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

Tegwen Marlais[1*], Jack Bickford-Smith[1], Carlos Talavera-López[2], Hai Le[1], Fatima Chowdhury[1], Michael A. Miles[1]

[1]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

[2]Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany

*Corresponding Author

E-mail: Tegwen.marlais@lshtm.ac.uk

Abstract

Human infection with the intestinal nematode *Strongyloides stercoralis* is persistent unless effectively treated, and potentially fatal in immunosuppressed individuals. Epidemiological data are lacking 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 analysed open access 'omics' data sets and used online predictors to identify *S. stercoralis* proteins that are likely 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 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].

S. 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

69 competence due to immunosuppression, such as corticosteroid treatment of co-morbidities, or
70 HTLV-1 co-infection, very large numbers of larvae may undergo this autoinfective cycle, causing
71 hyperinfection or disseminated strongyloidiasis, both with a high fatality rate [9-11]. This
72 autoinfective lifecycle means that infections will fully re-establish if worms are not completely
73 cleared from the host during treatment. Diagnosis of strongyloidiasis and validation of cure after
74 treatment are therefore imperative.

75 Treatment with the first-line drug ivermectin has a reported efficacy of between 57% and 100%.
76 However, accurate determination of cure depends on follow-up and the diagnostic method used [8,
77 10, 12]. Albendazole and mebendazole, used to treat infection with other STH are less effective or
78 ineffective against *Strongyloides* [13-15]. Moxidectin has shown effectiveness against *S. stercoralis*,
79 that is equivalent to ivermectin, in early trials and continues to be developed [16, 17].

80 Diagnosis of strongyloidiasis can be made by microscopy using agar plate culture [18], Baermann
81 funnel or spontaneous sedimentation, all of which isolate larvae from fresh stool [19, 20]. qPCR on
82 extracted stool DNA is used in research and highly resourced laboratories, and improves sensitivity
83 over microscopy [21-24]. However, sensitivity may be reduced by inadequate DNA extraction [25],
84 low stool volume used in testing (as little as 0.1 g compared to 1 gram or more used for culture or
85 larval concentration methods), and irregular larval excretion [12]. Serology, detecting antibodies to
86 either whole worm or recombinant antigens NIE and SsIR, has sensitivity of 70-98% and high
87 specificity [26] but can only distinguish cure months to years after effective treatment [27-29].
88 Therefore, there is a need for a rapid diagnostic test (RDT) that can be used for screening as well as
89 for confirmation of cure. An antigen-based assay would fulfil this need, because it may achieve high
90 sensitivity and specificity and test for active infection. Pilots of such assays have shown proof of
91 principle under research conditions using antibodies against *Strongyloides ratti* and *Strongyloides*
92 *venezuelensis* somatic or excretory/secretory (E/S) antigens [30-32]. However, identification of

specific *S. stercoralis* 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 antigens that may be less abundant or difficult to purify *in vitro*. 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, primarily Wormbase ParaSite, a resource for parasitic worm genomics curated by the Sanger Institute and the EMBL European Bioinformatics Institute. Full details of data sources are listed in Supporting Information file S1.

S. 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: presence in infected stool; specificity to *Strongyloides* and/or *S. stercoralis*; antigenicity, to facilitate raising sensitive antibodies. Datasets and computational analyses used to make this selection are detailed in Figure 1.

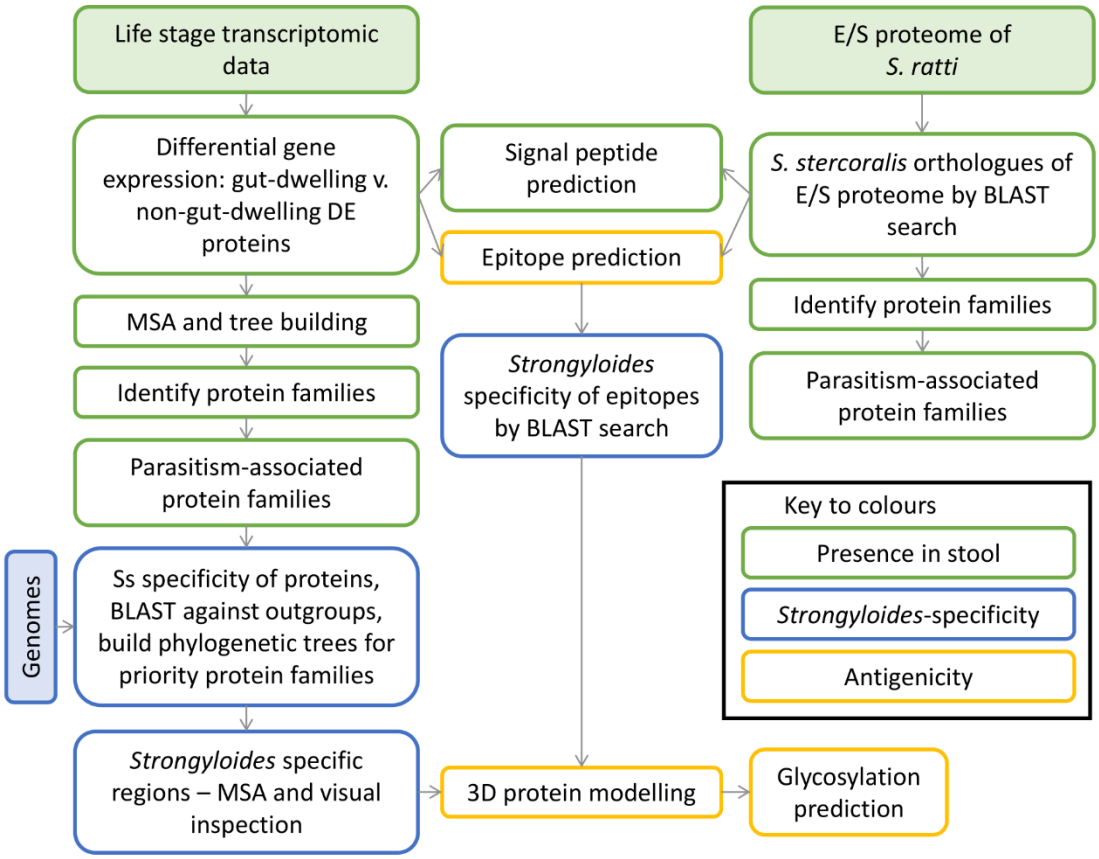


Figure 1. Schematic of the method for identifying coproantigens of *Strongyloides stercoralis* using 'omic' data and computational analyses.

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.

Evidence for the presence of candidate coproantigens in stool

Analysis of the *Strongyloides* genomes by Hunt et al. (2016) [35] was used to identify protein families associated with parasitism. Five parasitism-associated protein families ('priority protein families') were then 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.

Transcriptomics

To reveal candidate coproantigens, we used transcriptomic data from Stoltzfus *et al.* (2012) [36], which analysed transcriptomes of 7 life stages of *S. stercoralis* (Figure 2). 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 (Figure 2).

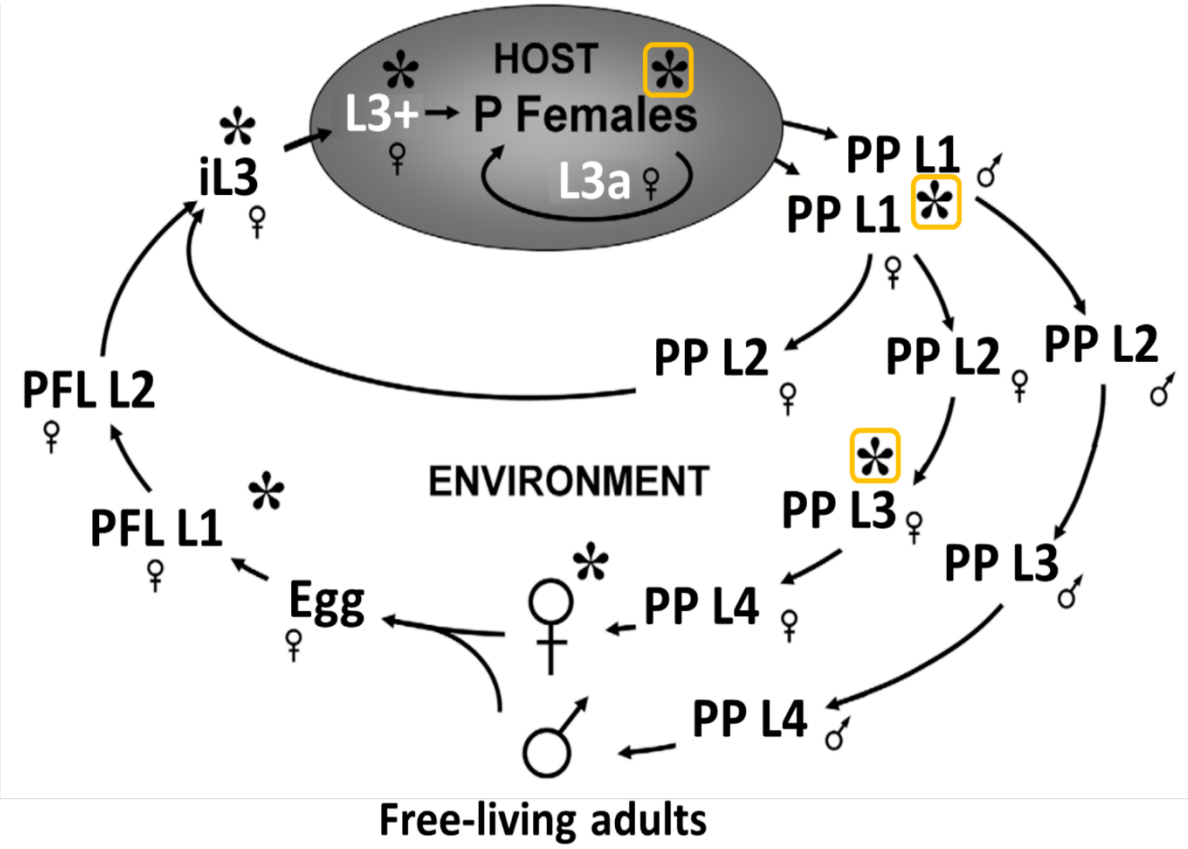


Figure 2. *S. stercoralis* life stages.

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) [36] under a CC BY license. Accession numbers are given for NCBI SRA for the triplicate reads in Supporting Information S1 Table.

We calculated relative abundance of transcripts using RSEM [37] and bowtie2 [38] 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 ebseq in RSEM. We selected genes that were differentially expressed with 100% confidence, in either direction, between gut-dwelling and non-gut-dwelling life stages.

Differentially expressed protein family identification

ClustalW [39] was used to perform multiple alignment of the differentially expressed (DE) genes and to produce a phylogenetic tree, which was annotated with iTOL [40]. 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.

DE proteins were firstly identified by submitting amino acid sequences of selected clusters on the tree to three domain-finding tools: Delta BLAST [41], InterPro [42] and ExPASy Prosite [43, 44] and a consensus of all three used to obtain probable protein identity. Secondly, DE protein sequences were submitted to BlastKOALA [45]. BlastKOALA protein identities were considered alongside previous consensus or used alone if there was no identity from three domain-finding tools.

Excretory/secretory proteomics

Source of proteomic data

Soblik *et al.* (2011) [46] submitted excretory/secretory (E/S) material of *S. ratti* parasitic females to mass spectrometry and identified the constituent proteins. In their presentation of the *Strongyloides 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 [47].

***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. *S. stercoralis* hits, in the form of accession 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 [48] 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]. Separately, differential gene expression data from analysis of the Stoltzfus *et al.* (2012) [36] dataset were extracted for the E/S orthologues that occurred in both datasets.

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 [49].

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.

A custom blast+ database was created from the genome-derived proteomes of selected outgroup species (Table 1). The outgroups were selected to represent parasitic and non-parasitic nematodes, as well as trematodes, cestodes and human. Human protein/coding sequence (CDS) data were

195 provided by the Human Genome Project at the Wellcome Trust Sanger Institute and was obtained

196 from Ensembl. Accession numbers for all data are available in Supporting Information file S1.

197

198

Table 1. Outgroup species used for phylogenetic comparison to *S. stercoralis*.

Species		Rationale for selection
1	*Ancylostoma duodenale*	Hookworm, prevalent nematode of humans
2	*Ascaris lumbricoides*	Prevalent nematode of humans
3	*Caenorhabditis elegans*	Free-living nematode and model organism
4	*Clonorchis sinensis*	Trematode of humans
5	*Enterobius vermicularis*	Prevalent nematode of humans
6	*Homo sapiens*	Human, present in all samples
7	*Necator americanus*	Hookworm, prevalent nematode of humans
8	*Onchocerca volvulus*	Non-gut nematode of humans
9	*Parastrongyloides trichosuri*	Relative of *S. stercoralis,* facultative parasite of opossums
10	*Strongyloides ratti*	Close relative of *S. stercoralis*, serological cross-reactivity with *S. stercoralis* [50]
11	*Strongyloides venezuelensis*	Close relative of *S. stercoralis*
12	*Strongyloides papillosus*	Close relative of *S. stercoralis*
13	*Syphacia muris*	Nematode of mice and rats, serological cross-reactivity with *Strongyloides* in experimental animal infection [51]
14	*Taenia solium*	Cestode of humans
15	*Trichuris trichiura*	Prevalent nematode of humans

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. 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 (S2 Table). 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 (S2 Table). 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 with iTOL [40] 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 submitted to BLASTP search against the NCBI nr, and Nematoda (taxid: 6231) databases, as relevant, to validate their specificity.

Antigenic potential of candidate coproantigens

Epitope prediction

BepiPred 1.0 [52] and bcepred [53] were used to predict epitopes within the DE proteins and E/S proteome orthologues. 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 [54] for certain candidate antigens. 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.* Sequences were searched using BLASTP against the NCBI non-redundant (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 [55] for 3D structure modelling against known crystal structures, using the intensive mode. UCSF Chimera [56] 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 [57] to account for the potential of a glycan to obscure protein antigen regions, or conversely to contribute to antigenicity. 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.

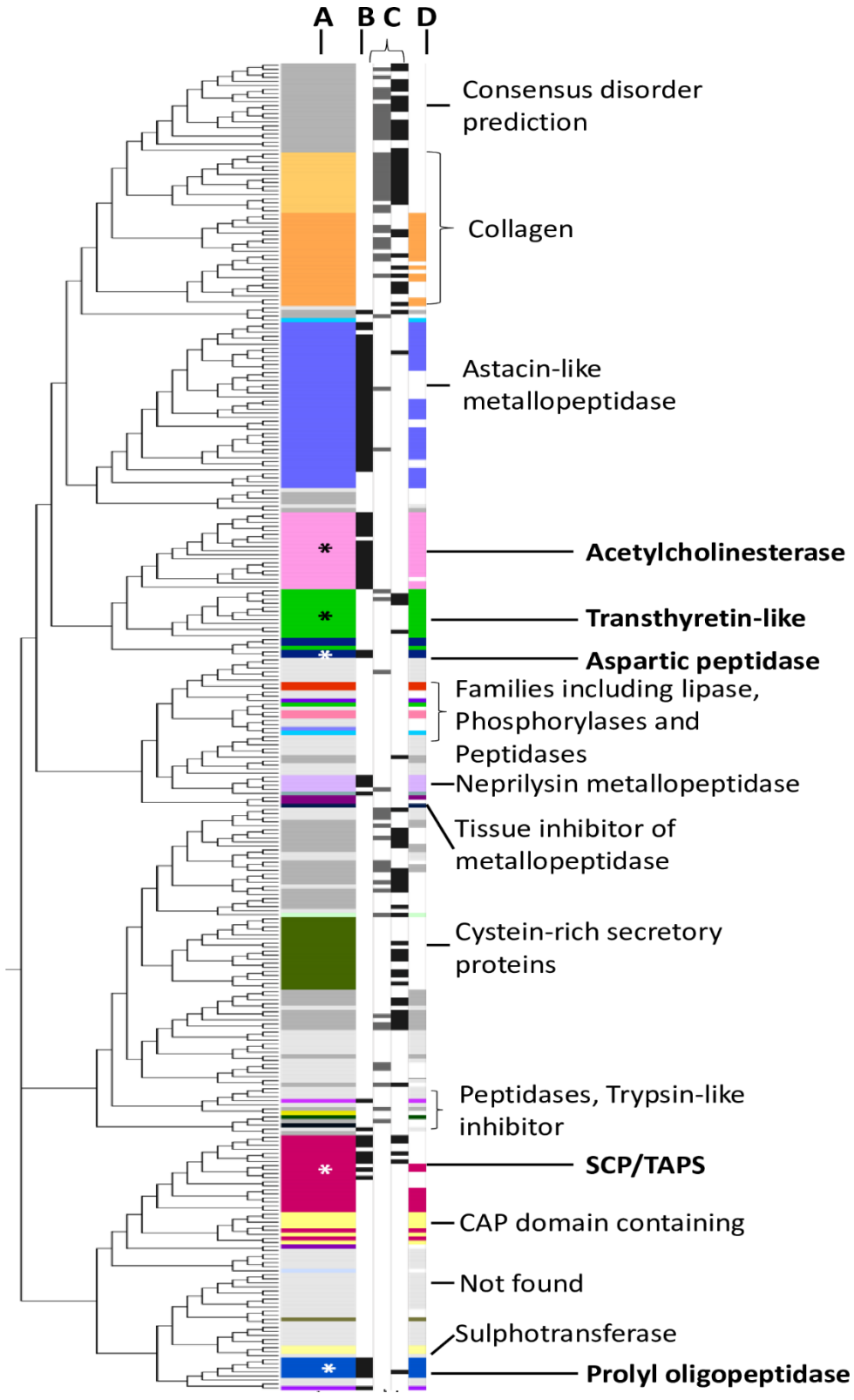
Results

Priority protein families

Hunt *et al.* (2016) [35] identified seven protein families associated with *Strongyloides* parasitism. For our coproantigen search, we focused on 5 of these, namely: sperm coating protein/Tpx-1/Ag5/PR-1/Sc7 (SCP/TAPS), transthyretin-like (TTL), acetylcholinesterase (AChE), prolyl oligopeptidase (POP) and aspartic peptidases.

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) [36] we found 328 which we were 100% confident of differential expression between gut-dwelling and non-gut-dwelling life stages according to our groupings (Figure 3). Of these, 198 (60.4%) contained a signal peptide. Of the 328 DE genes, 203 were expressed more in gut-dwelling life stages than non-gut, including some or all of the proteins in the 5 priority protein families analysed here. These were therefore the focus of our coproantigen search (Figure 3 and S3 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 (Figure 3).



277

278 **Figure 3. *S. stercoralis* proteins differentially expressed between gut-dwelling and non-gut-**
279 **dwelling life stages.**

280 The total range of DE proteins is included for comparison. Proteins are grouped by similarity and

families are identified. Any protein family with less than 2 BLAST hits present not labelled. Colours refer to features or families; the 5 priority protein families analysed (SCP/TAPS; TTL; AChE; aspartic peptidases, and POP) are labelled with an asterisk. Labels indicate layers representing: A, the 328 *S. stercoralis* DE proteins; B, proteins orthologous to *S. ratti* E/S proteins; C, proteins containing epitopes predicted by BcePred (left) and BepiPred (right); D, the 203 proteins that were DE in gut-dwelling life stages.

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. Original *S. ratti* E/S proteins were given as 582 accession numbers, however two were found to have 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* (S4 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.

To investigate gene expression of the E/S orthologues, we extracted data for the 1,057 *S. stercoralis* proteins from RSEM analysis of the entire 13,098 active *S. stercoralis* genes. 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] (S5 file). Relevant proteins were assigned to the five 'priority protein families' (Figure 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 (Figure 3 and S3 file). Within the 78 proteins, BepiPred predicted 125 epitopes, and within 62 proteins bcepred predicted 108 epitopes (S6 file). Fifty six epitopes contained overlap or identity between the two prediction tools (S6 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.

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 (S6 file).

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 (Figure 4).

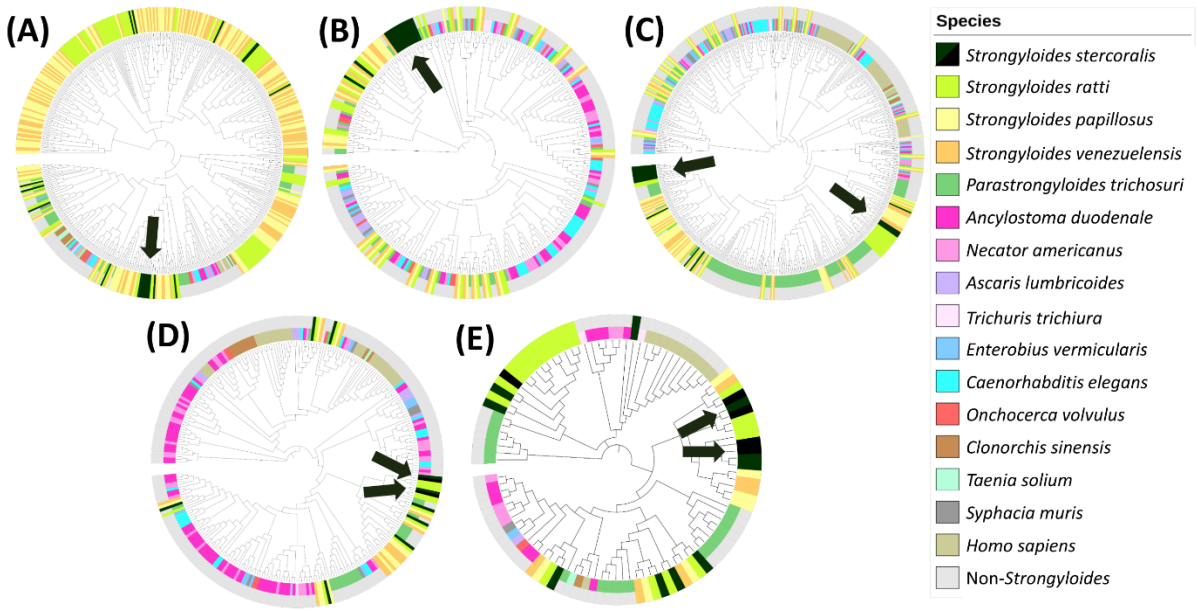


Figure 4. Phylogenetic alignment of *S.* stercoralis high priority family proteins with other *S. stercoralis* proteins and outgroups.

Proteins of the five high 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.

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. 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

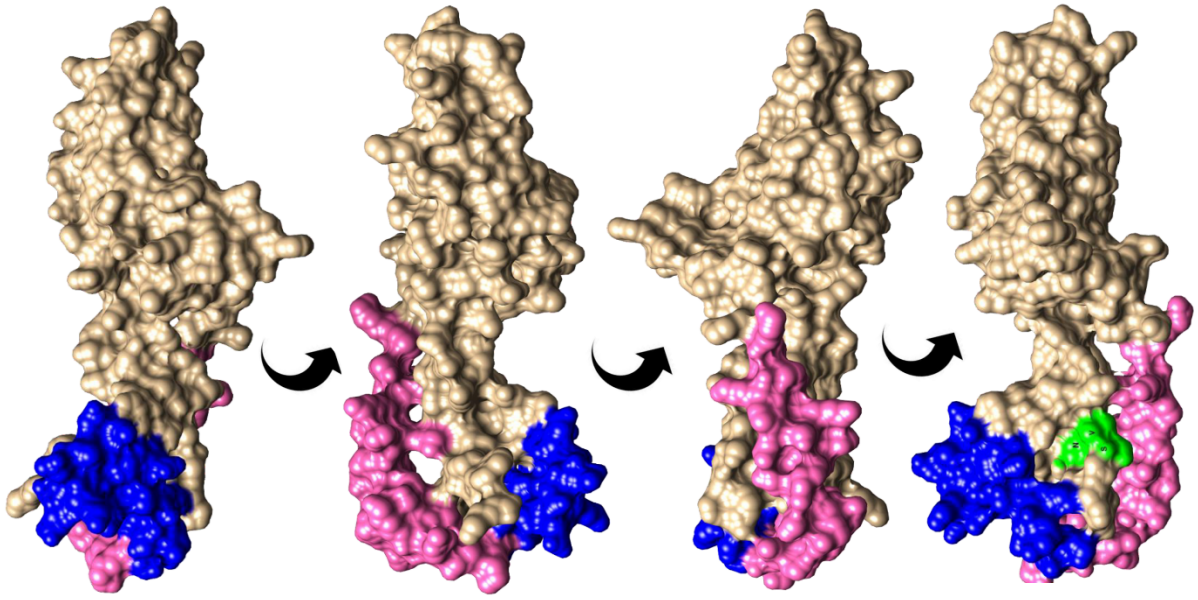
Twenty one differentially expressed proteins were identified as definite (n = 19) or possible (n = 2) members of the SCP/TAPS protein family (Figure 3 and S3 file). The two 'possible' SCP/TAPS proteins clustered separately among CAP-domain proteins (Figure 3) and were therefore not included as definite 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 (Figure 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 Figure 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 2).

Transthyretin-like coproantigen candidates

Fourteen (14) TTL proteins were differentially expressed between gut and non-gut-dwelling life stages with 100% confidence, all of which were expressed more in gut-dwelling life stages, almost exclusively in parasitic female worms (Figure 3 and S3 file). Twelve of these proteins grouped as a cluster while the other 2 were more similar to other protein families or features (Figure 3). When aligned against homologous proteins from the outgroup species, 11 TTL proteins clustered together (Figure 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,

368 the search revealed that the amino acid sequence VTCDGKPL in protein SSTP_0000485800 is
369 conserved across many nematode genera and should therefore be avoided in any candidate antigen.
370 Two of the TTL proteins did contain one or more regions of *Strongyloides* species or genus
371 specificity, giving a total of 6 TTL candidate coproantigens (Table 2).

372 In addition to high *Strongyloides*-specificity across its whole sequence of 177aa, TTL protein
373 SSTP_0000700800 also contained predicted epitopes. To investigate the position of these, the
374 protein was 3D modelled to a template generated from TTR-52 protein of *C. elegans* (PDB accession:
375 3UAF) (Figure 5). This model aligned to 89 residues (aa 2-90; 50% of the sequence) with 99.9%
376 confidence. This protein contained a predicted glycosylation site at position N165 and, although not
377 predicted to have a signal peptide, it was indicated as an 'extracellular or secreted' protein in
378 UniProt [58] and is a known E/S protein family, therefore it is more likely to be glycosylated.



379

380 **Figure 5. A transthyretin-like (TTL) protein of *S. stercoralis* in rotated view showing surface-**
381 **exposed epitope regions and N-linked glycosylation sites**
382 Accession number SSTP_0000700800. Blue, BepiPred predicted epitope aa's 124-143; pink, bcepred
383 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.

Acetylcholinesterase (AChE) coproantigen candidates

Nineteen AChE proteins were in the DE dataset, of which 18 were expressed higher in gut-dwelling stages, particularly in parasitic female worms (Figure 3). The other AChE protein was highly expressed in infectious and tissue-migrating larvae (S3 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 (Figure 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 2).

From the larger cluster in Figure 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 (Figure 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 (Figure 6).

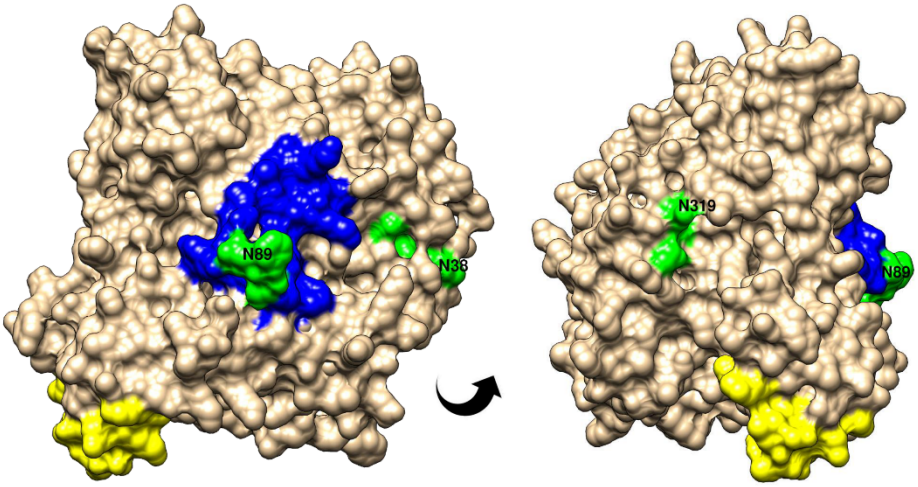


Figure 6. *S. stercoralis* acetylcholinesterase in rotated views showing surface-exposed epitope regions and N-linked glycosylation sites. Accession number SSTP_0000509400. Predicted features indicated in colour: blue, BepiPred 2.0 predicted epitope at residues 85-103; and yellow at 390-411; green, predicted N-linked glycosylation sites at N38 (NVT), N89 (NFS) and N319 (NLT). The site at N89 is within the blue epitope sequence.

Aspartic peptidase coproantigen candidates

Only 4 aspartic peptidases appeared in the DE proteins, all of which were expressed more 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 (Figure 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 2). 3D modelling of this protein could not be achieved with any reliability and was thus not used for epitope selection. There were no predicted epitopes in either of the DE aspartic peptidases.

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 (Figure 4E). Four of the DE proteins (SSTP_0000289100, 1108800, 1108500, 1019400) had higher species-specificity, clustering with other *S. stercoralis* homologues (Figure 4E, arrowed). One of these 4 proteins, SSTP_0000289100 contained two Bepipred epitopes and was very highly expressed in parasitic females, as well as being in the E/S orthologues (Figure 3 and S3 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 (SSTP_0001108800) (Table 2). 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 2 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 file S7.

Table 2. 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 file S7. PF: parasitic female, E/S: excretory/secretory.

Protein family	Gene	Residues	Length	Peptide is specific to *Strongyloides*	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
Transthryetin-like (TTL)	SSTP_0000700800	1-177	177	Yes	Yes- in PF	No	No	Yes
	SSTP_0001222000	22-50 51-86 65-159 87-116 125-160	29 36 95 30 36	Yes Yes Yes Yes Yes	Yes- in PF	Yes	No	Yes
Acetylcholinesterase (AChE)	SSTP_0000274700	17-43 63-104 187-207 243-266 268-306 308-328 372-451 551-607	27 42 21 24 39 21 80 57	Yes Yes Yes Yes Yes Yes Yes Yes	Yes- in PF	Yes	Yes	No
	SSTP_0000509400	85-103 390-411	19 22	Yes Yes	Yes- in PF	Yes	Yes	Yes
Aspartic peptidase	SSTP_0000164500	107-140	34	Yes	Yes- in PF	Yes	Yes	No
Prolyl oligopeptidase (POP)	SSTP_0001108800	221-266 384-421 422-507 741-779	46 38 86 39	Yes Yes Yes Yes	Yes- in PF	No	Yes	No

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 [27, 59]. Incomplete cure and reinfection post-treatment may occur [8, 60, 61]. 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* [62], *Fasciola* [63], *Echinococcus* [64], *Strongyloides ratti* [32], *S. venezuelensi*s [30], *Opisthorcis* [65], *Toxocara* [66] and *Entamoeba histolytica* [67], 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 [68]. 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 [36]. Seven protein families have been identified by other studies as expanded in the genomes of parasitic nematodes, and upregulated in parasitic life stages [35]. Studies of the *Ancylostoma* hookworm E/S proteome and *S. venezuelensis* somatic larval proteome detected some of the same families (SCP/TAPS, proteases and TTL) indicating likely presence in stool [69, 70]. 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, 46].

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 [50].

Priority protein families and species specificity

Phylogenetic trees of *S. stercoralis* DE proteins and their homologues from a selection of outgroups 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 [35]. 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, Ancylostoma ceylanicum* and *Necator americanus,* in which they are numerous. Multiple of these SCP/TAPS proteins 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.

Transthyretin-like proteins

Transthyretin-like proteins are a nematode-specific protein family and are known to be in E/S material of many species [71]. 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 [68]. The exact function of TTL proteins remains unknown in animal parasitic nematodes, but evidence from plant parasites suggests a role in subduing the immune response to favour worm survival [72].

Acetylcholinesterase

Secreted AChE has a possible role in enabling certain parasitic helminths to evade host expulsion mechanisms from mucosal surfaces [35, 73]. 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 (S5 file). Secreted AChE differs from the neuromuscular protein in structure, gene family and substrate, being less specific to acetylcholine [73]. 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 a role in haemoglobin digestion in other parasitic helminths, therefore secretion may be linked to feeding within the host gut epithelium where the adult worm resides [74, 75]. 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 DE 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 [76]. 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 [46]. 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 [77].

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 [78]. 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 [54].

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 [79]. 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 [55], 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 [80, 81].

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 [82], and LAM of *Mycobacterium tuberculosis* [83]. They have also been implicated in lysate seroantigen of *S. stercoralis* [84]. In helminths, glycan structures may not only be species-specific, but also life-stage specific [85]. 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 [85]. 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 [86]. As an alternative, glycans could be excluded altogether by

577 synthesising peptides or expressing recombinant proteins in bacteria, thus focusing the antigen
578 search purely on proteins, as we have done here.

579 Limitations and future work

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

593 Another limitation is prediction of antigens that are present in the stool. In our pipeline, a protein
594 being excreted or secreted by a gut-dwelling stage of the parasite has been used as a proxy for
595 presence in the stool. This rests on the assumption that the protein passes through the gut, is
596 present in the stool in sufficient quantity and with antigenic properties unchanged by digestion or
597 denaturation. As *S. stercoralis* establishes and matures in the small intestine, E/S proteins will not be
598 subject to the denaturation and enzymatic degradation in the stomach but may be affected by
599 enzymes in the small intestine. However, post-parasitic larvae, which hatch from eggs laid in the
600 intestine, migrate through the small and large intestine, and can be found in faeces, so any E/S
601 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.

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 potential proteins for coproantigen detection assays include: enolase, common to *Schistosoma japonicum* [87], *Echinostoma* and *Fasciola* [88], *Onchocerca* [89] and *Trichuris* [90], as well as in among *S. stercoralis* E/S orthologous proteins reported here and with constitutive high expression across all life stages (S5 file); protein 14-3-3 from E/S and somatic extract of *Strongyloides* [91, 92], *Ascaris* [93]*, Schistosoma* and *Ancylostoma* [70]. In addition, collagen, which forms some 80% of the outer cuticle of the nematode and was prominent among the DE proteins [94]. As an antigen, collagen would have to be carefully analysed for species specificity and the influence of glycosylation on its availability to capture antibodies [95]. A summary of E/S proteomic studies was presented by Ditgen *et al.* (2014) [96], and analysis of *S. venezuelensis* somatic proteins was made by Fonseca *et al*. (2019) [69] which may inform further antigen searches. Vaccine candidates for *S. stercoralis*: sodium potassium ATPase (SsEAT), tropomyosin, and a galectin (LEC-5) [97, 98] may also generate effective antibodies for antigen capture, if the relevant antigens are detectable in stool.

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 [99]. This has impacted on diagnosis and vaccination for other parasitic infections [100, 101]. Diversity can be readily investigated by amplifying and sequencing the genes of interest from a wide geographic range of samples. 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 [102].

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.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Conceived and designed the study: MAM, TM, CTL. Carried out the study: TM, CTL, JB-S, HL, FC. Wrote the manuscript: TM. Reviewed and edited the manuscript: MAM, JB-S, CTL.

Acknowledgements

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

Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

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).

References

1. Schär F, Trostdorf 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

670 2. Forrer A, Khieu V, Schar F, Vounatsou P, Chammartin F, Marti H, et al. *Strongyloides*
671 *stercoralis* and hookworm co-infection: spatial distribution and determinants in Preah Vihear
672 Province, Cambodia. Parasit Vectors. 2018; 11(1):33. https://doi.org/10.1186/s13071-017-2604-8
673 3. Buonfrate D, Baldissera M, Abrescia F, Bassetti M, Caramaschi G, Giobbia M, et al.
674 Epidemiology of *Strongyloides stercoralis* in northern Italy: results of a multicentre case-control
675 study, February 2013 to July 2014. Euro Surveill. 2016; 21(31). https://doi.org/10.2807/1560-
676 7917.es.2016.21.31.30310
677 4. Becker SL, Sieto B, Silue KD, Adjossan L, Kone S, Hatz C, et al. Diagnosis, clinical features, and
678 self-reported morbidity of *Strongyloides stercoralis* and hookworm infection in a co-endemic setting.
679 PLoS Negl Trop Dis. 2011; 5(8):e1292. https://doi.org/10.1371/journal.pntd.0001292
680 5. Minciullo PL, Cascio A, Gangemi S. Association between urticaria and nematode infections.
681 Allergy Asthma Proc. 2018; 39(2):86-95. https://doi.org/10.2500/aap.2018.38.4104
682 6. Abrescia FF, Falda A, Caramaschi G, Scalzini A, Gobbi F, Angheben A, et al. Reemergence of
683 strongyloidiasis, northern Italy. Emerg Infect Dis. 2009; 15(9):1531-3.
684 https://doi.org/10.3201/eid1509.090191
685 7. Atkins NS, Lindo JF, Lee MG, Conway DJ, Bailey JW, Robinson RD, et al. Humoral responses in
686 human strongyloidiasis: correlations with infection chronicity. Trans R Soc Trop Med Hyg. 1997;
687 91(5):609-13.
688 8. Repetto SA, Ruybal P, Batalla E, López C, Fridman V, Sierra M, et al. Strongyloidiasis outside
689 endemic areas: long-term parasitological and clinical follow-up after ivermectin treatment. Clin
690 Infect Dis. 2018:cix1069-cix. https://doi.org/10.1093/cid/cix1069
691 9. Buonfrate D, Requena-Mendez A, Angheben A, Munoz J, Gobbi F, Van Den Ende J, et al.
692 Severe strongyloidiasis: a systematic review of case reports. BMC Infect Dis. 2013; 13:78.
693 https://doi.org/10.1186/1471-2334-13-78
694 10. Luvira V, Watthanakulpanich D, Pittisuttithum P. Management of *Strongyloides stercoralis*: a
695 puzzling parasite. Int Health. 2014; 6(4):273-81. https://doi.org/10.1093/inthealth/ihu058

11. Tefé-Silva C, R. Machado E, Faccioli LH, Ramos S. Hyperinfection syndrome in strongyloidiasis. Curr Top Trop Med2012.

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

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

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. Buchter V, Hofmann D, Häberli C, Keiser J. Characterization of moxidectin against *Strongyloides ratti*: *in vitro* and i*n vivo* activity and pharmacokinetics in the rat model. ACS Infect Dis. 2021; 7(5):1069-76. https://doi.org/10.1021/acsinfecdis.0c00435

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. Lopez M, Morales ML, Konana M, Hoyer P, Pineda-Reyes R, White AC, Jr., et al. Kato-Katz and Lumbreras rapid sedimentation test to evaluate helminth prevalence in the setting of a school-

based deworming program. Pathog Glob Health. 2016; 110(3):130-4. https://doi.org/10.1080/20477724.2016.1187361

20. Tello R, Terashima A, Marcos LA, Machicado J, Canales M, Gotuzzo E. Highly effective and inexpensive parasitological technique for diagnosis of intestinal parasites in developing countries: Spontaneous sedimentation technique in tube. Int J Parasitol. 2012; 16(6):e414-e6. https://doi.org/10.1016/j.ijid.2011.12.017

21. Becker SL, Piraisoody N, Kramme S, Marti H, Silue KD, Panning M, et al. Real-time PCR for detection of *Strongyloides stercoralis* in human stool samples from Cote d'Ivoire: diagnostic accuracy, inter-laboratory comparison and patterns of hookworm co-infection. Acta Trop. 2015; 150:210-7. https://doi.org/10.1016/j.actatropica.2015.07.019

22. Incani RN, Ferrer E, Hoek D, Ramak R, Roelfsema J, Mughini-Gras L, et al. Diagnosis of intestinal parasites in a rural community of Venezuela: advantages and disadvantages of using microscopy or RT-PCR. Acta Trop. 2016. https://doi.org/10.1016/j.actatropica.2016.12.014

23. Meurs L, Polderman AM, Vinkeles Melchers NV, Brienen EA, Verweij JJ, Groosjohan B, et al. Diagnosing polyparasitism in a high-prevalence setting in Beira, Mozambique: detection of intestinal parasites in fecal samples by microscopy and real-time PCR. PLoS Negl Trop Dis. 2017; 11(1):e0005310. https://doi.org/10.1371/journal.pntd.0005310

24. Repetto SA, Ruybal P, Solana ME, Lopez C, Berini CA, Alba Soto CD, et al. Comparison between PCR and larvae visualization methods for diagnosis of *Strongyloides stercoralis* out of endemic area: a proposed algorithm. Acta Trop. 2016; 157:169-77. https://doi.org/10.1016/j.actatropica.2016.02.004

25. Repetto SA, Alba Soto CD, Cazorla SI, Tayeldin ML, Cuello S, Lasala MB, et al. An improved DNA isolation technique for PCR detection of *Strongyloides stercoralis* in stool samples. Acta Trop. 2013; 126(2):110-4. https://doi.org/10.1016/j.actatropica.2013.02.003

745 26. Buonfrate D, Formenti F, Perandin F, Bisoffi Z. Novel approaches to the diagnosis of
746 *Strongyloides stercoralis* infection. Clin Microbiol Infect. 2015.
747 https://doi.org/10.1016/j.cmi.2015.04.001
748 27. Buonfrate D, Sequi M, Mejia R, Cimino RO, Krolewiecki AJ, Albonico M, et al. Accuracy of five
749 serologic tests for the follow up of *Strongyloides stercoralis* infection. PLoS Negl Trop Dis. 2015;
750 9(2):e0003491. https://doi.org/10.1371/journal.pntd.0003491
751 28. Lindo JF, Atkins NS, Lee MG, Robinson RD, Bundy DA. Parasite-specific serum IgG following
752 successful treatment of endemic strongyloidiasis using ivermectin. Trans R Soc Trop Med Hyg. 1996;
753 90(6):702-3.
754 29. Vargas P, Krolewiecki AJ, Echazú A, Juarez M, Cajal P, Gil JF, et al. Serologic monitoring of
755 public health interventions against *Strongyloides stercoralis*. Am J Trop Med Hyg. 2017; 97(1):166-
756 72. https://doi.org/10.4269/ajtmh.16-0857
757 30. Goncalves AL, Silva CV, Ueta MT, Costa-Cruz JM. A new faecal antigen detection system for
758 *Strongyloides venezuelensis* diagnosis in immunosuppressed rats. Exp Parasitol. 2010; 125(4):338-41.
759 https://doi.org/10.1016/j.exppara.2010.02.015
760 31. Nageswaran C, Craig PS, Devaney E. Coproantigen detection in rats experimentally infected
761 with *Strongyloides ratti*. Parasitology. 1994; 108 (Pt 3):335-42.
762 32. Sykes AM, McCarthy JS. A coproantigen diagnostic test for *Strongyloides* infection. PLoS Negl
763 Trop Dis. 2011; 5(2):e955. https://doi.org/10.1371/journal.pntd.0000955
764 33. Rappuoli R. Reverse vaccinology. Curr Opin Microbiol. 2000; 3(5):445-50.
765 34. Hunt VL, Tsai IJ, Coghlan A, Reid AJ, Holroyd N, Foth BJ, et al. The genomic basis of parasitism
766 in the *Strongyloides* clade of nematodes. Nat Genet. 2016. https://doi.org/10.1038/ng.3495
767 35. Hunt VL, Tsai IJ, Selkirk ME, Viney M. The genome of *Strongyloides* spp. gives insights into
768 protein families with a putative role in nematode parasitism. Parasitology. 2016:1-16.
769 https://doi.org/10.1017/s0031182016001554

770 36. Stoltzfus JD, Minot S, Berriman M, Nolan TJ, Lok JB. RNAseq analysis of the parasitic
771 nematode *Strongyloides stercoralis* reveals divergent regulation of canonical dauer pathways. PLoS
772 Negl Trop Dis. 2012; 6(10):e1854. https://doi.org/10.1371/journal.pntd.0001854
773 37. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without
774 a reference genome. BMC Bioinformatics. 2011; 12:323. https://doi.org/10.1186/1471-2105-12-323
775 38. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of
776 short DNA sequences to the human genome. Genome Biol. 2009; 10(3):R25.
777 https://doi.org/10.1186/gb-2009-10-3-r25
778 39. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive
779 multiple sequence alignment through sequence weighting, position-specific gap penalties and
780 weight matrix choice. Nucleic Acids Res. 1994; 22(22):4673-80.
781 40. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and
782 annotation of phylogenetic and other trees. Nucleic Acids Res. 2016; 44(W1):W242-5.
783 https://doi.org/10.1093/nar/gkw290
784 41. Boratyn GM, Schaffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced
785 lookup time accelerated BLAST. Biol Direct. 2012; 7:12. https://doi.org/10.1186/1745-6150-7-12
786 42. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the
787 integrative protein signature database. Nucleic Acids Res. 2009; 37(Database issue):D211-5.
788 https://doi.org/10.1093/nar/gkn785
789 43. Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, et al.
790 PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids
791 Res. 2010; 38(Database issue):D161-6. https://doi.org/10.1093/nar/gkp885
792 44. Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing
793 developments at PROSITE. Nucleic Acids Res. 2013; 41(Database issue):D344-7.
794 https://doi.org/10.1093/nar/gks1067

795 45. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource
796 for gene and protein annotation. Nucleic Acids Res. 2016; 44(D1):D457-62.
797 https://doi.org/10.1093/nar/gkv1070
798 46. Soblik H, Younis AE, Mitreva M, Renard BY, Kirchner M, Geisinger F, et al. Life cycle stage-
799 resolved proteomic analysis of the excretome/secretome from *Strongyloides ratti*—identification of
800 stage-specific proteases. Mol Cell Proteomics. 2011; 10(12):M111.010157.
801 https://doi.org/https://doi.org/10.1074/mcp.M111.010157
802 47. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
803 Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078-9.
804 https://doi.org/10.1093/bioinformatics/btp352
805 48. Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. 2007-2015.
806 Available from: http://bioinfogp.cnb.csic.es/tools/venny/index.html.
807 49. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides
808 from transmembrane regions. Nat Methods. 2011; 8(10):785-6.
809 https://doi.org/10.1038/nmeth.1701
810 50. Eamudomkarn C, Sithithaworn P, Sithithaworn J, Kaewkes S, Sripa B, Itoh M. Comparative
811 evaluation of *Strongyloides ratti* and *S. stercoralis* larval antigen for diagnosis of strongyloidiasis in an
812 endemic area of opisthorchiasis. Parasitol Res. 2015. https://doi.org/10.1007/s00436-015-4458-3
813 51. Sousa JE, Carvalho EF, Levenhagen MA, Faria LS, Goncalves-Pires Mdo R, Costa-Cruz JM.
814 Serological cross-reactivity between *Strongyloides venezuelensis* and *Syphacia muris* in Wistar rats
815 (*Rattus norvegicus*). Parasitol Int. 2016; 65(2):137-45. https://doi.org/10.1016/j.parint.2015.11.005
816 52. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes.
817 Immunome Res. 2006; 2:2. https://doi.org/10.1186/1745-7580-2-2
818 53. Saha S, Raghava GPS, editors. BcePred: prediction of continuous B-cell epitopes in antigenic
819 sequences using physico-chemical properties. Artificial Immune Systems; 2004; Berlin, Heidelberg:
820 Springer Berlin Heidelberg.

821 54. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-
822 cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017; 45(W1):W24-W9.
823 https://doi.org/10.1093/nar/gkx346
824 55. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein
825 modeling, prediction and analysis. Nat Protoc. 2015; 10(6):845-58.
826 https://doi.org/10.1038/nprot.2015.053
827 56. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF
828 Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;
829 25(13):1605-12. https://doi.org/10.1002/jcc.20084
830 57. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the
831 correlation to protein function. Pac Symp Biocomput. 2002:310-22.
832 58. The UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res.
833 2021; 49(D1):D480-D9. https://doi.org/10.1093/nar/gkaa1100
834 59. Bisoffi Z, Buonfrate D, Sequi M, Mejia R, Cimino RO, Krolewiecki AJ, et al. Diagnostic accuracy
835 of five serologic tests for *Strongyloides stercoralis* infection. PLoS Negl Trop Dis. 2014; 8(1):e2640.
836 https://doi.org/10.1371/journal.pntd.0002640
837 60. Khieu V, Hattendorf J, Schar F, Marti H, Char MC, Muth S, et al. *Strongyloides stercoralis*
838 infection and re-infection in a cohort of children in Cambodia. Parasitol Int. 2014; 63(5):708-12.
839 https://doi.org/10.1016/j.parint.2014.06.001
840 61. Forrer A, Khieu V, Schindler C, Schar F, Marti H, Char MC, et al. Ivermectin treatment and
841 sanitation effectively reduce *Strongyloides stercoralis* infection risk in rural communities in
842 Cambodia. PLoS Negl Trop Dis. 2016; 10(8):e0004909.
843 https://doi.org/10.1371/journal.pntd.0004909
844 62. Lagatie O, Verheyen A, Van Hoof K, Lauwers D, Odiere MR, Vlaminck J, et al. Detection of
845 *Ascaris lumbricoides* infection by ABA-1 coproantigen ELISA. PLoS Negl Trop Dis. 2020;
846 14(10):e0008807. https://doi.org/10.1371/journal.pntd.0008807

847 63. Sarkari B, Khabisi SA. Immunodiagnosis of human fascioliasis: an update of concepts and
848 performances of the serological assays. J Clin Diagn Res. 2017; 11(6):Oe05-oe10.
849 https://doi.org/10.7860/jcdr/2017/26066.10086
850 64. Morel N, Lassabe G, Elola S, Bondad M, Herrera S, Mari C, et al. A monoclonal antibody-
851 based copro-ELISA kit for canine echinococcosis to support the PAHO effort for hydatid disease
852 control in South America. PLoS Negl Trop Dis. 2013; 7(1):e1967.
853 https://doi.org/10.1371/journal.pntd.0001967
854 65. Teimoori S, Arimatsu Y, Laha T, Kaewkes S, Sereerak P, Sripa M, et al. Chicken IgY-based
855 coproantigen capture ELISA for diagnosis of human opisthorchiasis. Parasitol Int. 2016.
856 https://doi.org/10.1016/j.parint.2015.10.011
857 66. Elsemore DA, Geng J, Cote J, Hanna R, Lucio-Forster A, Bowman DD. Enzyme-linked
858 immunosorbent assays for coproantigen detection of *Ancylostoma caninum* and *Toxocara canis* in
859 dogs and *Toxocara cati* in cats. J Vet Diagn Invest. 2017; 29(5):645-53.
860 https://doi.org/10.1177/1040638717706098
861 67. Haque R, Kress K, Wood S, Jackson TF, Lyerly D, Wilkins T, et al. Diagnosis of pathogenic
862 *Entamoeba histolytica* infection using a stool ELISA based on monoclonal antibodies to the
863 galactose-specific adhesin. J Infect Dis. 1993; 167(1):247-9.
864 68. Culma MF. *Strongyloides stercoralis* proteome: A reverse approach to the identification of
865 potential immunogenic candidates. Microb Pathog. 2021; 152:104545.
866 https://doi.org/10.1016/j.micpath.2020.104545
867 69. Fonseca PDM, Corral MA, Cosenza-Contreras M, Meisel D, Melo GB, Antunes MMS, et al.
868 Shotgun proteomics of *Strongyloides venezuelensis* infective third stage larvae: Insights into host-
869 parasite interaction and novel targets for diagnostics. Mol Biochem Parasitol. 2019; 235:111249.
870 https://doi.org/10.1016/j.molbiopara.2019.111249

871 70. Mulvenna J, Hamilton B, Nagaraj SH, Smyth D, Loukas A, Gorman JJ. Proteomics analysis of
872 the excretory/secretory component of the blood-feeding stage of the hookworm, *Ancylostoma*
873 *caninum*. Mol Cell Proteomics. 2009; 8(1):109-21. https://doi.org/10.1074/mcp.M800206-MCP200
874 71. Saverwyns H, Visser A, Van Durme J, Power D, Morgado I, Kennedy MW, et al. Analysis of the
875 transthyretin-like (TTL) gene family in *Ostertagia ostertagi* – Comparison with other strongylid
876 nematodes and *Caenorhabditis elegans*. Int J Parasitol. 2008; 38(13):1545-56.
877 https://doi.org/https://doi.org/10.1016/j.ijpara.2008.04.004
878 72. Lin B, Zhuo K, Chen S, Hu L, Sun L, Wang X, et al. A novel nematode effector suppresses plant
879 immunity by activating host reactive oxygen species-scavenging system. New Phytol. 2016;
880 209(3):1159-73. https://doi.org/10.1111/nph.13701
881 73. Selkirk ME, Lazari O, Matthews JB. Functional genomics of nematode acetylcholinesterases.
882 Parasitology. 2005; 131(S1):S3-S18. https://doi.org/10.1017/S0031182005008206
883 74. Brindley PJ, Kalinna BH, Wong JYM, Bogitsh BJ, King LT, Smyth DJ, et al. Proteolysis of human
884 hemoglobin by schistosome cathepsin D. Mol Biochem Parasitol. 2001; 112(1):103-12.
885 https://doi.org/https://doi.org/10.1016/S0166-6851(00)00351-0
886 75. Rebello KM, McKerrow JH, Mota EM, O´Donoghue AJ, Neves-Ferreira AGC. Activity profiling
887 of peptidases in *Angiostrongylus costaricensis* first-stage larvae and adult worms. PLoS Negl Trop Dis.
888 2018; 12(10):e0006923. https://doi.org/10.1371/journal.pntd.0006923
889 76. Fajtová P, Štefanić S, Hradilek M, Dvořák J, Vondrášek J, Jílková A, et al. Prolyl oligopeptidase
890 from the blood fluke *Schistosoma mansoni*: from functional analysis to anti-schistosomal inhibitors.
891 PLoS Negl Trop Dis. 2015; 9(6):e0003827. https://doi.org/10.1371/journal.pntd.0003827
892 77. Abduhaleem N, Mamuda A, Mustapha T, Majid RA, Lung LTT, Unyah NZ. Evaluation of
893 monoclonal mntibody (IgG2bMAb) for detection of coproantigen from experimentally infected rats
894 with *Strongyloides ratti*. Annu Res Rev Biol. 2019; 32(2):1-9.
895 https://doi.org/https://doi.org/10.9734/arrb/2019/v32i230082

896 78. Goodswen SJ, Kennedy PJ, Ellis JT. Vacceed: a high-throughput in silico vaccine candidate
897 discovery pipeline for eukaryotic pathogens based on reverse vaccinology. Bioinformatics. 2014;
898 30(16):2381-3. https://doi.org/10.1093/bioinformatics/btu300
899 79. Russi RC, Bourdin E, Garcia MI, Veaute CMI. *In silico* prediction of T- and B-cell epitopes in
900 PmpD: first step towards to the design of a *Chlamydia trachomatis* vaccine. Biomed J. 2018;
901 41(2):109-17. https://doi.org/10.1016/j.bj.2018.04.007
902 80. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate
903 prediction of protein structures and interactions using a three-track neural network. Science. 2021;
904 373(6557):871-6. https://doi.org/10.1126/science.abj8754
905 81. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate
906 protein structure prediction with AlphaFold. Nature. 2021; 596(7873):583-9.
907 https://doi.org/10.1038/s41586-021-03819-2
908 82. Corstjens PL, De Dood CJ, Kornelis D, Fat EM, Wilson RA, Kariuki TM, et al. Tools for
909 diagnosis, monitoring and screening of *Schistosoma* infections utilizing lateral-flow based assays and
910 upconverting phosphor labels. Parasitology. 2014; 141(14):1841-55.
911 https://doi.org/10.1017/s0031182014000626
912 83. Lawn SD. Point-of-care detection of lipoarabinomannan (LAM) in urine for diagnosis of HIV-
913 associated tuberculosis: a state of the art review. BMC Infect Dis. 2012; 12:103.
914 https://doi.org/10.1186/1471-2334-12-103
915 84. Ines EJ, Silva ML, Souza JN, Teixeira MC, Soares NM. The role of glycosylated epitopes in the
916 serodiagnosis of *Strongyloides stercoralis* infection. Diagn Microbiol Infect Dis. 2013; 76(1):31-5.
917 https://doi.org/10.1016/j.diagmicrobio.2013.01.016
918 85. Hokke CH, van Diepen A. Helminth glycomics - glycan repertoires and host-parasite
919 interactions. Mol Biochem Parasitol. 2017; 215:47-57.
920 https://doi.org/10.1016/j.molbiopara.2016.12.001

921 86. Lok JB, Shao H, Massey HC, Li X. Transgenesis in *Strongyloides* and related parasitic
922 nematodes: historical perspectives, current functional genomic applications and progress towards
923 gene disruption and editing. Parasitology. 2016:1-16. https://doi.org/10.1017/s0031182016000391
924 87. Gao H, Xiao D, Song L, Zhang W, Shen S, Yin X, et al. Assessment of the diagnostic efficacy of
925 enolase as an indication of active infection of *Schistosoma japonicum*. Parasitol Res. 2016;
926 115(1):151-64. https://doi.org/10.1007/s00436-015-4730-6
927 88. Marcilla A, Trelis M, Cortes A, Sotillo J, Cantalapiedra F, Minguez MT, et al. Extracellular
928 vesicles from parasitic helminths contain specific excretory/secretory proteins and are internalized in
929 intestinal host cells. PLoS One. 2012; 7(9):e45974. https://doi.org/10.1371/journal.pone.0045974
930 89. Eberle R, Brattig NW, Trusch M, Schluter H, Achukwi MD, Eisenbarth A, et al. Isolation,
931 identification and functional profile of excretory-secretory peptides from *Onchocerca ochengi*. Acta
932 Trop. 2015; 142:156-66. https://doi.org/10.1016/j.actatropica.2014.11.015
933 90. Liu Z, Kelleher A, Tabb S, Wei J, Pollet J, Hotez PJ, et al. Identification, characterization, and
934 structure of Tm16 from *Trichuris muris*. J Parasitol Res. 2017; 2017:4342789.
935 https://doi.org/10.1155/2017/4342789
936 91. Rodpai R, Intapan PM, Thanchomnang T, Sanpool O, Janwan P, Laummaunwai P, et al.
937 Identification of antigenic proteins in *Strongyloides stercoralis* by proteomic analysis. Parasitol Res.
938 2017. https://doi.org/10.1007/s00436-017-5443-9
939 92. Marcilla A, Sotillo J, Perez-Garcia A, Igual-Adell R, Valero ML, Sanchez-Pino MM, et al.
940 Proteomic analysis of *Strongyloides stercoralis* L3 larvae. Parasitology. 2010; 137(10):1577-83.
941 https://doi.org/10.1017/s0031182010000314
942 93. Wang T, Van Steendam K, Dhaenens M, Vlaminck J, Deforce D, Jex AR, et al. Proteomic
943 analysis of the excretory-secretory products from larval stages of *Ascaris suum* reveals high
944 abundance of glycosyl hydrolases. PLoS Negl Trop Dis. 2013; 7(10):e2467.
945 https://doi.org/10.1371/journal.pntd.0002467
946 94. Kingston IB. Nematode collagen genes. Parasitol Today. 1991; 7(1):11-5.

947 95. Blaxter ML, Page AP, Rudin W, Maizels RM. Nematode surface coats: Actively evading
948 immunity. Parasitol Today. 1992; 8(7):243-7. https://doi.org/https://doi.org/10.1016/0169-
949 4758(92)90126-M
950 96. Ditgen D, Anandarajah EM, Meissner KA, Brattig N, Wrenger C, Liebau E. Harnessing the
951 helminth secretome for therapeutic immunomodulators. Biomed Res Int. 2014; 2014:964350.
952 https://doi.org/10.1155/2014/964350
953 97. Abraham D, Hess JA, Mejia R, Nolan TJ, Lok JB, Lustigman S, et al. Immunization with the
954 recombinant antigen Ss-IR induces protective immunity to infection with *Strongyloides stercoralis* in
955 mice. Vaccine. 2011; 29(45):8134-40. https://doi.org/10.1016/j.vaccine.2011.08.030
956 98. Kerepesi LA, Keiser PB, Nolan TJ, Schad GA, Abraham D, Nutman TB. DNA immunization with
957 Na+-K+ ATPase (Sseat-6) induces protective immunity to larval *Strongyloides stercoralis* in mice.
958 Infect Immun. 2005; 73(4):2298-305. https://doi.org/10.1128/iai.73.4.2298-2305.2005
959 99. Qiang S, Bin Z, Shu-hua X, Zheng F, Hotez P, Hawdon JM. Variation between ASP-1 molecules
960 from *Ancylostoma caninum* in China and the United States. J Parasitol. 2000; 86(1):181-5.
961 https://doi.org/10.1645/0022-3395(2000)086[0181:vbamfa]2.0.co;2
962 100. Bhattacharyya T, Boelaert M, Miles MA. Comparison of visceral leishmaniasis diagnostic
963 antigens in African and Asian *Leishmania donovani* reveals extensive diversity and region-specific
964 polymorphisms. PLoS Negl Trop Dis. 2013; 7(2):e2057.
965 https://doi.org/10.1371/journal.pntd.0002057
966 101. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic diversity and
967 protective efficacy of the RTS,S/AS01 malaria vaccine. N Engl J Med. 2015; 373(21):2025-37.
968 https://doi.org/10.1056/NEJMoa1505819
969 102. Jaleta TG, Zhou S, Bemm FM, Schär F, Khieu V, Muth S, et al. Different but overlapping
970 populations of *Strongyloides stercoralis* in dogs and humans—Dogs as a possible source for zoonotic
971 strongyloidiasis. PLoS Negl Trop Dis. 2017; 11(8):e0005752.
972 https://doi.org/10.1371/journal.pntd.0005752

Supporting information

S1 Table. Data sources used. CDS: coding sequence.

S2 Table: Number of outgroup protein hits for S. stercoralis protein families from the differentially expressed (DE) dataset.

S3 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 counbts where red is high and white is low.

S4 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

S5 file: *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

997 predicted epitopes according to two tools: BepiPred 1.0 and bcepred are indicated as well as
998 proteins predicted by SignalP to contain a signal peptide. Accession numbers starting ERR refer to
999 RNA-seq data in NCBI SRA and those starting SSTP are *S. stercoralis* gene transcripts which may be
1000 found in UniProt or WormBase ParaSite.
1001 Life-stage abbreviations are explained in the manuscript text.
1002 **S6 file: Predicted epitopes in *S. stercoralis* proteins using BepiPred and bcepred.** ES: Orthologues of
1003 *S. ratti* E/S proteome. DE: Differentially expressed proteins between gut-dwelling and non-gut-
1004 dwelling life stages.
1005 **S7 file: Candidate coproantigen sequences in fasta format.** FASTA headers are in the format:
1006 accession number, start aa, end aa.
1007

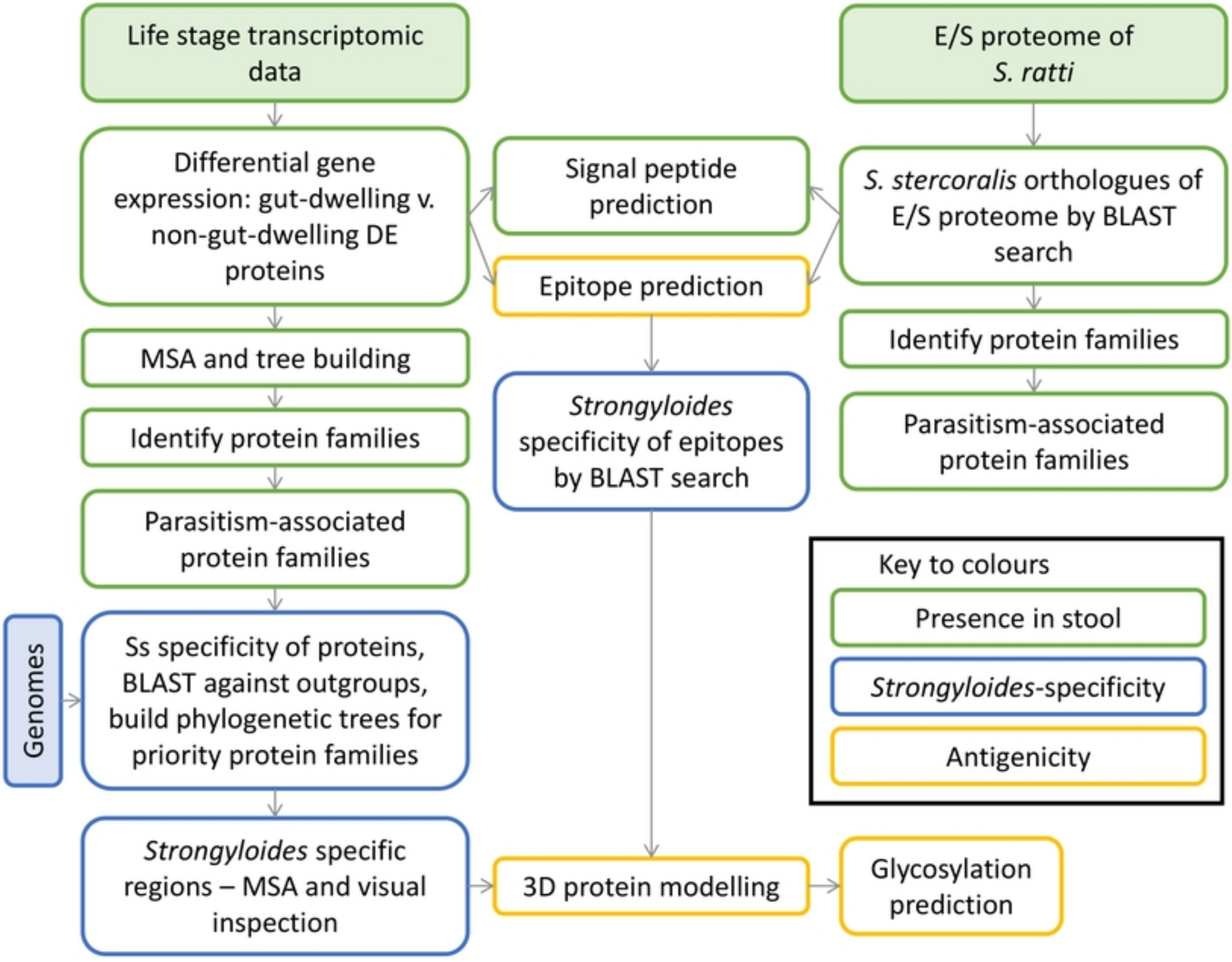

Life stage transcriptomic data
E/S proteome of S. ratti
Differential gene expression: gut-dwelling v. non-gut-dwelling DE proteins
Signal peptide prediction
S. stercoralis orthologues of E/S proteome by BLAST search
Epitope prediction
MSA and tree building
Identify protein families
Strongyloides specificity of epitopes by BLAST search
Identify protein families
Parasitism-associated protein families
Parasitism-associated protein families
Key to colours
Presence in stool
Strongyloides-specificity
Antigenicity
Genomes
Ss specificity of proteins, BLAST against outgroups, build phylogenetic trees for priority protein families
Strongyloides specific regions – MSA and visual inspection
3D protein modelling
Glycosylation prediction

Figure 1

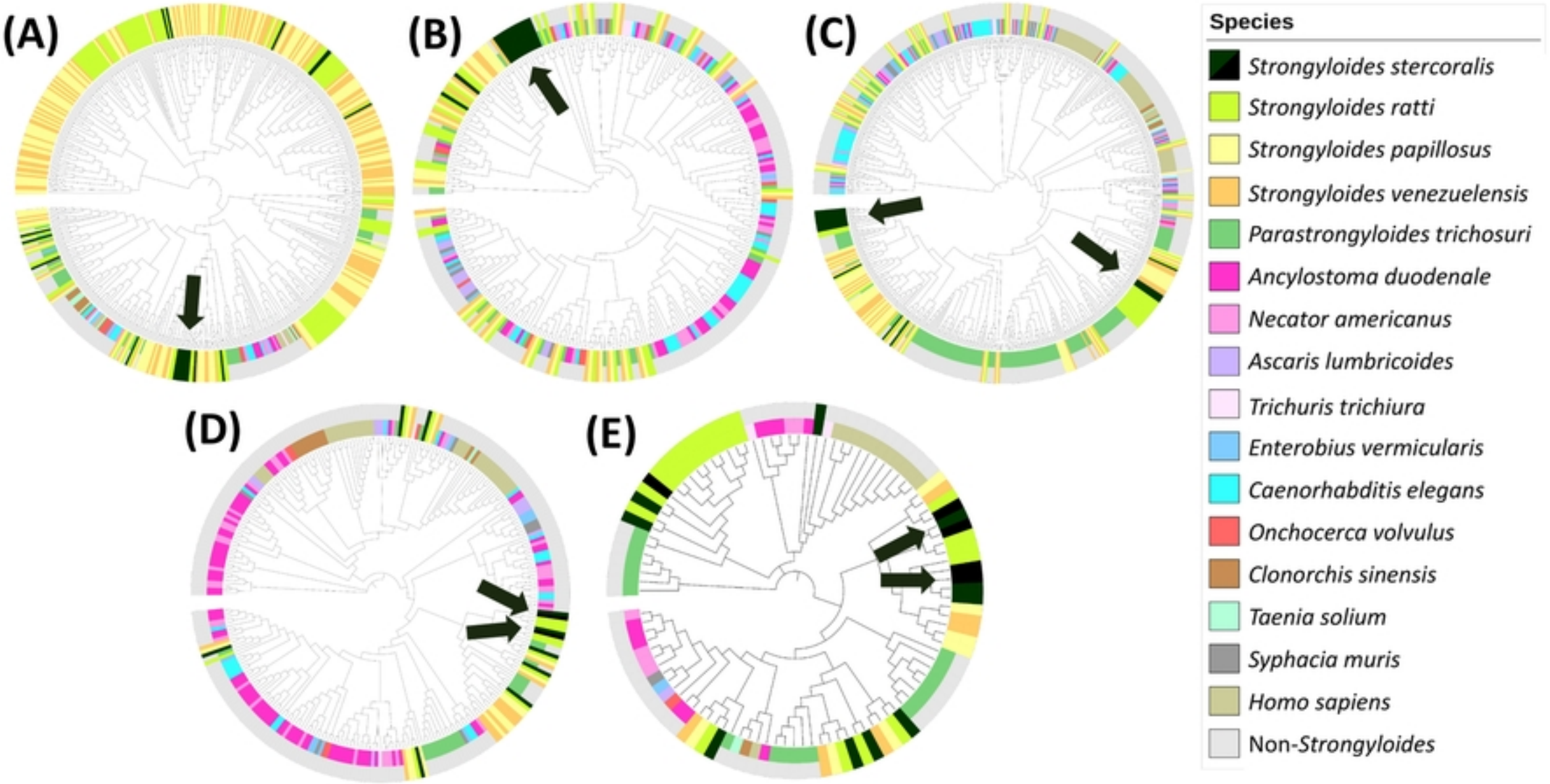

(A)
(B)
(C)
(D)
(E)
Species
Strongyloides stercoralis
Strongyloides ratti
Strongyloides papillosus
Strongyloides venezuelensis
Parastrongyloides trichosuri
Ancylostoma duodenale
Necator americanus
Ascaris lumbricoides
Trichuris trichiura
Enterobius vermicularis
Caenorhabditis elegans
Onchocerca volvulus
Clonorchis sinensis
Taenia solium
Syphacia muris
Homo sapiens
Non-Strongyloides

Figure 4

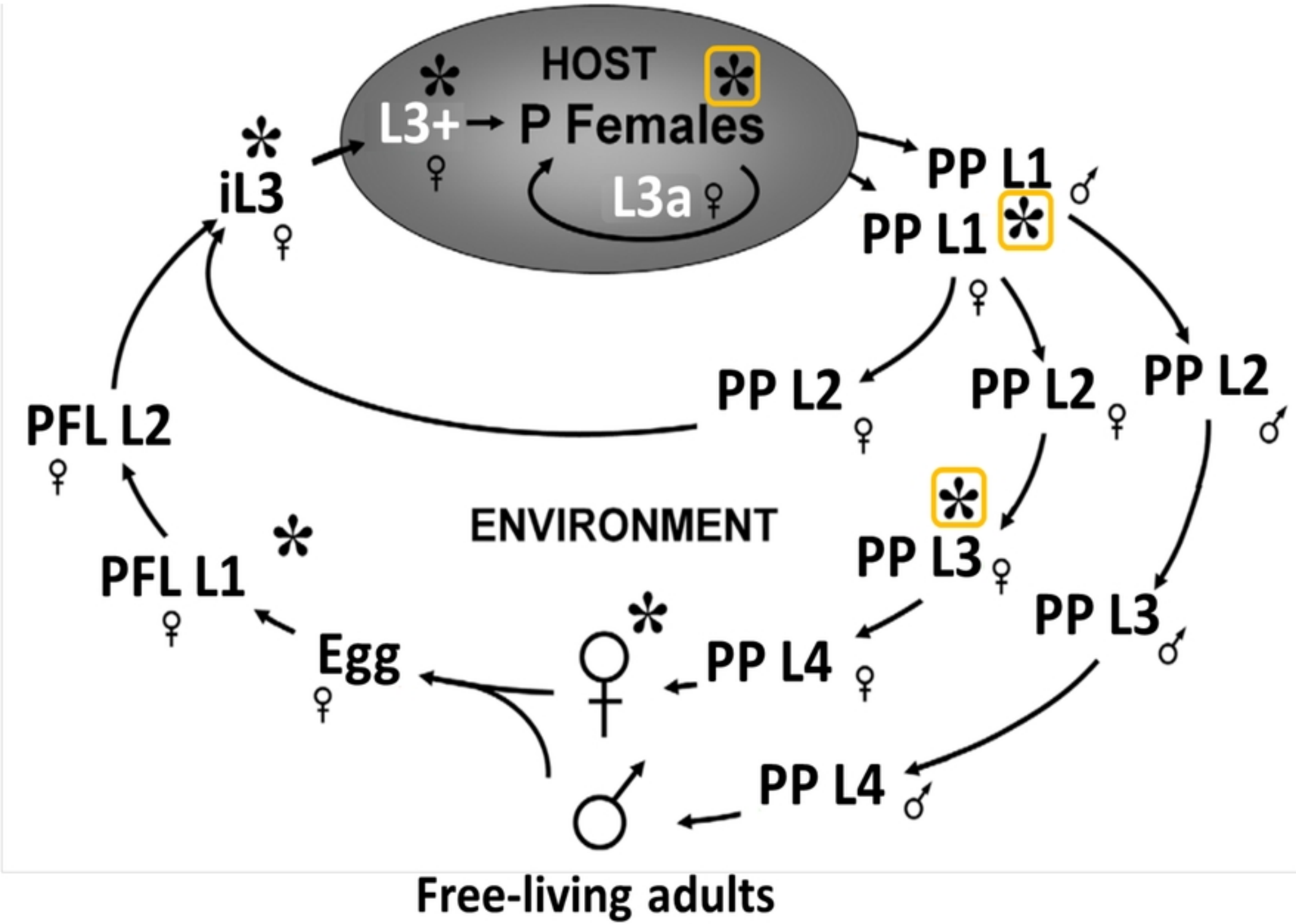

HOST
L3+
P Females
L3a
iL3
PP L1
PP L1
PP L2
PP L2
PP L2
PFL L2
ENVIRONMENT
PP L3
PP L3
PFL L1
Egg
PP L4
PP L4
Free-living adults

Figure 2

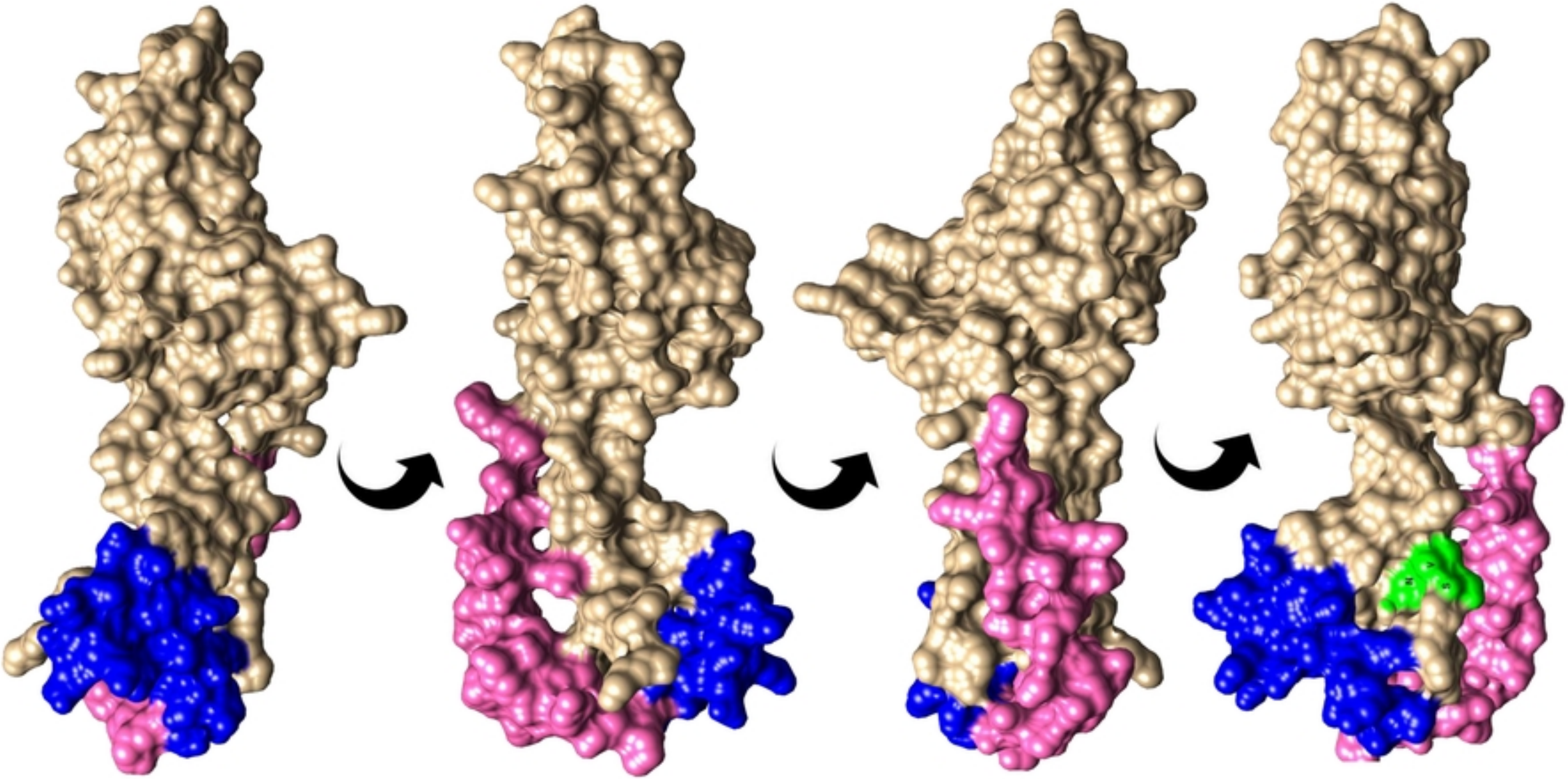


Figure 5

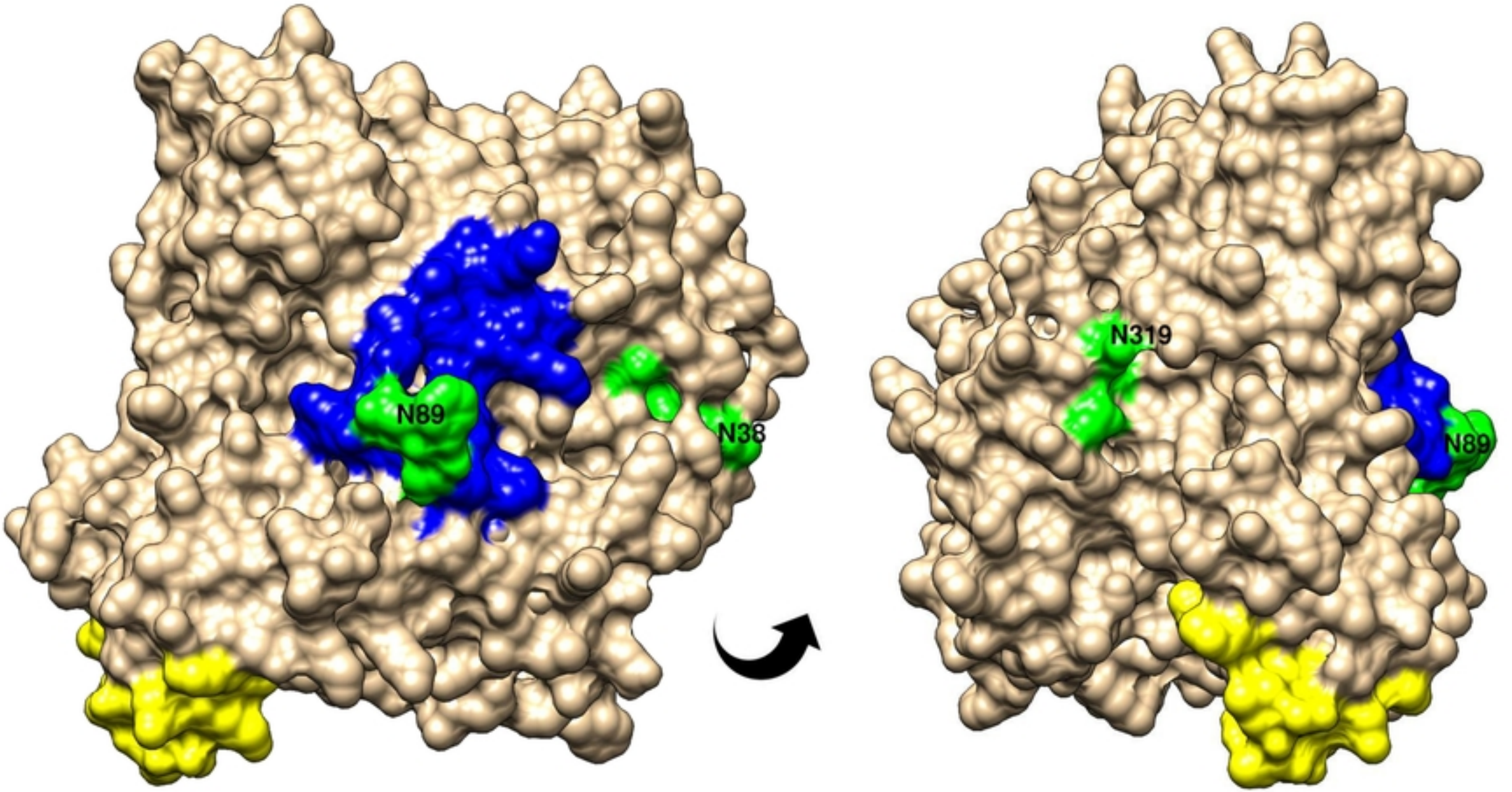


Figure 6

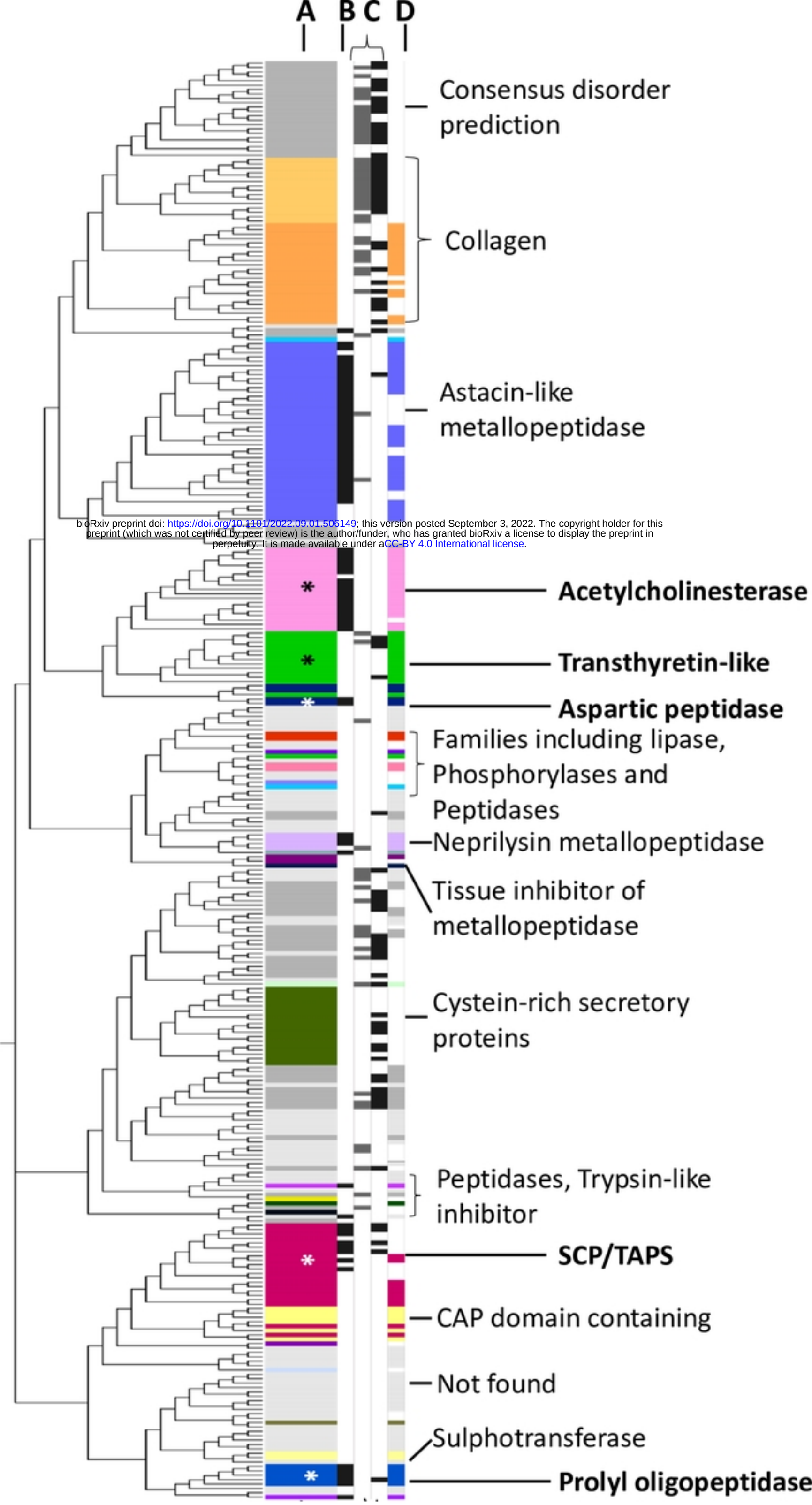


Figure 3