/j.pt.2023.01.009> PMID: 36804383
18. Koga K, Kasuya S, Khamboonruang C, Sukhavat K, Ieda M, Takatsuka N, et al. A modified agar plate method for detection of *Strongyloides stercoralis*. *Am J Trop Med Hyg*. 1991; 45(4):518–21.
19. Buonfrate D, Requena-Mendez A, Angheben A, Cinquini M, Cruciani M, Fittipaldo A, et al. Accuracy of molecular biology techniques for the diagnosis of *Strongyloides stercoralis* infection-A systematic review and meta-analysis. *PLoS Negl Trop Dis*. 2018; 12(2):e0006229. <https://doi.org/10.1371/journal.pntd.0006229> PMID: 29425193
20. Cools P, van Lieshout L, Koelewijn R, Addiss D, Ajampur SSR, Ayana M, et al. First international external quality assessment scheme of nucleic acid amplification tests for the detection of Schistosoma and soil-transmitted helminths, including Strongyloides: A pilot study. *PLoS Negl Trop Dis*. 2020; 14(6):e0008231. <https://doi.org/10.1371/journal.pntd.0008231>
21. Tamarozzi F, Longoni SS, Mazzi C, Pettene S, Montresor A, Mahanty S, et al. Diagnostic accuracy of a novel enzyme-linked immunoassay for the detection of IgG and IgG4 against *Strongyloides stercoralis* based on the recombinant antigens NIE/SslR. *Parasit Vectors*. 2021; 14(1):412. <https://doi.org/10.1186/s13071-021-04916-x> PMID: 34407876
22. Kalantari N, Chehrazi M, Ghaffari S, Gorgani-Firouzjaee T. Serological assays for the diagnosis of *Strongyloides stercoralis* infection: a systematic review and meta-analysis of diagnostic test accuracy. *Trans R Soc Trop Med Hyg*. 2020. <https://doi.org/10.1093/trstmh/trz135> PMID: 32052848
23. Bisoffi Z, Buonfrate D, Sequi M, Mejia R, Cimino RO, Krolewiecki AJ, et al. Diagnostic accuracy of five serologic tests for *Strongyloides stercoralis* infection. *PLoS Negl Trop Dis*. 2014; 8(1):e2640. <https://doi.org/10.1371/journal.pntd.0002640> PMID: 24427320
24. Buonfrate D, Sequi M, Mejia R, Cimino RO, Krolewiecki AJ, Albonico M, et al. Accuracy of five serologic tests for the follow up of *Strongyloides stercoralis* infection. *PLoS Negl Trop Dis*. 2015; 9(2):e0003491. <https://doi.org/10.1371/journal.pntd.0003491> PMID: 25668740
25. Lindo JF, Atkins NS, Lee MG, Robinson RD, Bundy DA. Parasite-specific serum IgG following successful treatment of endemic strongyloidiasis using ivermectin. *Trans R Soc Trop Med Hyg*. 1996; 90(6):702–3. [https://doi.org/10.1016/s0035-9203\(96\)90444-7](https://doi.org/10.1016/s0035-9203(96)90444-7) PMID: 9015524
26. Vargas P, Krolewiecki AJ, Echazú A, Juárez M, Cajal P, Gil JF, et al. Serologic monitoring of public health interventions against *Strongyloides stercoralis*. *Am J Trop Med Hyg*. 2017; 97(1):166–72. <https://doi.org/10.4269/ajtmh.16-0857>
27. Nageswaran C, Craig PS, Devaney E. Coproantigen detection in rats experimentally infected with *Strongyloides ratti*. *Parasitology*. 1994; 108 (Pt 3):335–42.
28. Sykes AM, McCarthy JS. A coproantigen diagnostic test for *Strongyloides* infection. *PLoS Negl Trop Dis*. 2011; 5(2):e955. <https://doi.org/10.1371/journal.pntd.0000955> PMID: 21347447
29. Goncalves AL, Silva CV, Ueta MT, Costa-Cruz JM. A new faecal antigen detection system for *Strongyloides venezuelensis* diagnosis in immunosuppressed rats. *Exp Parasitol*. 2010; 125(4):338–41. <https://doi.org/10.1016/j.exppara.2010.02.015>
30. Chaves LA, Goncalves AL, Paula FM, Silva NM, Silva CV, Costa-Cruz JM, et al. Comparison of parasitological, immunological and molecular methods for evaluation of fecal samples of immunosuppressed rats experimentally infected with *Strongyloides venezuelensis*. *Parasitology*. 2015; 142(14):1715–21. <https://doi.org/10.1017/S0031182015001298> PMID: 26442899
31. El-Badry AA. ELISA-based coproantigen in human strongyloidiasis: a diagnostic method correlating with worm burden. *J Egypt Soc Parasitol*. 2009; 39(3):757–68. PMID: 20120743
32. Balachandra D, Ahmad H, Arifin N, Noordin R. Direct detection of *Strongyloides* infection via molecular and antigen detection methods. *Eur J Clin Microbiol Infect Dis*. 2021; 40(1):27–37. <https://doi.org/10.1007/s10096-020-03949-x> PMID: 32729057
33. Rappuoli R. Reverse vaccinology. *Curr Opin Microbiol*. 2000; 3(5):445–50. [https://doi.org/10.1016/s1369-5274\(00\)00119-3](https://doi.org/10.1016/s1369-5274(00)00119-3) PMID: 11050440
34. Hunt VL, Tsai IJ, Coghlan A, Reid AJ, Holroyd N, Foth BJ, et al. The genomic basis of parasitism in the *Strongyloides* clade of nematodes. *Nat Genet*. 2016. <https://doi.org/10.1038/ng.3495>
35. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite – a comprehensive resource for helminth genomics. *Mol Biochem Parasitol*. 2017; 215:2–10. <https://doi.org/10.1016/j.molbiopara.2016.11.005> PMID: 27899279

36. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, et al. WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Research*. 2015; 44(D1):D774–D80. <https://doi.org/10.1093/nar/gkv1217> PMID: 26578572
37. Stoltzfus JD, Minot S, Berriman M, Nolan TJ, Lok JB. RNAseq analysis of the parasitic nematode *Strongyloides stercoralis* reveals divergent regulation of canonical dauer pathways. *PLoS Negl Trop Dis*. 2012; 6(10):e1854. <https://doi.org/10.1371/journal.pntd.0001854> PMID: 23145190
38. Hunt VL, Tsai IJ, Selkirk ME, Viney M. The genome of *Strongyloides* spp. gives insights into protein families with a putative role in nematode parasitism. *Parasitology*. 2016:1–16. <https://doi.org/10.1017/s0031182016001554>
39. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011; 12:323. <https://doi.org/10.1186/1471-2105-12-323> PMID: 21816040
40. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10(3):R25. <https://doi.org/10.1186/gb-2009-10-3-r25> PMID: 19261174
41. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22(22):4673–80. <https://doi.org/10.1093/nar/22.22.4673> PMID: 7984417
42. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016; 44(W1):W242–5. <https://doi.org/10.1093/nar/gkw290> PMID: 27095192
43. Boratyn GM, Schaffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. *Biol Direct*. 2012; 7:12. <https://doi.org/10.1186/1745-6150-7-12> PMID: 22510480
44. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res*. 2009; 37(Database issue):D211–5. <https://doi.org/10.1093/nar/gkn785> PMID: 18940856
45. Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, et al. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res*. 2010; 38(Database issue):D161–6. <https://doi.org/10.1093/nar/gkp885> PMID: 19858104
46. Sigrist CJ, de Castro E, Cerutti L, Cuče BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. *Nucleic Acids Res*. 2013; 41(Database issue):D344–7. <https://doi.org/10.1093/nar/gks1067> PMID: 23161676
47. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016; 44(D1):D457–62. <https://doi.org/10.1093/nar/gkv1070> PMID: 26476454
48. Soblik H, Younis AE, Mitreva M, Renard BY, Kirchner M, Geisinger F, et al. Life cycle stage-resolved proteomic analysis of the excretome/secretome from *Strongyloides ratti*—identification of stage-specific proteases. *Mol Cell Proteomics*. 2011; 10(12):M111.010157. <https://doi.org/https://doi.org/10.1074/mcp.M111.010157>
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
50. Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. 2007–2015. Available from: <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.
51. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011; 8(10):785–6. <https://doi.org/10.1038/nmeth.1701> PMID: 21959131
52. Eamudomkarn C, Sithithaworn P, Sithithaworn J, Kaewkes S, Sripa B, Itoh M. Comparative evaluation of *Strongyloides ratti* and *S. stercoralis* larval antigen for diagnosis of strongyloidiasis in an endemic area of opisthorchiasis. *Parasitol Res*. 2015. <https://doi.org/10.1007/s00436-015-4458-3>
53. Sousa JE, Carvalho EF, Levenhagen MA, Faria LS, Goncalves-Pires Mdo R, Costa-Cruz JM. Serological cross-reactivity between *Strongyloides venezuelensis* and *Syphacia muris* in Wistar rats (*Rattus norvegicus*). *Parasitol Int*. 2016; 65(2):137–45. <https://doi.org/10.1016/j.parint.2015.11.005> PMID: 26601618
54. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res*. 2006; 2:2. <https://doi.org/10.1186/1745-7580-2-2> PMID: 16635264

55. Saha S, Raghava GPS, editors. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. Artificial Immune Systems; 2004; Berlin, Heidelberg: Springer Berlin Heidelberg.
56. Jespersen MC, Peters B, Nielsen M, Marcotili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017; 45(W1):W24–W9. <https://doi.org/10.1093/nar/gkx346> PMID: 28472356
57. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015; 10(6):845–58. <https://doi.org/10.1038/nprot.2015.053> PMID: 25950237
58. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25(13):1605–12. <https://doi.org/10.1002/jcc.20084> PMID: 15264254
59. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002:310–22. PMID: 11928486
60. The UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021; 49(D1):D480–D9. <https://doi.org/10.1093/nar/gkaa1100> PMID: 33237286
61. Khieu V, Hattendorf J, Schar F, Marti H, Char MC, Muth S, et al. *Strongyloides stercoralis* infection and re-infection in a cohort of children in Cambodia. Parasitol Int. 2014; 63(5):708–12. <https://doi.org/10.1016/j.parint.2014.06.001>
62. Forrer A, Khieu V, Schindler C, Schar F, Marti H, Char MC, et al. Ivermectin treatment and sanitation effectively reduce *Strongyloides stercoralis* infection risk in rural communities in Cambodia. PLoS Negl Trop Dis. 2016; 10(8):e0004909. <https://doi.org/10.1371/journal.pntd.0004909> PMID: 27548286
63. Lagatie O, Verheyen A, Van Hoof K, Lauwers D, Odiere MR, Vlaminck J, et al. Detection of *Ascaris lumbricoides* infection by ABA-1 coproantigen ELISA. PLoS Negl Trop Dis. 2020; 14(10):e0008807. <https://doi.org/10.1371/journal.pntd.0008807> PMID: 33057357
64. Sarkari B, Khabisi SA. Immunodiagnosis of human fascioliasis: an update of concepts and performances of the serological assays. J Clin Diagn Res. 2017; 11(6):Oe05–oe10. <https://doi.org/10.7860/JCDR/2017/26066.10086> PMID: 28764235
65. Morel N, Lassabe G, Elola S, Bondad M, Herrera S, Mari C, et al. A monoclonal antibody-based copro-ELISA kit for canine echinococcosis to support the PAHO effort for hydatid disease control in South America. PLoS Negl Trop Dis. 2013; 7(1):e1967. <https://doi.org/10.1371/journal.pntd.0001967> PMID: 23326610
66. Teimoori S, Arimatsu Y, Laha T, Kaewkes S, Sereerak P, Sripa M, et al. Chicken IgY-based coproantigen capture ELISA for diagnosis of human opisthorchiasis. Parasitol Int. 2016. <https://doi.org/10.1016/j.parint.2015.10.011> PMID: 27140305
67. Elsemore DA, Geng J, Cote J, Hanna R, Lucio-Forster A, Bowman DD. Enzyme-linked immunosorbent assays for coproantigen detection of *Ancylostoma caninum* and *Toxocara canis* in dogs and *Toxocara cati* in cats. J Vet Diagn Invest. 2017; 29(5):645–53. <https://doi.org/10.1177/1040638717706098>
68. Haque R, Kress K, Wood S, Jackson TF, Lyerly D, Wilkins T, et al. Diagnosis of pathogenic *Entamoeba histolytica* infection using a stool ELISA based on monoclonal antibodies to the galactose-specific adhesin. J Infect Dis. 1993; 167(1):247–9.
69. Culma MF. *Strongyloides stercoralis* proteome: A reverse approach to the identification of potential immunogenic candidates. Microb Pathog. 2021; 152:104545. <https://doi.org/10.1016/j.micpath.2020.104545>
70. Dishnica K, Piubelli C, Manfredi M, Kondaveeti RT, Longoni SS, Degani M, et al. Novel insights into the somatic proteome of *Strongyloides stercoralis* infective third-stage larvae. Parasit Vectors. 2023; 16(1):45. <https://doi.org/10.1186/s13071-023-05675-7> PMID: 36721249
71. Fonseca PDM, Corral MA, Cosenza-Contreras M, Meisel D, Melo GB, Antunes MMS, et al. Shotgun proteomics of *Strongyloides venezuelensis* infective third stage larvae: Insights into host-parasite interaction and novel targets for diagnostics. Mol Biochem Parasitol. 2019; 235:111249. <https://doi.org/10.1016/j.molbiopara.2019.111249>
72. Mulvenna J, Hamilton B, Nagaraj SH, Smyth D, Loukas A, Gorman JJ. Proteomics analysis of the excretory/secretory component of the blood-feeding stage of the hookworm, *Ancylostoma caninum*. Mol Cell Proteomics. 2009; 8(1):109–21. <https://doi.org/10.1074/mcp.M800206-MCP200>
73. Logan J, Pearson MS, Manda SS, Choi YJ, Field M, Eichenberger RM, et al. Comprehensive analysis of the secreted proteome of adult *Necator americanus* hookworms. PLoS Negl Trop Dis. 2020; 14(5):e0008237. <https://doi.org/10.1371/journal.pntd.0008237>

74. Morante T, Shepherd C, Constantinoiu C, Loukas A, Sotillo J. Revisiting the Ancylostoma Caninum Secretome Provides New Information on Hookworm-Host Interactions. *Proteomics*. 2017; 17(23–24). <https://doi.org/10.1002/ptm.201700186> PMID: 29052354
75. Tritten L, Ballesteros C, Beech R, Geary TG, Moreno Y. Mining nematode protein secretomes to explain lifestyle and host specificity. *PLoS Negl Trop Dis*. 2021; 15(9):e0009828. <https://doi.org/10.1371/journal.pntd.0009828> PMID: 34587193
76. Lin B, Zhuo K, Chen S, Hu L, Sun L, Wang X, et al. A novel nematode effector suppresses plant immunity by activating host reactive oxygen species-scavenging system. *New Phytol*. 2016; 209(3):1159–73. <https://doi.org/10.1111/nph.13701> PMID: 26484653
77. Shi H, Huang X, Chen X, Yang Y, Wu F, Yao C, et al. Haemonchus contortus Transthyretin-Like Protein TTR-31 Plays Roles in Post-Embryonic Larval Development and Potentially Apoptosis of Germ Cells. *Front Cell Dev Biol*. 2021; 9:753667. <https://doi.org/10.3389/fcell.2021.753667> PMID: 34805162
78. Treitz C, Cassidy L, Höckendorf A, Leippe M, Tholey A. Quantitative proteome analysis of Caenorhabditis elegans upon exposure to nematocidal Bacillus thuringiensis. *J Proteomics*. 2015; 113:337–50. <https://doi.org/10.1016/j.jpro.2014.09.027> PMID: 25452134
79. Selkirk ME, Lazari O, Matthews JB. Functional genomics of nematode acetylcholinesterases. *Parasitology*. 2005; 131(S1):S3–S18. <https://doi.org/10.1017/S0031182005008206> PMID: 16569291
80. Silman I. The multiple biological roles of the cholinesterases. *Prog Biophys Mol Biol*. 2021; 162:41–56. <https://doi.org/10.1016/j.pbiomolbio.2020.12.001> PMID: 33307019
81. Rebello KM, McKerrow JH, Mota EM, O'Donoghue AJ, Neves-Ferreira AGC. Activity profiling of peptidases in *Angiostrongylus costaricensis* first-stage larvae and adult worms. *PLoS Negl Trop Dis*. 2018; 12(10):e0006923. <https://doi.org/10.1371/journal.pntd.0006923> PMID: 30379807
82. Fajtová P, Štefanić S, Hradilek M, Dvořák J, Vondrášek J, Jílková A, et al. Prolyl oligopeptidase from the blood fluke *Schistosoma mansoni*: from functional analysis to anti-schistosomal inhibitors. *PLoS Negl Trop Dis*. 2015; 9(6):e0003827. <https://doi.org/10.1371/journal.pntd.0003827>
83. Abduhaleem N, Mamuda A, Mustapha T, Majid RA, Lung LTT, Unyah NZ. Evaluation of monoclonal antibody (IgG2bMAb) for detection of coproantigen from experimentally infected rats with *Strongyloides ratti*. *Annu Res Rev Biol*. 2019; 32(2):1–9. <https://doi.org/https://doi.org/10.9734/arrb/2019/v32i230082>
84. Goodswen SJ, Kennedy PJ, Ellis JT. Vacceed: a high-throughput in silico vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology. *Bioinformatics*. 2014; 30(16):2381–3. <https://doi.org/10.1093/bioinformatics/btu300> PMID: 24790156
85. Russi RC, Bourdin E, Garcia MI, Veaute CMI. In silico prediction of T- and B-cell epitopes in PmpD: first step towards the design of a *Chlamydia trachomatis* vaccine. *Biomed J*. 2018; 41(2):109–17. <https://doi.org/10.1016/j.bj.2018.04.007>
86. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021; 373(6557):871–6. <https://doi.org/10.1126/science.abj8754> PMID: 34282049
87. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021; 596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
88. Corstjens PL, De Dood CJ, Kornelis D, Fat EM, Wilson RA, Kariuki TM, et al. Tools for diagnosis, monitoring and screening of *Schistosoma* infections utilizing lateral-flow based assays and upconverting phosphor labels. *Parasitology*. 2014; 141(14):1841–55. <https://doi.org/10.1017/s0031182014000626>
89. Lawn SD. Point-of-care detection of lipoarabinomannan (LAM) in urine for diagnosis of HIV-associated tuberculosis: a state of the art review. *BMC Infect Dis*. 2012; 12:103. <https://doi.org/10.1186/1471-2334-12-103> PMID: 22536883
90. Ines EJ, Silva ML, Souza JN, Teixeira MC, Soares NM. The role of glycosylated epitopes in the serodiagnosis of *Strongyloides stercoralis* infection. *Diagn Microbiol Infect Dis*. 2013; 76(1):31–5. <https://doi.org/10.1016/j.diagmicrobio.2013.01.016>
91. Hokke CH, van Diepen A. Helminth glycomics—glycan repertoires and host-parasite interactions. *Mol Biochem Parasitol*. 2017; 215:47–57. <https://doi.org/10.1016/j.molbiopara.2016.12.001> PMID: 27939587
92. Lok JB, Shao H, Massey HC, Li X. Transgenesis in *Strongyloides* and related parasitic nematodes: historical perspectives, current functional genomic applications and progress towards gene disruption and editing. *Parasitology*. 2016:1–16. <https://doi.org/10.1017/s0031182016000391>

93. Gao H, Xiao D, Song L, Zhang W, Shen S, Yin X, et al. Assessment of the diagnostic efficacy of enolase as an indication of active infection of *Schistosoma japonicum*. *Parasitol Res*. 2016; 115(1):151–64. <https://doi.org/10.1007/s00436-015-4730-6>
94. Marcilla A, Trelis M, Cortes A, Sotillo J, Cantalapedra F, Minguez MT, et al. Extracellular vesicles from parasitic helminths contain specific excretory/secretory proteins and are internalized in intestinal host cells. *PLoS One*. 2012; 7(9):e45974. <https://doi.org/10.1371/journal.pone.0045974> PMID: 23029346
95. Eberle R, Brattig NW, Trusch M, Schluter H, Achukwi MD, Eisenbarth A, et al. Isolation, identification and functional profile of excretory-secretory peptides from *Onchocerca ochengi*. *Acta Trop*. 2015; 142:156–66. <https://doi.org/10.1016/j.actatropica.2014.11.015>
96. Liu Z, Kelleher A, Tabb S, Wei J, Pollet J, Hotez PJ, et al. Identification, characterization, and structure of Tm16 from *Trichuris muris*. *J Parasitol Res*. 2017; 2017:4342789. <https://doi.org/10.1155/2017/4342789>
97. Rodpai R, Intapan PM, Thanchomnang T, Sanpool O, Janwan P, Laummaunwai P, et al. Identification of antigenic proteins in *Strongyloides stercoralis* by proteomic analysis. *Parasitol Res*. 2017. <https://doi.org/10.1007/s00436-017-5443-9>
98. Marcilla A, Sotillo J, Perez-Garcia A, Igual-Adell R, Valero ML, Sanchez-Pino MM, et al. Proteomic analysis of *Strongyloides stercoralis* L3 larvae. *Parasitology*. 2010; 137(10):1577–83. <https://doi.org/10.1017/s0031182010000314>
99. Wang T, Van Steendam K, Dhaenens M, Vlaminck J, Deforce D, Jex AR, et al. Proteomic analysis of the excretory-secretory products from larval stages of *Ascaris suum* reveals high abundance of glycosyl hydrolases. *PLoS Negl Trop Dis*. 2013; 7(10):e2467. <https://doi.org/10.1371/journal.pntd.0002467> PMID: 24098821
100. Kingston IB. Nematode collagen genes. *Parasitol Today*. 1991; 7(1):11–5. [https://doi.org/10.1016/0169-4758\(91\)90077-2](https://doi.org/10.1016/0169-4758(91)90077-2) PMID: 15463376
101. Blaxter ML, Page AP, Rudin W, Maizels RM. Nematode surface coats: Actively evading immunity. *Parasitol Today*. 1992; 8(7):243–7. [https://doi.org/10.1016/0169-4758\(92\)90126-m](https://doi.org/10.1016/0169-4758(92)90126-m) PMID: 15463630
102. Maeda Y, Palomares-Rius JE, Hino A, Afrin T, Mondal SI, Nakatake A, et al. Secretome analysis of *Strongyloides venezuelensis* parasitic stages reveals that soluble and insoluble proteins are involved in its parasitism. *Parasit Vectors*. 2019; 12(1):21. <https://doi.org/10.1186/s13071-018-3266-x> PMID: 30626426
103. Abraham D, Hess JA, Mejia R, Nolan TJ, Lok JB, Lustigman S, et al. Immunization with the recombinant antigen Ss-IR induces protective immunity to infection with *Strongyloides stercoralis* in mice. *Vaccine*. 2011; 29(45):8134–40. <https://doi.org/10.1016/j.vaccine.2011.08.030>
104. Kerepesi LA, Keiser PB, Nolan TJ, Schad GA, Abraham D, Nutman TB. DNA immunization with Na⁺-K⁺ ATPase (Sseat-6) induces protective immunity to larval *Strongyloides stercoralis* in mice. *Infect Immun*. 2005; 73(4):2298–305. <https://doi.org/10.1128/IAI.73.4.2298-2305.2005> PMID: 15784574
105. Ramanathan R, Burbelo PD, Groot S, Iadarola MJ, Neva FA, Nutman TB. A luciferase immunoprecipitation systems assay enhances the sensitivity and specificity of diagnosis of *Strongyloides stercoralis* infection. *J Infect Dis*. 2008; 198(3):444–51. <https://doi.org/10.1086/589718>
106. Ravi V, Ramachandran S, Thompson RW, Andersen JF, Neva FA. Characterization of a recombinant immunodiagnostic antigen (NIE) from *Strongyloides stercoralis* L3-stage larvae. *Mol Biochem Parasitol*. 2002; 125(1–2):73–81. [https://doi.org/10.1016/s0166-6851\(02\)00214-1](https://doi.org/10.1016/s0166-6851(02)00214-1) PMID: 12467975
107. Balachandra D, Rahumatullah A, Lim TS, Mustafa FH, Ahmad H, Anuar NS, et al. A new antigen detection ELISA for the diagnosis of *Strongyloides* infection. *Acta Trop*. 2021; 221:105986. <https://doi.org/10.1016/j.actatropica.2021.105986> PMID: 34058161
108. Qiang S, Bin Z, Shu-hua X, Zheng F, Hotez P, Hawdon JM. Variation between ASP-1 molecules from *Ancylostoma caninum* in China and the United States. *J Parasitol*. 2000; 86(1):181–5. [https://doi.org/10.1645/0022-3395\(2000\)086\[0181:vbamfa\]2.0.co;2](https://doi.org/10.1645/0022-3395(2000)086[0181:vbamfa]2.0.co;2)
109. Bhattacharyya T, Boelaert M, Miles MA. Comparison of visceral leishmaniasis diagnostic antigens in African and Asian *Leishmania donovani* reveals extensive diversity and region-specific polymorphisms. *PLoS Negl Trop Dis*. 2013; 7(2):e2057. <https://doi.org/10.1371/journal.pntd.0002057> PMID: 23469296
110. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic diversity and protective efficacy of the RTS,S/AS01 malaria vaccine. *N Engl J Med*. 2015; 373(21):2025–37. <https://doi.org/10.1056/NEJMoa1505819> PMID: 26488565
111. Jaleta TG, Zhou S, Bemm FM, Schär F, Khieu V, Muth S, et al. Different but overlapping populations of *Strongyloides stercoralis* in dogs and humans—Dogs as a possible source for zoonotic strongyloidiasis. *PLoS Negl Trop Dis*. 2017; 11(8):e0005752. <https://doi.org/10.1371/journal.pntd.0005752> PMID: 28793306