A reverse vaccinology approach identifies putative vaccination targets in the zoonotic nematode *Ascaris*.

Evangelista[a*], Francisco Miguel Dias; van Vliet[a], Arnoud H. M.; Lawton[b], Scott P.; Betson[a*], Martha

[a] School of Veterinary Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford GU2 7AL, United Kingdom

[b] Epidemiology Research Unit, Department of Veterinary and Animal Sciences, Northern Faculty, Scotland's Rural University College (SRUC), An Lòchran, Inverness, IV2 5NA, United Kingdom

* Corresponding author

Email:

Francisco Miguel Dias Evangelista: f.evangelista@surrey.ac.uk

Arnoud H. M. van Vliet: a.vanvliet@surrey.ac.uk

Scott P. Lawton: scott.Lawton@sruc.ac.uk

Corresponding author:

Martha Betson

School of Veterinary Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford GU2 7AL, United Kingdom

Phone: +44 1483 689821

Email: m.betson@surrey.ac.uk

Abstract

Word count: 214

Ascariasis is the most prevalent helminthic disease affecting both humans and pigs and is caused by the roundworms *Ascaris lumbricoides* and *Ascaris suum*. While preventive chemotherapy continues to be the most common control method, recent reports of anthelminthic resistance highlight the need for development of a vaccine against ascariasis. The aim of this study was to use a reverse vaccinology approach to identify potential vaccine candidates for *Ascaris*. Three *Ascaris* proteomes predicted from whole-genome sequences were analysed. Candidate proteins were identified using open-access bioinformatic tools (e.g. Vacceed, VaxiJen, Bepipred 2.0) which test for different characteristics such as sub-cellular location, T-cell and B-cell molecular binding, antigenicity, allergenicity and phylogenetic relationship with other nematode proteins. From over 100,000 protein sequences analysed, four transmembrane proteins were predicted to be non-allergen antigens and potential vaccine candidates. The four proteins are a Piezo protein, two voltage-dependent calcium channels and a protocadherin-like protein, are all expressed in either the muscle or ovaries of both *Ascaris* species, and all contained high affinity epitopes for T-cells and B-cells. The use of a reverse vaccinology approach allowed the prediction of four new potential vaccination targets against ascariasis in humans and pigs. These targets can now be further tested in *in vitro* and *in vivo* assays to prove efficacy in both pigs and humans.

Key words: *Ascaris lumbricoides*; *Ascaris suum*; Neglected Tropical Diseases; Nematodes; Zoonoses; Vaccine.

1. Introduction

Word count: 3820

The giant roundworm *Ascaris* is the most prevalent soil-transmitted helminth (STH) infection in humans, being responsible for 0.861 million disability-adjusted life years (DALYs) worldwide (Kyu et al., 2018). In parallel, *Ascaris* infections remain an issue on many pig farms worldwide. This parasite is especially important in farms with lower levels of biosecurity, such as organic farms, and/or in lower to medium income countries, where backyard farming is common (Dold and Holland, 2011; Katakam et al., 2016). Being a zoonotic disease, the close contact between humans and pigs in many highly endemic areas increases the risk *Ascaris* transmission among both hosts (Dold and Holland, 2011), (Easton et al., 2020).

Preventive chemotherapy, bolstered by implementation of water, sanitation and hygiene (WASH) protocols, continues to be the mainstay of ascariasis control in humans, as advocated in the World Health Organization (WHO) roadmap for the control of Neglected Tropical Diseases (NTDs) (World Health Organization, 2020). Reduced efficacy and resistance to benzimidazole drugs have already been reported in human *Ascaris* infections (Furtado et al., 2019; Krücken et al., 2017), and this directly affects the implementation of control protocols in endemic areas, necessitating the development of effective vaccines and vaccination protocols. Vaccination against *Ascaris* has had some degree of efficacy in mouse and pig animal models, but no candidate vaccine has undergone human clinical trials yet (Hotez et al., 2016). The proteins used in the past vaccination assays were either secreted proteins or those retrieved from crude extracts of adult worms. Recently, vaccination with a chimeric protein led to up to 73% larval burden reduction in mice, in contrast to a 99.8% reduction when mice were repeatedly infected with *Ascaris* sterile eggs (de Castro et al., 2021;

Gazzinelli-Guimarães et al., 2018; Tsuji et al., 2001, 2004). These results highlight that there are other antigens that could be tested. These antigens could then be incorporated in a multi-component vaccine along with the already known vaccination targets to stimulate a more complete immune response.

Several annotated genomes are now available for *Ascaris lumbricoides* and *Ascaris suum*, which enable the search for putative vaccine candidates using a reverse vaccinology analysis. A reverse vaccinology approach combines genome information and bioinformatic tools for identification of vaccine candidates, and has been successfully applied to other nematodes before, such as *Toxocara canis* and *Trichuris muris* (Salazar Garcés et al., 2020; Zawawi et al., 2020). Such methodology allows researchers to uncover vaccine targets from predicted proteomes by assessing if proteins have useful characteristics. Different analyses, such as protein sub-cellular location and the prediction of B-cell and T-cell epitopes, are examples of steps used in these approaches that help select proteins for further testing. This is especially important when funding is limited, and vaccination trials need to be focused. The aim of this study was to apply an *in silico* methodology to analyse protein sequences predicted from three *Ascaris* annotated proteomes to identify potential new vaccination targets that could be used in vaccination assays against *Ascaris*.

2. Materials and methods

2.1. Data Retrieval

The annotated proteomes for three assembled genomes of *Ascaris* spp. were retrieved from the WormBase ParaSite database (Howe et al., 2017). The *A. lumbricoides* proteome from BioProject PRJEB4950 (International Helminth Genomes Consortium, 2019) has 23,604 protein sequences.

The *A. suum* proteomes from BioProject PRJNA80881 (Jex et al., 2011) and BioProject PRJNA62057 (Wang et al., 2017, 2012) have 18,542 and 57,968 protein sequences, respectively.

2.2. Protein subcellular location analysis

The retrieved proteomes were first visualised and analysed with BioEdit v7.2.5 (Hall, 1999). A total of 2,604 protein sequences that included unknown amino acids (aa) (indicated with the symbol 'X') were excluded from further analysis as they tend to be a consequence of poor annotations and most of the bioinformatics tools used do not analyse protein sequences with unknown amino acids.

The framework Vacceed v2.1 (Goodswen et al., 2014) was used to identify potential vaccine candidates. The tools employed in Vacceed analysis were: WoLF PSORT v0.2 (Horton et al., 2007), DeepLoc v1.0 (Almagro Armenteros et al., 2017), SignalP v5.0 (José Juan Almagro Armenteros et al., 2019), TargetP v2.0 (Jose Juan Almagro Armenteros et al., 2019), TMHMM v2.0 (Krogh et al., 2001) and Phobius v1.01 (Käll et al., 2004). Proteins were ranked with scores between 0 and 1, from low immunogenicity (final score = 0) to high immunogenicity potential (final score = 1). Higher scores were given to proteins which were predicted to be secreted, signal or transmembrane peptides. The proteins sequences with a Vacceed score of ≥ 0.750 were retrieved (Palmieri et al., 2017) and later tested for epitope binding to CD4+ T helper (Th) cells.

As Vacceed makes use of different bioinformatic tools, there was the need to check how much each tool influences the final protein scores. After retrieving the Vacceed scores from runs including all programs, the process was repeated six times excluding one program per run (i.e. first run with all the bioinformatic tools except TMHMM, second with all the bioinformatic tools except DeepLoc, and so on) (Palmieri et al., 2017). The proteins scores were then compared between the different assessments through a Pearson correlation test using R v3.6.3 (R Core Team, 2020), RStudio

v1.2.5033 (RStudio Team, 2019) (with the package *corrplot* (Wei and Simko, 2017)) and IBM SPSS Statistics v26.

2.3. CD4+ Th cell binding predictions

The protein sequences retrieved with Vacceed were submitted to the standalone version of the Immune Epitope Database (IEDB) Major Histocompatibility Complex class II (MHC-II) binding predictor v2.22.3 (MHCII-IEDB, available at http://tools.iedb.org/mhcii/). This tool employs neural networks trained on IEDB experimentally validated epitopes to predict and quantify the binding affinity between a given peptide/antigen epitope and a selected MHC-II molecule recognized by CD4+ Th cells. The IEDB-recommended 2.22 prediction method was used, comprising the Consensus approach, NN-align, SMM-align, CombLib and Sturniolo methods, or NetMHCIIpan, if any of the previous methods was not available for the selected MHC-II allele (Fleri et al., 2017). This method was used against the 27 human leukocyte antigen (HLA) allele reference set, covering around 99% of the worlds human population (Greenbaum et al., 2011). The default peptide epitope length of 15 aa with a core of 9 aa was selected. The protein sequences which had epitopes ranked from zero to one (from the maximum of 100) and were simultaneously predicted to bind to all the 27 alleles in the used reference set were selected for further testing. These MHC-II binding predictions were only made for human alleles due to the lack of *in silico* tools that make binding predictions for swine MHC-II alleles.

2.4. Allergenicity, antigenicity and function predictions

The protein sequences retrieved with MHCII-IEDB were tested for antigenicity and allergenicity. Protein antigenicity was analysed using IEDB Kolaskar and Tongaonkar antigenicity scale and VaxiJen 2.0 (Doytchinova and Flower, 2007; Kolaskar and Tongaonkar, 1990). The IEDB Kolaskar

and Tongaonkar antigenicity scale method, available at http://tools.iedb.org/bcell/, was used with default setting and threshold of 1.000 and the VaxiJen 2.0 tool, server accessible at http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html, was used with the default threshold of 0.5. Allergenicity was evaluated using AllerTop 2.0 and AllergenFP (Dimitrov et al., 2014, 2013). The AllerTop 2.0 server, https://www.ddg-pharmfac.net/AllerTOP/index.html, and the AllergenFP tool, http://www.ddg-pharmfac.net/AllergenFP/index.html, analyse protein sequences and compare them to a training set of 2,427 known allergens and 2,427 non-allergens using two different algorithms for the better recognition of allergens and non-allergens, respectively for each tool. Proteins that were classified as both potential antigens and non-allergens were considered as good vaccination targets.

The function of the selected proteins, their protein family and tissue location in both *A. lumbricoides* and *A. suum* were analysed based on the most recent genomic and transcriptomic studies (Easton et al., 2020; Wang et al., 2017). These assays were able to map the location where all the genes and, therefore, proteins were transcribed in both *Ascaris* species as well as in what stages of the parasite lifecycle. This information is important as it allows us to infer what roles these proteins could play in the *Ascaris* lifecycle and how useful they would be as components in a vaccine.

2.5. CD4+ Th cell binding epitope selection

To further streamline the epitope selection process, we integrated the information gathered using the Phobius and MHCII-IEDB tools to select extracellular CD4+ Th binding epitopes. We used Phobius to assess the extracellular domains of the predicted proteins and identify if the previously predicted epitopes were present in those areas. Proteins that lacked the full 15 aa epitopes in these

areas were disregarded. For each protein, the two non-redundant epitopes (without overlapping aa) predicted to bind to the largest combined number of unique alleles in MHCII-IEDB tool analysis were retrieved. These epitopes were submitted to a BLASTp (Basic Local Alignment Search Tool protein) search (available at https://blast.ncbi.nlm.nih.gov/Blast.cgi) to check for identity in humans and pigs. Epitopes with 100% identity to human or pig epitopes were discarded. The BLASTp analysis was conducted with a threshold of e-value=0.05. Each epitope was also submitted to Allertop 2.0 (Dimitrov et al., 2014) to confirm they were not potential allergens. The best expressed protein transcript for each vaccine target, as found in the most recent genomic studies (Easton et al., 2020; Wang et al., 2017), was submitted to Protter to draw the two-dimensional (2D) protein structure using the information retrieved with Phobius (Käll et al., 2004; Omasits et al., 2014).

2.6. B-cell linear epitope prediction and selection

The presence of B-cell linear epitopes was assessed using Bepipred 2.0 webserver (available at http://www.cbs.dtu.dk/services/BepiPred/) (Jespersen et al., 2017). Using the same methodology as in the previous step, proteins that had epitopes found in extracellular domains, exposed to the host immunological system, predicted to be non-allergens, and present in most of the protein transcripts of the different genes were considered good vaccination targets. The two highest scored non-allergen epitopes, with a size between 8 and 40 aa, were retrieved for each protein.

2.7. Protein phylogenetic analysis

To assess the relationship between the predicted vaccine targets and their orthologues in other nematodes, a phylogenetic analysis of the predicted vaccine targets was performed including predicted orthologues/similar proteins present in annotated genomes of other parasitic nematodes and *Caenorhabditis elegans* available in WormBase ParaSite database (Howe et al., 2017). Proteins

were selected according to the list of orthologues provided by WormBase and, if no orthologue was present for a given nematode, a BlastP analysis was carried out and proteins were retrieved according to the combination of query coverage, identity percentage and e-value. The protein sequences were aligned using MUSCLE (http://www.ebi.ac.uk/Tools/muscle/index.html), the phylogenetic analysis was performed with Maximum-likelihood method using the JTT+G substitution model in MEGA-X (Kumar et al., 2018) and the predicted trees were visualized in iTOL (Letunic and Bork, 2021). Nodal support was tested using bootstrap values, which were calculated with 500 replicates.

3. Results

3.1. Initial protein selection

From the 100,114 protein sequences extracted from the combined *A. suum* and *A. lumbricoides* proteomes, 97,510 were analysed with the Vacceed framework for protein subcellular location. A total of 28.085 proteins were selected for further analysis (Fig.1), from which 5,984 protein sequences were from *A. lumbricoides*, and 5,004 and 17,097 were *A. suum* protein sequences from the two *A. suum* proteomes, BioProject PRJNA80881 and BioProject PRJNA62057, respectively. The BioProject PRJNA62057 has a higher number of proteins in its proteome due to the way it was assembled, where all the expressed transcripts were kept instead of being merged into a single protein sequence. The Vacceed final scores for each protein in the three proteomes are provided in Supplementary Table 1.

The correlation assessment of the Vacceed framework demonstrated a high correlation between the results of the different Vacceed runs (Figure 1). The removal of TMHMM from the framework

196 resulted in a slightly lowered correlation score when compared with the original Vacceed run
197 (r=0.97).

198 (Insert Figure 1)

199 **3.2. Selection of vaccine targets and identification of epitopes**

200 Before the CD4+ Th cell binding prediction was carried out, a further 2,906 duplicate sequences
201 were identified and removed from the *A. suum* PRJNA62057 dataset. A total of 25,179 protein
202 sequences were then analysed for epitope binding to MHC-II alleles. When analysing the final
203 output, 32 protein sequences had epitopes predicted to bind to all the 27 MHC-II alleles in the
204 reference set with a rank between 0.01 and 1. From this set of 32 protein sequences, 9 proteins
205 belonged to the *A. lumbricoides* proteome, while the other 23 proteins were from the two *A. suum*
206 proteomes, 3 from BioProject PRJNA80881 and 20 BioProject PRJNA62057. Only three of these
207 protein sequences were predicted to be secreted proteins. A table detailing the WormBase protein
208 identifiers and protein aa length for these 32 candidate proteins is provided in Supplementary Table
209 2.

210 The analysis with AllerTop 2.0 and AllergenFP did not predict any allergens in the 32 protein
211 sequences. However, antigen prediction tools classified four proteins sequences as non-antigenic,
212 and these proteins were disregarded. The remaining protein sequences that had similar results
213 between their respective orthologues in *A. lumbricoides* and *A. suum* were grouped into five
214 clusters, represented by distinct genes in all three *Ascaris* genomes. The five genes and their
215 respective transcripts are predicted to be responsible for the transcription of four membrane
216 transporters and one cell adhesion protein.

Four genes and respective proteins were selected as promising vaccination targets based on epitope exposure to the host immunological system: "Voltage-dependent T-type calcium channel subunit alpha" (ATtype [WormBase gene identifiers: *GS_24322*, *ALUE_0000418301* and *AgB13X_g094*]), "Piezo-type mechanosensitive ion channel component" (APiezo [WormBase gene identifiers: *GS_03113*, *ALUE_0000666901* and *AgR007_g063*]), "Voltage-dependent L-type calcium channel subunit alpha" (ALtype [WormBase gene identifiers: *GS_04697*, *ALUE_0001482301* and *AgR007_g282*]), and "Protocadherin-like" (AProto [WormBase gene identifiers: *GS_06422*, *ALUE_0000418601* and *AgB13X_g096*]). According to the most recent transcriptomic data, both ATtype and AProto are highly expressed in the ovaries while APiezo and ALtype are highly expressed in the muscle of adults (Easton et al., 2020; Wang et al., 2017). These proteins were predicted to have both CD4+ Th cell and B cell binding epitopes in extracellular areas (Fig. 2). For each protein, two non-allergen CD4+ Th cell binding epitopes were selected based on the results of the Allertop 2.0, MHCII-IEDB tool and their presence in the extracellular areas of the protein (Table 1). Two B-cell epitopes were also selected for each target. Supplementary Table 3 lists the epitopes found for each predicted vaccine target using MHCII-IEDB that scored between 0 and 1 and the MHC-II alleles they were predicted to bind to (only epitopes that were predicted to bind to two or more MHC-II alleles were retrieved). A workflow diagram of the analysis can be seen in Figure 3.

(Insert Table 1 and Figure 2)

(Insert Figure 3)

3.3. Phylogenetic relationships of vaccine targets identified across nematode species

To analyse the potential relationship between the predicted vaccine targets and proteins in other nematodes, orthologs of the selected vaccine targets were retrieved from other nematode proteomes present in WormBase Parasite (Supplementary Table 4), aligned and used for the generation of a phylogenetic tree. There is a general clustering of the proteins within each nematode clade. The exception occurred in the APiezo orthologues where clade III nematodes were separated in two different groups, with the Ascaridomorpha nematodes (e.g. *Ascaris*, *Toxocara* and *Parascaris*) being more closely related to the nematodes in Clade IV and the Spiruromorpha nematodes (e.g. *Onchocerca*, *Dirofilaria*, *Loa*, *Brugia* and *Wuchereria*) being closer to the Clade V nematodes. These relationships can be seen in Figure 4.

(Insert Figure 4)

4. Discussion

In this study we have used bioinformatic approaches to identify four different *Ascaris* proteins that could be included in multi-epitope vaccines against human and pig *Ascaris* infections. These genes are highly expressed in two distinct regions of the parasite: the ovaries and early egg stages, in the case of ATtype and Aproto, and in the muscle, the case for APiezo and ALtype (Easton et al., 2020; Wang et al., 2017). While both ATtype and ALtype are predicted to be calcium channels that promote calcium import with muscle and smooth muscle contraction through a voltage mechanism, APiezo is a mechanosensitive calcium channel that in *C. elegans* was shown to affect reproductive tissue development and malfunction (Bai et al., 2020). One interesting relationship shown in Figure 4.B is how APiezo appears to be more closely related between blood feeding parasites, such as *Necator americanus* and filarial parasites such as *Loa loa*, suggesting a role in adaptation to contact

with blood. The AProto protein is the most unique as to our knowledge, as protocadherins have not been reported before in nematodes. The structural domains appear to be more closely related to that of Flamingo/Stan cadherins due to the presence of both laminin-G and EGF-like receptors with seven transmembrane domains (with these last highlighted in Figure 2D) (Easton et al., 2020; Hardin et al., 2013). Being highly transcribed in the ovaries and early egg stages, while predicted to be responsible for homophilic cell adhesion, AProto could have a role in early oocyte development. Although present in regions usually considered difficult recognize by the host immune system, such as the parasite's muscle and ovaries, an IgG immunoblot assay showed that this is possible, as proteins highly transcribed in the muscle, ovaries and intestines of the parasite were recognized by serum from pigs infected with *Ascaris* (González-Miguel et al., 2014). It is also interesting to highlight the close relationship between the predicted vaccine targets and the orthologues in other nematodes, suggesting that these orthologues might also be useful in the control of the respective species.

Previous attempts to generate a vaccine against ascariasis used crude extracts, recombinant proteins and, more recently, chimeric proteins (de Castro et al., 2021; Gazzinelli-Guimarães et al., 2018; Tsuji et al., 2004, 2001). The highest lung larvae burden reduction achieved using a multi-epitope or recombinant protein vaccine was 73.5%, when using a chimeric protein containing B-cell epitopes of the As14, As16 and As37 *Ascaris* proteins (de Castro et al., 2021). In comparison, vaccination against trichuriasis in mice resulted in up to 97% reduction of adult nematodes (Gomez-Samblas et al., 2017). The vaccine against *Trichuris* is based on recombinant proteins and showed efficacies vastly superior to similar vaccines against ascariasis. With vaccine development against ascariasis lagging behind other parasitic diseases, there is a need to discover other antigens to be tested as vaccine candidates. The reverse vaccinology approach used in our study is based on

genomic and proteomic data to predict which proteins may be usable as vaccine candidates prior to new *in vivo* studies. This methodology allows researchers to focus down vaccination assays to a smaller set of proteins, effectively reducing costs and time. This reverse vaccinology approach has been successfully applied to identify vaccination targets for other nematodes, for example *T. canis* (Salazar Garcés et al., 2020) and *T. muris* (Zawawi et al., 2020).

Our reverse vaccinology analysis used all the proteins predicted in the three *Ascaris* proteomes. Previous *in silico* studies on vaccine candidate prediction in *Ascaris* focused exclusively on secreted proteins (Ebner et al., 2020). The workflow applied in this study allows the testing of all the proteins predicted from a genome analysis, without automatically excluding non-secreted proteins. In a recent study, an *A. lumbricoides* multi-epitope vaccine candidate was developed using *in silico* methodology and proteins were selected based on their binding to the HLA-DRB1*07:01 and HLA-DRB1*15:01 MHC-II alleles (Kaur et al., 2021). Although these MHC-II alleles are known to recognise *Ascaris* antigens in humans, they only cover up to 15% of the human population in areas where human ascariasis is endemic (Ebner et al., 2020). This might prove detrimental in *in vivo* studies that cover population that do not have these MHC-II alleles, limiting its usefulness.

Each of the three genomes included in the analysis were predicted to have over 5,000 proteins that could be further investigated as good vaccine targets according to Vacceed, which corresponds to 25-29% of all the proteins present. As the number of secreted proteins in *Ascaris* is predicted to be 254 proteins (Ebner et al., 2020), this suggests that the number of potential vaccination targets might be vastly superior to the ones that are usually investigated in these species and other helminths. The final candidates are all predicted to be non-secreted proteins. This contrasts with most of the previously studied vaccine candidates, except for the muscle membrane-bound As37 protein (Versteeg et al., 2020). As37 was not predicted to be a good target in Vacceed (with scores

of 0.001 in all three genomes), showing some limitations to the method we used. However, our predictions and the protection achieved with the As37 recombinant protein support the idea that only targeting secreted proteins can be detrimental to the selection of good vaccination targets in *Ascaris*. Recent work in *Toxocara canis,* a parasite of the same family as *Ascaris*, showed that membrane proteins are capable to induce protection in a mouse model, reinforcing the idea that it is an error to disregard these proteins when looking for vaccination targets in nematodes (Salazar Garcés et al., 2020).

The underlying host's immune responses against *Ascaris* are still only partially understood (Zawawi and Else, 2020). This is a disadvantage when selecting the right tools to predict which proteins could be useful as vaccination targets. MHC-II molecules appear to have a prominent role in the control of nematode infections in mammals, including *A. lumbricoides* infections in humans (Garamszegi and Nunn, 2011; Zawawi and Else, 2020). This role makes the discovery of epitopes that bind to these molecules a priority for the design of multi-epitope/subunit-based vaccines. The MHCII-IEDB tool was chosen for this purpose as it was used in selecting epitopes for vaccination assays against other nematodes and is one of the most accurate tools available (Ebner et al., 2020; Zawawi et al., 2020). A reference set of 27 different alleles was chosen due to the fact that heterogenicity throughout the human population leads to different immune responses(Ebner et al., 2020). Thus, we wanted to selected proteins that would be able to induce a helpful immune reaction in a large portion of the population, and not just focus on one specific allele. Unfortunately, this biases towards selecting larger proteins as they contain a larger number of epitopes. These predictions were made only for human alleles due to unavailability of similar open-access bioinformatic tools for swine MHC-II alleles. This is a limitation of the reverse vaccinology approach to discovering vaccine targets in pigs. Bepipred 2.0 predicted the presence of B cell

328 binding epitopes in same proteins which would enhance the host immune response against these
329 proteins in their native form in the parasite.

330 Whilst the proteins we selected have been predicted to be useful and have similarities to other
331 proteins tested in vaccination assays against other nematodes, there is still the need to test them *in*
332 *vitro* and *in vivo*. The next step should be to confirm *in vitro* that both humans and pigs are able to
333 recognize these proteins targets and their respective epitopes. Should these proteins prove useful in
334 stimulating an immune reaction in the host, we propose that the proteins and respective epitopes
335 identified in this work should be incorporated into a multi-epitope vaccine which, ideally, would
336 include CD4+ Th cell and B-cell binding epitopes from other proteins, such as As14, As16 and
337 As37. Ideally, such a vaccine should be tested using a pig model to assess its potential effect on
338 both larvae and adult *Ascaris*, impossible to assess using a mouse model. This is more relevant
339 when testing for the utility of incorporating ATtype and AProto epitopes due to protein higher
340 expression in the ovaries of adult female *Ascaris* (Easton et al., 2020; Wang et al., 2017). Should
341 these proteins not prove useful, a protocol optimization should be done to identify new targets so
342 that, in the future, it might also be applied to other species that are in need of further studies.

343 In conclusion, this study highlights the role that reverse vaccinology and *in silico* methodology can
344 play in identification of vaccine candidates for parasitic diseases. The genome-wide approach,
345 without bias towards secreted proteins, led to the prediction of four novel candidates that were not
346 identified in previous studies but show the promise of promoting a useful immune response in
347 vaccination assays against *A. lumbricoides* and *A. suum*. These proteins should now be tested *in*
348 *vitro* and in combination with already known vaccine targets. Ultimately, the findings of this study
349 will support the future development of a vaccine against both ascariasis in humans and pigs, thus

promoting the health of both populations by reducing the need to use Mass-Drug administrations and decreasing the risk of anthelmintic resistance appearing.

Acknowledgements

The authors would like to thank Dr. Stephen Goodswen for assistance and troubleshooting with the Vacceed tool and Dr. Pritesh Taylor for help with computer resources. Graphical abstract created with BioRender.com.

Conflict of interest: All authors report no conflict of interest.

Author contributions:

Concept and design: FE, AV, SL and MB

Data acquisition and analysis: FE

Data interpretation: FE, AV, SL and MB

Manuscript draft: FE

Critical review and editing of the manuscript: FE, AV, SL and MB

Financial support: This study was funded by the Doctoral College Studentship Award from the University of Surrey awarded to Francisco Miguel Dias Evangelista.

References

Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., Nielsen, H., 2019. Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance 2, e201900429. https://doi.org/10.26508/lsa.201900429

Almagro Armenteros, J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H., Winther, O., 2017. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33, 3387–3395. https://doi.org/10.1093/bioinformatics/btx431

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37, 420–423. https://doi.org/10.1038/s41587-019-0036-z

Bai, X., Bouffard, J., Lord, A., Brugman, K., Sternberg, P.W., Cram, E.J., Golden, A., 2020. *Caenorhabditis elegans* PIEZO channel coordinates multiple reproductive tissues to govern ovulation. Elife 9, e53603. https://doi.org/10.7554/eLife.53603

de Castro, J.C., de Almeida, L. V., Cardoso, M.S., Oliveira, F.M.S., Nogueira, D.S., Reis-Cunha, J.L., Magalhaes, L.M.D., Zhan, B., Bottazzi, M.E., Hotez, P.J., Bueno, L.L., Bartholomeu, D.C., Fujiwara, R.T., 2021. Vaccination with chimeric protein induces protection in murine model against ascariasis. Vaccine 39, 394–401. https://doi.org/10.1016/j.vaccine.2020.11.046

Dimitrov, I., Bangov, I., Flower, D.R., Doytchinova, I., 2014. AllerTOP v.2—a server for in silico prediction of allergens. J. Mol. Model. 20, 2278. https://doi.org/10.1007/s00894-014-2278-5

Dimitrov, I., Naneva, L., Doytchinova, I., Bangov, I., 2013. AllergenFP: allergenicity prediction by

descriptor fingerprints. Bioinformatics 30, 846–851. https://doi.org/10.1093/bioinformatics/btt619

Dold, C., Holland, C. V, 2011. *Ascaris* and ascariasis. Microbes Infect. 13, 632–637. https://doi.org/10.1016/j.micinf.2010.09.012

Doytchinova, I.A., Flower, D.R., 2007. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine 25, 856–866. https://doi.org/https://doi.org/10.1016/j.vaccine.2006.09.032

Easton, A., Gao, S., Lawton, S.P., Bennuru, S., Khan, A., Dahlstrom, E., Oliveira, R.G., Kepha, S., Porcella, S.F., Webster, J., Anderson, R., Grigg, M.E., Davis, R.E., Wang, J., Nutman, T.B., 2020. Molecular evidence of hybridization between pig and human *Ascaris* indicates an interbred species complex infecting humans. Elife 9, e61562. https://doi.org/10.7554/eLife.61562

Ebner, F., Morrison, E., Bertazzon, M., Midha, A., Hartmann, S., Freund, C., Álvaro-Benito, M., 2020. CD4+ Th immunogenicity of the Ascaris spp. secreted products. npj Vaccines 5, 25. https://doi.org/10.1038/s41541-020-0171-z

Fleri, W., Paul, S., Dhanda, S.K., Mahajan, S., Xu, X., Peters, B., Sette, A., 2017. The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. Front. Immunol. 8, 278. https://doi.org/10.3389/fimmu.2017.00278

Furtado, L.F.V., Medeiros, C. da S., Zuccherato, L.W., Alves, W.P., de Oliveira, V.N.G.M., da Silva, V.J., Miranda, G.S., Fujiwara, R.T., Rabelo, É.M.L., 2019. First identification of the benzimidazole resistance-associated F200Y SNP in the beta-tubulin gene in Ascaris

lumbricoides. PLoS One 14, e0224108.

Garamszegi, L.Z., Nunn, C.L., 2011. Parasite-mediated evolution of the functional part of the MHC in primates. J. Evol. Biol. 24, 184–195. https://doi.org/10.1111/j.1420-9101.2010.02156.x

Gazzinelli-Guimarães, A.C., Gazzinelli-Guimarães, P.H., Nogueira, D.S., Oliveira, F.M.S., Barbosa, F.S., Amorim, C.C.O., Cardoso, M.S., Kraemer, L., Caliari, M.V., Akamatsu, M.A., Ho, P.L., Jones, K.M., Weatherhead, J., Bottazzi, M.E., Hotez, P.J., Zhan, B., Bartholomeu, D.C., Russo, R.C., Bueno, L.L., Fujiwara, R.T., 2018. IgG Induced by Vaccination With *Ascaris suum* Extracts Is Protective Against Infection. Front. Immunol.

Gomez-Samblas, M., García-Rodríguez, J.J., Trelis, M., Bernal, D., Lopez-Jaramillo, F.J., Santoyo-Gonzalez, F., Vilchez, S., Espino, A.M., Bolás-Fernández, F., Osuna, A., 2017. Self-adjuvanting C18 lipid vinil sulfone-PP2A vaccine: study of the induced immunomodulation against Trichuris muris infection. Open Biol. 7, 170031. https://doi.org/10.1098/rsob.170031

González-Miguel, J., Morchón, R., Gussoni, S., Bossetti, E., Hormaeche, M., Kramer, L.H., Simón, F., 2014. Immunoproteomic approach for identification of *Ascaris suum* proteins recognized by pigs with porcine ascariasis. Vet. Parasitol. 203, 343–348. https://doi.org/https://doi.org/10.1016/j.vetpar.2014.03.031

Goodswen, S.J., Kennedy, P.J., Ellis, J.T., 2014. Vacceed: a high-throughput in silico vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology. Bioinformatics 30, 2381–2383. https://doi.org/10.1093/bioinformatics/btu300

Greenbaum, J., Sidney, J., Chung, J., Brander, C., Peters, B., Sette, A., 2011. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different

supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics 63, 325–335. https://doi.org/10.1007/s00251-011-0513-0

Hall, T.A., 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hardin, J., Lynch, A., Loveless, T., Pettitt, J., 2013. Chapter Eleven - Cadherins and Their Partners in the Nematode Worm Caenorhabditis elegans, in: van Roy, F.B.T.-P. in M.B. and T.S. (Ed.), The Molecular Biology of Cadherins. Academic Press, pp. 239–262. https://doi.org/https://doi.org/10.1016/B978-0-12-394311-8.00011-X

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., Nakai, K., 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35, W585–W587. https://doi.org/10.1093/nar/gkm259

Hotez, P.J., Strych, U., Lustigman, S., Bottazzi, M.E., 2016. Human anthelminthic vaccines: Rationale and challenges. Vaccine 34, 3549–3555. https://doi.org/https://doi.org/10.1016/j.vaccine.2016.03.112

Howe, K.L., Bolt, B.J., Shafie, M., Kersey, P., Berriman, M., 2017. WormBase ParaSite − a comprehensive resource for helminth genomics. Mol. Biochem. Parasitol. 215, 2–10. https://doi.org/https://doi.org/10.1016/j.molbiopara.2016.11.005

International Helminth Genomes Consortium, 2019. Comparative genomics of the major parasitic worms. Nat. Genet. 51, 163—174. https://doi.org/10.1038/s41588-018-0262-1

Jespersen, M.C., Peters, B., Nielsen, M., Marcatili, P., 2017. BepiPred-2.0: improving sequence-

448 based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 45, W24–
449 W29. https://doi.org/10.1093/nar/gkx346

450 Jex, A.R., Liu, S., Li, B., Young, N.D., Hall, R.S., Li, Y., Yang, L., Zeng, N., Xu, X., Xiong, Z.,
451 Chen, F., Wu, X., Zhang, G., Fang, X., Kang, Y., Anderson, G.A., Harris, T.W., Campbell,
452 B.E., Vlaminck, J., Wang, T., Cantacessi, C., Schwarz, E.M., Ranganathan, S., Geldhof, P.,
453 Nejsum, P., Sternberg, P.W., Yang, H., Wang, J., Wang, J., Gasser, R.B., 2011. *Ascaris suum*
454 draft genome. Nature 479, 529–533. https://doi.org/10.1038/nature10553

455 Käll, L., Krogh, A., Sonnhammer, E.L.L., 2004. A Combined Transmembrane Topology and Signal
456 Peptide Prediction Method. J. Mol. Biol. 338, 1027–1036.
457 https://doi.org/https://doi.org/10.1016/j.jmb.2004.03.016

458 Katakam, K.K., Thamsborg, S.M., Dalsgaard, A., Kyvsgaard, N.C., Mejer, H., 2016. Environmental
459 contamination and transmission of Ascaris suum in Danish organic pig farms. Parasit. Vectors
460 9, 80. https://doi.org/10.1186/s13071-016-1349-0

461 Kaur, R., Arora, N., Rawat, S.S., Keshri, A.K., Singh, N., Show, S.K., Kumar, P., Mishra, A.,
462 Prasad, A., 2021. Immunoinformatics driven construction of multi-epitope vaccine candidate
463 against *Ascaris lumbricoides* using its entire immunogenic epitopes. Expert Rev. Vaccines 1–
464 13. https://doi.org/10.1080/14760584.2021.1974298

465 Kolaskar, A.S., Tongaonkar, P.C., 1990. A semi-empirical method for prediction of antigenic
466 determinants on protein antigens. FEBS Lett. 276, 172–174. https://doi.org/10.1016/0014-
467 5793(90)80535-Q

468 Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane

protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305, 567–580. https://doi.org/10.1006/jmbi.2000.4315

Krücken, J., Fraundorfer, K., Mugisha, J.C., Ramünke, S., Sifft, K.C., Geus, D., Habarugira, F., Ndoli, J., Sendegeya, A., Mukampunga, C., Bayingana, C., Aebischer, T., Demeler, J., Gahutu, J.B., Mockenhaupt, F.P., von Samson-Himmelstjerna, G., 2017. Reduced efficacy of albendazole against *Ascaris lumbricoides* in Rwandan schoolchildren. Int. J. Parasitol. Drugs Drug Resist. 7, 262–271. https://doi.org/10.1016/j.ijpddr.2017.06.001

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Kyu, H.H., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R.S., Abebe, M., Abebe, Z., Abil, O.Z., Aboyans, V., Abrham, A.R., Abu-Raddad, L.J., Abu-Rmeileh, N.M.E., Accrombessi, M.M.K., Acharya, D., Acharya, P., Ackerman, I.N., Adamu, A.A., Adebayo, O.M., Adekanmbi, V., Ademi, Z., Adetokunboh, O.O., Adib, M.G., Adsuar, J.C., Afanvi, K.A., Afarideh, M., Afshin, A., Agarwal, G., Agesa, K.M., Aggarwal, R., Aghayan, S.A., Agrawal, A., Ahmadi, A., Ahmadi, M., Ahmadieh, H., Ahmed, M.B., Ahmed, S., Aichour, A.N., Aichour, I., Aichour, M.T.E., Akinyemiju, T., Akseer, N., Al-Aly, Z., Al-Eyadhy, A., Al-Mekhlafi, H.M., Al-Raddadi, R.M., Alahdab, F., Alam, K., Alam, T., Alashi, A., Alavian, S.M., Alene, K.A., Alijanzadeh, M., Alizadeh-Navaei, R., Aljunid, S.M., Alkerwi, A., Alla, F., Allebeck, P., Alonso, J., Alsharif, U., Altirkawi, K., Alvis-Guzman, N., Aminde, L.N., Amini, E., Amiresmaili, M., Ammar, W., Amoako, Y.A., Anber, N.H., Andrei, C.L., Androudi, S.,

Animut, M.D., Anjomshoa, M., Ansha, M.G., Antonio, C.A.T., Anwari, P., Arabloo, J., Aremu, O., Ärnlöv, J., Arora, A., Arora, M., Artaman, A., Aryal, K.K., Asayesh, H., Ataro, Z., Ausloos, M., Avila-Burgos, L., Avokpaho, E.F.G.A., Awasthi, A., Ayala Quintanilla, B.P., Ayer, R., Azzopardi, P.S., Babazadeh, A., Badali, H., Balakrishnan, K., Bali, A.G., Banach, M., Banoub, J.A.M., Barac, A., Barboza, M.A., Barker-Collo, S.L., Bärnighausen, T.W., Barquera, S., Barrero, L.H., Bazargan-Hejazi, S., Bedi, N., Beghi, E., Behzadifar, M., Behzadifar, M., Bekele, B.B., Bekru, E.T., Belachew, A.B., Belay, Y.A., Bell, M.L., Bello, A.K., Bennett, D.A., Bensenor, I.M., Berhane, A., Bernabe, E., Bernstein, R.S., Beuran, M., Beyranvand, T., Bhala, N., Bhatt, S., Bhaumik, S., Bhutta, Z.A., Biadgo, B., Biehl, M.H., Bijani, A., Bikbov, B., Bilano, V., Bililign, N., Bin Sayeed, M.S., Bisanzio, D., Bjørge, T., Bleyer, A., Bobasa, E.M., Bou-Orm, I.R., Boufous, S., Bourne, R., Brady, O.J., Brant, L.C., Brayne, C., Brazinova, A., Breitborde, N.J.K., Brenner, H., Briant, P.S., Briko, A.N., Britton, G., Brugha, T., Buchbinder, R., Busse, R., Butt, Z.A., Cahuana-Hurtado, L., Campuzano Rincon, J.C., Cano, J., Cárdenas, R., Carrero, J.J., Carter, A., Carvalho, F., Castañeda-Orjuela, C.A., Castillo Rivas, J., Castro, F., Catalá-López, F., Cercy, K.M., Cerin, E., Chaiah, Y., Chang, J.-C., Charlson, F.J., Chattu, V.K., Chiang, P.P.-C., Chitheer, A., Choi, J.-Y.J., Christensen, H., Christopher, D.J., Chung, S.-C., Cicuttini, F.M., Cirillo, M., Collado-Mateo, D., Cooper, C., Cortesi, P.A., Cortinovis, M., Cousin, E., Criqui, M.H., Cromwell, E.A., Cross, M., Crump, J.A., Daba, A.K., Dachew, B.A., Dadi, A.F., Dandona, L., Dandona, R., Dargan, P.I., Daryani, A., Das Gupta, R., Das Neves, J., Dasa, T.T., Davitoiu, D.V., De La Hoz, F.P., De Leo, D., De Neve, J.-W., De Steur, H., Degefa, M.G., Degenhardt, L., Deiparine, S., Demoz, G.T., Denova-Gutiérrez, E., Deribe, K., Dervenis, N., Des Jarlais, D.C., Dey, S., Dharmaratne, S.D., Dhimal, M., Dinberu, M.T., Dirac, M.A., Djalalinia, S., Doan, L., Dokova, K., Doku, D.T., Dorsey, E.R., Doyle, K.E., Driscoll, T.R., Dubey, M., Dubljanin, E., Duken,

E.E., Duncan, B.B., Duraes, A.R., Ebrahimi, H., Ebrahimpour, S., Echko, M.M., Edessa, D., Edvardsson, D., Effiong, A., Eggen, A.E., Ehrlich, J.R., El Bcheraoui, C., El-Khatib, Z., Elyazar, I.R.F., Enayati, A., Endalifer, M.L., Endries, A.Y., Er, B., Erskine, H.E., Eskandarieh, S., Esteghamati, A., Esteghamati, S., Fakhim, H., Faramarzi, M., Fareed, M., Farhadi, F., Farid, T.A., Farinha, C.S.E. sá, Farioli, A., Faro, A., Farzadfar, F., Fazaeli, A.A., Feigin, V.L., Fentahun, N., Fereshtehnejad, S.-M., Fernandes, E., Fernandes, J.C., Ferrari, A.J., Ferreira, M.L., Filip, I., Fischer, F., Fitzmaurice, C., Foigt, N.A., Foreman, K.J., Frank, T.D., Fukumoto, T., Fullman, N., Fürst, T., Furtado, J.M., Gakidou, E., Gall, S., Gallus, S., Ganji, M., Garcia-Basteiro, A.L., Gardner, W.M., Gebre, A.K., Gebremedhin, A.T., Gebremichael, T.G., Gelano, T.F., Geleijnse, J.M., Genova-Maleras, R., Geramo, Y.C.D., Gething, P.W., Gezae, K.E., Ghadami, M.R., Ghadiri, K., Ghasemi-Kasman, M., Ghimire, M., Ghoshal, A.G., Gill, P.S., Gill, T.K., Ginawi, I.A., Giussani, G., Gnedovskaya, E. V, Goldberg, E.M., Goli, S., Gómez-Dantés, H., Gona, P.N., Gopalani, S.V., Gorman, T.M., Goulart, A.C., Goulart, B.N.G., Grada, A., Grosso, G., Gugnani, H.C., Guillemin, F., Guo, Y., Gupta, P.C., Gupta, R., Gupta, R., Gupta, T., Gutiérrez, R.A., Gyawali, B., Haagsma, J.A., Hachinski, V., Hafezi-Nejad, N., Haghparast Bidgoli, H., Hagos, T.B., Hailegiyorgis, T.T., Haj-Mirzaian, A., Haj-Mirzaian, A., Hamadeh, R.R., Hamidi, S., Handal, A.J., Hankey, G.J., Hao, Y., Harb, H.L., Harikrishnan, S., Haririan, H., Haro, J.M., Hassankhani, H., Hassen, H.Y., Havmoeller, R., Hay, R.J., Hay, S.I., Hedayatizadeh-Omran, A., Heibati, B., Hendrie, D., Henok, A., Heredia-Pi, I., Herteliu, C., Heydarpour, F., Heydarpour, P., Hibstu, D.T., Hoek, H.W., Hoffman, H.J., Hole, M.K., Homaie Rad, E., Hoogar, P., Hosgood, H.D., Hosseini, S.M., Hosseinzadeh, M., Hostiuc, M., Hostiuc, S., Hotez, P.J., Hoy, D.G., Hsairi, M., Htet, A.S., Huang, J.J., Iburg, K.M., Ikeda, C.T., Ilesanmi, O.S., Irvani, S.S.N., Irvine, C.M.S., Islam, S.M.S., Islami, F., Jacobsen, K.H., Jahangiry, L., Jahanmehr, N., Jain, S.K., Jakovljevic, M., James, S.L.,

539 Jayatilleke, A.U., Jeemon, P., Jha, R.P., Jha, V., Ji, J.S., Johnson, C.O., Jonas, J.B.,
540 Jonnagaddala, J., Jorjoran Shushtari, Z., Joshi, A., Jozwiak, J.J., Jungari, S.B., Jürisson, M.,
541 Kabir, Z., Kadel, R., Kahsay, A., Kalani, R., Kanchan, T., Kar, C., Karami, M., Karami Matin,
542 B., Karch, A., Karema, C., Karimi, N., Karimi, S.M., Kasaeian, A., Kassa, D.H., Kassa, G.M.,
543 Kassa, T.D., Kassebaum, N.J., Katikireddi, S.V., Kaul, A., Kawakami, N., Kazemi, Z.,
544 Karyani, A.K., Keighobadi, M.M., Keiyoro, P.N., Kemmer, L., Kemp, G.R., Kengne, A.P.,
545 Keren, A., Khader, Y.S., Khafaei, B., Khafaie, M.A., Khajavi, A., Khalid, N., Khalil, I.A.,
546 Khan, E.A., Khan, M.S., Khan, M.A., Khang, Y.-H., Khater, M.M., Khazaei, M., Khoja, A.T.,
547 Khosravi, A., Khosravi, M.H., Kiadaliri, A.A., Kidanemariam, Z.T., Kiirithio, D.N., Kim, C.-
548 I., Kim, D., Kim, Y.-E., Kim, Y.J., Kimokoti, R.W., Kinfu, Y., Kisa, A., Kissimova-Skarbek,
549 K., Knudsen, A.K.S., Kocarnik, J.M., Kochhar, S., Kokubo, Y., Kolola, T., Kopec, J.A.,
550 Kosen, S., Kotsakis, G.A., Koul, P.A., Koyanagi, A., Krishan, K., Krishnaswami, S., Krohn,
551 K.J., Kuate Defo, B., Kucuk Bicer, B., Kumar, G.A., Kumar, M., Kuzin, I., Lad, D.P., Lad,
552 S.D., Lafranconi, A., Lalloo, R., Lallukka, T., Lami, F.H., Lang, J.J., Langan, S.M., Lansingh,
553 V.C., Latifi, A., Lau, K.M.-M., Lazarus, J. V, Leasher, J.L., Ledesma, J.R., Lee, P.H., Leigh,
554 J., Leili, M., Leshargie, C.T., Leung, J., Levi, M., Lewycka, S., Li, S., Li, Y., Liang, X., Liao,
555 Y., Liben, M.L., Lim, L.-L., Lim, S.S., Limenih, M.A., Linn, S., Liu, S., Looker, K.J., Lopez,
556 A.D., Lorkowski, S., Lotufo, P.A., Lozano, R., Lucas, T.C.D., Lunevicius, R., Lyons, R.A.,
557 Ma, S., Macarayan, E.R.K., Mackay, M.T., Maddison, E.R., Madotto, F., Maghavani, D.P.,
558 Mai, H.T., Majdan, M., Majdzadeh, R., Majeed, A., Malekzadeh, R., Malta, D.C., Mamun,
559 A.A., Manda, A.-L., Manguerra, H., Mansournia, M.A., Mantilla Herrera, A.M., Mantovani,
560 L.G., Maravilla, J.C., Marcenes, W., Marks, A., Martins-Melo, F.R., Martopullo, I., März, W.,
561 Marzan, M.B., Massano, J., Massenburg, B.B., Mathur, M.R., Maulik, P.K., Mazidi, M.,
562 McAlinden, C., McGrath, J.J., McKee, M., McMahon, B.J., Mehata, S., Mehrotra, R., Mehta,

563 K.M., Mehta, V., Mejia-Rodriguez, F., Mekonen, T., Melese, A., Melku, M., Memiah, P.T.N.,
564 Memish, Z.A., Mendoza, W., Mengistu, G., Mensah, G.A., Mereta, S.T., Meretoja, A.,
565 Meretoja, T.J., Mestrovic, T., Miazgowski, B., Miazgowski, T., Millear, A.I., Miller, T.R.,
566 Mini, G.K., Mirarefin, M., Mirica, A., Mirrakhimov, E.M., Misganaw, A.T., Mitchell, P.B.,
567 Mitiku, H., Moazen, B., Mohajer, B., Mohammad, K.A., Mohammadi, M., Mohammadifard,
568 N., Mohammadnia-Afrouzi, M., Mohammed, M.A., Mohammed, S., Mohebi, F., Mokdad,
569 A.H., Molokhia, M., Monasta, L., Montañez, J.C., Moosazadeh, M., Moradi, G., Moradi, M.,
570 Moradi-Lakeh, M., Moradinazar, M., Moraga, P., Morawska, L., Moreno Velásquez, I.,
571 Morgado-Da-Costa, J., Morrison, S.D., Moschos, M.M., Mousavi, S.M., Mruts, K.B., Muche,
572 A.A., Muchie, K.F., Mueller, U.O., Muhammed, O.S., Mukhopadhyay, S., Muller, K.,
573 Mumford, J.E., Murthy, G.V.S., Musa, K.I., Mustafa, G., Nabhan, A.F., Nagata, C., Nagel, G.,
574 Naghavi, M., Naheed, A., Nahvijou, A., Naik, G., Najafi, F., Nam, H.S., Nangia, V., Nansseu,
575 J.R., Neamati, N., Negoi, I., Negoi, R.I., Neupane, S., Newton, C.R.J., Ngunjiri, J.W., Nguyen,
576 A.Q., Nguyen, G., Nguyen, H.T., Nguyen, H.L.T., Nguyen, H.T., Nguyen, L.H., Nguyen, M.,
577 Nguyen, N.B., Nguyen, S.H., Nichols, E., Ningrum, D.N.A., Nixon, M.R., Nomura, S.,
578 Noroozi, M., Norrving, B., Noubiap, J.J., Nouri, H.R., Shiadeh, M.N., Nowroozi, M.R.,
579 Nsoesie, E.O., Nyasulu, P.S., Odell, C.M., Ofori-Asenso, R., Ogbo, F.A., Oh, I.-H., Oladimeji,
580 O., Olagunju, A.T., Olagunju, T.O., Olivares, P.R., Olsen, H.E., Olusanya, B.O., Olusanya,
581 J.O., Ong, K.L., Ong, S.K., Oren, E., Ortiz, A., Ota, E., Otstavnov, S.S., Øverland, S.,
582 Owolabi, M.O., P A, M., Pacella, R., Pakhare, A.P., Pakpour, A.H., Pana, A., Panda-Jonas, S.,
583 Park, E.-K., Park, J., Parry, C.D.H., Parsian, H., Pasdar, Y., Patel, S., Patil, S.T., Patle, A.,
584 Patton, G.C., Paturi, V.R., Paudel, D., Paulson, K.R., Pearce, N., Pereira, A., Pereira, D.M.,
585 Perico, N., Pesudovs, K., Petzold, M., Pham, H.Q., Phillips, M.R., Pigott, D.M., Pillay, J.D.,
586 Piradov, M.A., Pirsaheb, M., Pishgar, F., Plana-Ripoll, O., Polinder, S., Popova, S., Postma,

587 M.J., Pourshams, A., Poustchi, H., Prabhakaran, D., Prakash, S., Prakash, V., Prasad, N.,
588 Purcell, C.A., Qorbani, M., Quistberg, D.A., Radfar, A., Rafay, A., Rafiei, A., Rahim, F.,
589 Rahimi, K., Rahimi, Z., Rahimi-Movaghar, A., Rahimi-Movaghar, V., Rahman, M., Rahman,
590 M.H.U., Rahman, M.A., Rahman, S.U., Rai, R.K., Rajati, F., Ranjan, P., Rao, P.C., Rasella,
591 D., Rawaf, D.L., Rawaf, S., Reddy, K.S., Reiner, R.C., Reitsma, M.B., Remuzzi, G., Renzaho,
592 A.M.N., Resnikoff, S., Rezaei, S., Rezai, M.S., Ribeiro, A.L.P., Roberts, N.L.S., Robinson,
593 S.R., Roever, L., Ronfani, L., Roshandel, G., Rostami, A., Roth, G.A., Rothenbacher, D.,
594 Rubagotti, E., Sachdev, P.S., Sadat, N., Sadeghi, E., Saeedi Moghaddam, S., Safari, H., Safari,
595 Y., Safari-Faramani, R., Safdarian, M., Safi, S., Safiri, S., Sagar, R., Sahebkar, A., Sahraian,
596 M.A., Sajadi, H.S., Salam, N., Salama, J.S., Salamati, P., Saleem, Z., Salimi, Y., Salimzadeh,
597 H., Salomon, J.A., Salvi, S.S., Salz, I., Samy, A.M., Sanabria, J., Sanchez-Niño, M.D.,
598 Santomauro, D.F., Santos, I.S., Santos, J.V., Santric Milicevic, M.M., Sao Jose, B.P., Sardana,
599 M., Sarker, A.R., Sarmiento-Suárez, R., Sarrafzadegan, N., Sartorius, B., Sarvi, S., Sathian, B.,
600 Satpathy, M., Sawant, A.R., Sawhney, M., Saxena, S., Schaeffner, E., Schmidt, M.I.,
601 Schneider, I.J.C., Schutte, A.E., Schwebel, D.C., Schwendicke, F., Scott, J.G., Sekerija, M.,
602 Sepanlou, S.G., Serván-Mori, E., Seyedmousavi, S., Shabaninejad, H., Shafieesabet, A.,
603 Shahbazi, M., Shaheen, A.A., Shaikh, M.A., Shams-Beyranvand, M., Shamsi, M., Sharafi, H.,
604 Sharafi, K., Sharif, M., Sharif-Alhoseini, M., Sharma, J., Sharma, R., She, J., Sheikh, A., Shi,
605 P., Shibuya, K., Shiferaw, M.S., Shigematsu, M., Shiri, R., Shirkoohi, R., Shiue, I.,
606 Shokoohinia, Y., Shokraneh, F., Shoman, H., Shrime, M.G., Si, S., Siabani, S., Sibai, A.M.,
607 Siddiqi, T.J., Sigfusdottir, I.D., Sigurvinsdottir, R., Silva, D.A.S., Silva, J.P., Silveira, D.G.A.,
608 Singam, N.S.V., Singh, J.A., Singh, N.P., Singh, V., Sinha, D.N., Skiadaresi, E., Skirbekk, V.,
609 Sliwa, K., Smith, D.L., Smith, M., Soares Filho, A.M., Sobaih, B.H., Sobhani, S., Soofi, M.,
610 Sorensen, R.J.D., Soriano, J.B., Soyiri, I.N., Sposato, L.A., Sreeramareddy, C.T., Srinivasan,

611 V., Stanaway, J.D., Starodubov, V.I., Stein, D.J., Steiner, C., Steiner, T.J., Stokes, M.A.,
612 Stovner, L.J., Subart, M.L., Sudaryanto, A., Sufiyan, M.B., Sulo, G., Sunguya, B.F., Sur, P.J.,
613 Sykes, B.L., Sylaja, P.N., Sylte, D.O., Szoeke, C.E.I., Tabarés-Seisdedos, R., Tabuchi, T.,
614 Tadakamadla, S.K., Tandon, N., Tassew, S.G., Tavakkoli, M., Taveira, N., Taylor, H.R.,
615 Tehrani-Banihashemi, A., Tekalign, T.G., Tekelemedhin, S.W., Tekle, M.G., Temsah, M.-H.,
616 Temsah, O., Terkawi, A.S., Tessema, B., Teweldemedhin, M., Thankappan, K.R., Theis, A.,
617 Thirunavukkarasu, S., Thomas, N., Tilahun, B., To, Q.G., Tonelli, M., Topor-Madry, R.,
618 Torre, A.E., Tortajada-Girbés, M., Touvier, M., Tovani-Palone, M.R., Towbin, J.A., Tran,
619 B.X., Tran, K.B., Troeger, C.E., Tsadik, A.G., Tsoi, D., Tudor Car, L., Tyrovolas, S., Ukwaja,
620 K.N., Ullah, I., Undurraga, E.A., Updike, R.L., Usman, M.S., Uthman, O.A., Vaduganathan,
621 M., Vaezi, A., Valdez, P.R., Varavikova, E., Varughese, S., Vasankari, T.J.,
622 Venketasubramanian, N., Villafaina, S., Violante, F.S., Vladimirov, S.K., Vlassov, V., Vollset,
623 S.E., Vos, T., Vosoughi, K., Vujcic, I.S., Wagnew, F.S., Waheed, Y., Wang, Y., Wang, Y.-P.,
624 Weiderpass, E., Weintraub, R.G., Weiss, D.J., Weldegebreal, F., Weldegwergs, K.G.,
625 Werdecker, A., West, T.E., Westerman, R., Whiteford, H.A., Widecka, J., Wijeratne, T.,
626 Williams, H.C., Wilner, L.B., Wilson, S., Winkler, A.S., Wiyeh, A.B., Wiysonge, C.S., Wolfe,
627 C.D.A., Woolf, A.D., Wyper, G.M.A., Xavier, D., Xu, G., Yadgir, S., Yahyazadeh Jabbari,
628 S.H., Yamada, T., Yan, L.L., Yano, Y., Yaseri, M., Yasin, Y.J., Yeshaneh, A., Yimer, E.M.,
629 Yip, P., Yisma, E., Yonemoto, N., Yoon, S.-J., Yotebieng, M., Younis, M.Z., Yousefifard, M.,
630 Yu, C., Zadnik, V., Zaidi, Z., Zaman, S. Bin, Zamani, M., Zandian, H., Zar, H.J., Zenebe,
631 Z.M., Zhou, M., Zipkin, B., Zodpey, S., Zucker, I., Zuhlke, L.J., Murray, C.J.L., 2018. Global,
632 regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and
633 healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic
634 analysis for the Global Burden of Disease Study 2017. Lancet. https://doi.org/10.1016/S0140-

635 6736(18)32335-3

636 Letunic, I., Bork, P., 2021. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree
637 display and annotation. Nucleic Acids Res. 49, W293–W296.
638 https://doi.org/10.1093/nar/gkab301

639 Omasits, U., Ahrens, C.H., Müller, S., Wollscheid, B., 2014. Protter: interactive protein feature
640 visualization and integration with experimental proteomic data. Bioinformatics 30, 884–886.
641 https://doi.org/10.1093/bioinformatics/btt607

642 Palmieri, N., Shrestha, A., Ruttkowski, B., Beck, T., Vogl, C., Tomley, F., Blake, D.P., Joachim,
643 A., 2017. The genome of the protozoan parasite *Cystoisospora suis* and a reverse vaccinology
644 approach to identify vaccine candidates. Int. J. Parasitol. 47, 189–202.
645 https://doi.org/10.1016/j.ijpara.2016.11.007

646 R Core Team, 2020. R: A Language and Environment for Statistical Computing.

647 RStudio Team, 2019. RStudio: Integrated Development Environment for R.

648 Salazar Garcés, L.F., Santiago, L.F., Santos, S.P. de O., Jaramillo Hernández, D.A., da Silva, M.B.,
649 Alves, V. dos S., Silveira, E.F., Barrouin-Melo, S.M., Cooper, P.J., Pacheco, L.G.C., Pinheiro,
650 C. da S., Alcantara-Neves, N.M., 2020. Immunogenicity and protection induced by
651 recombinant Toxocara canis proteins in a murine model of toxocariasis. Vaccine 38, 4762–
652 4772. https://doi.org/10.1016/j.vaccine.2020.04.072

653 Tsuji, N., Miyoshi, T., Islam, M.K., Isobe, T., Yoshihara, S., Arakawa, T., Matsumoto, Y.,
654 Yokomizo, Y., 2004. Recombinant *Ascaris* 16-Kilodalton Protein-Induced Protection against

Ascaris suum Larval Migration after Intranasal Vaccination in Pigs. J. Infect. Dis. 190, 1812–1820. https://doi.org/10.1086/425074

Tsuji, N., Suzuki, K., Kasuga-Aoki, H., Matsumoto, Y., Arakawa, T., Ishiwata, K., Isobe, T., 2001. Intranasal immunization with recombinant *Ascaris suum* 14-kilodalton antigen coupled with cholera toxin B subunit induces protective immunity to *A. suum* infection in mice. Infect. Immun. 69, 7285–7292. https://doi.org/10.1128/IAI.69.12.7285-7292.2001

Versteeg, L., Wei, J., Liu, Z., Keegan, B., Fujiwara, R.T., Jones, K.M., Asojo, O., Strych, U., Bottazzi, M.E., Hotez, P.J., Zhan, B., 2020. Protective immunity elicited by the nematode-conserved As37 recombinant protein against Ascaris suum infection. PLoS Negl. Trop. Dis. 14, e0008057.

Wang, J., Gao, S., Mostovoy, Y., Kang, Y., Zagoskin, M., Sun, Y., Zhang, B., White, L.K., Easton, A., Nutman, T.B., Kwok, P.-Y., Hu, S., Nielsen, M.K., Davis, R.E., 2017. Comparative genome analysis of programmed DNA elimination in nematodes. Genome Res. 27, 2001—2014. https://doi.org/10.1101/gr.225730.117

Wang, J., Mitreva, M., Berriman, M., Thorne, A., Magrini, V., Koutsovoulos, G., Kumar, S., Blaxter, M.L., Davis, R.E., 2012. Silencing of Germline-Expressed Genes by DNA Elimination in Somatic Cells. Dev. Cell 23, 1072–1080. https://doi.org/https://doi.org/10.1016/j.devcel.2012.09.020

Wei, T., Simko, V., 2017. R package “corrplot”: Visualization of a Correlation Matrix.

World Health Organization, 2020. Ending the neglect to attain the Sustainable Development Goals – A road map for neglected tropical diseases 2021–2030.

676 Zawawi, A., Else, K.J., 2020. Soil-Transmitted Helminth Vaccines: Are We Getting Closer? Front.
677 Immunol.

678 Zawawi, A., Forman, R., Smith, H., Mair, I., Jibril, M., Albaqshi, M.H., Brass, A., Derrick, J.P.,
679 Else, K.J., 2020. In silico design of a T-cell epitope vaccine candidate for parasitic helminth
680 infection. PLOS Pathog. 16, e1008243.

681

Figure and table captions

Figure 1 – Vacceed assessment. A. Vacceed final scores and respective number of proteins for all analysed protein sequences. B. Pearson correlation plot of the different Vacceed assessments. Each value corresponds to the Pearson correlation coefficient between all the protein sequence scores in two different runs. (All tools = Vacceed with all the used tools: No DeepLoc = Vacceed without the tool DeepLoc; the other labels follow the same principle).

Figure 2 – 2D representation of the topology of predicted vaccination targets using Phobius predictions. The horizontal bar in the middle represents the membrane, above it is the extracellular domain, and below is the intracellular domain. Highlighted are the Cd4+ Th epitopes chosen to be incorporated in future vaccination assays. A. "*AgB13X_g094_t05*" as the best expressed protein transcript for ATtype. B. "*AgR007_g063_t01*" as the best expressed transcript for APiezo. C. "*AgR007_g282_t05*" as the best expressed protein transcript for ALtype. D. "*AgB13X_g096_t06*" as the best expressed protein transcript for AProto.

Figure 3 – Workflow and summary diagram of the reverse vaccinology approach used in this study to identify and select potential vaccine targets in *Ascaris lumbricoides* and *A. suum*.

Figure 4 – Maximum likelihood phylogenies using the predicted vaccine targets and orthologues found in other nematodes. The trees were inferred using bootstrap values with 500 replicates. Values on the nodes represent the percentage of bootstrap support values and nodes without bootstrap values were supported by 100% of the 500 replicates. Each sequence is identified by the nematode species followed by the respective BioProject (proteins transcripts used in these analyses are individualized in Supplementary Table 4). A. Phylogenetic tree of ATtype orthologues. B. Phylogenetic tree of APiezo orthologues. C. Phylogenetic tree of ALtype orthologues. D. Phylogenetic tree of AProto orthologues.

Table 1 - Proteins and respective epitopes identified as potential vaccination targets against *Ascaris lumbricoides* and *Ascaris suum*. Amino acid, aa; *Ascaris suum*, *A. suum*; *Ascaris lumbricoides*, *A lumbricoides*.

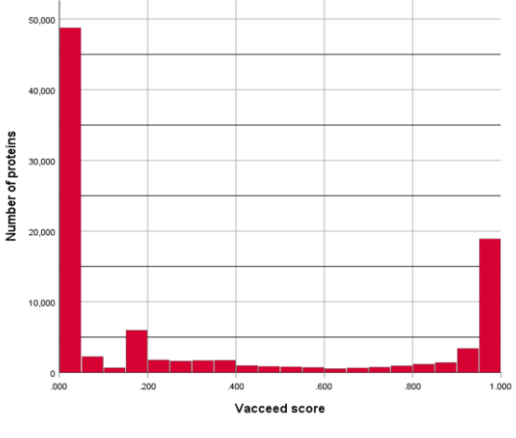
Supplementary Table 1 – The Vacceed final scores for each proteome in the three *Ascaris* proteomes.

Supplementary Table 2 - Protein sequences predicted to have epitopes that bind to all 27 MHC-II alleles, used as a reference set, with a score between 0.01 and 1 in the MHCII-IEDB tool.

Supplementary Table 3 - The epitopes found for each predicted vaccine target using MHCII-IEDB that scored between 0 and 1 and the MHC-II alleles they were predicted to bind to (only epitopes that were predicted to bind to two or more MHC-II alleles were retrieved).

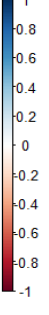
Supplementary Table 4 – List of orthologues used in the phylogenetic analysis.

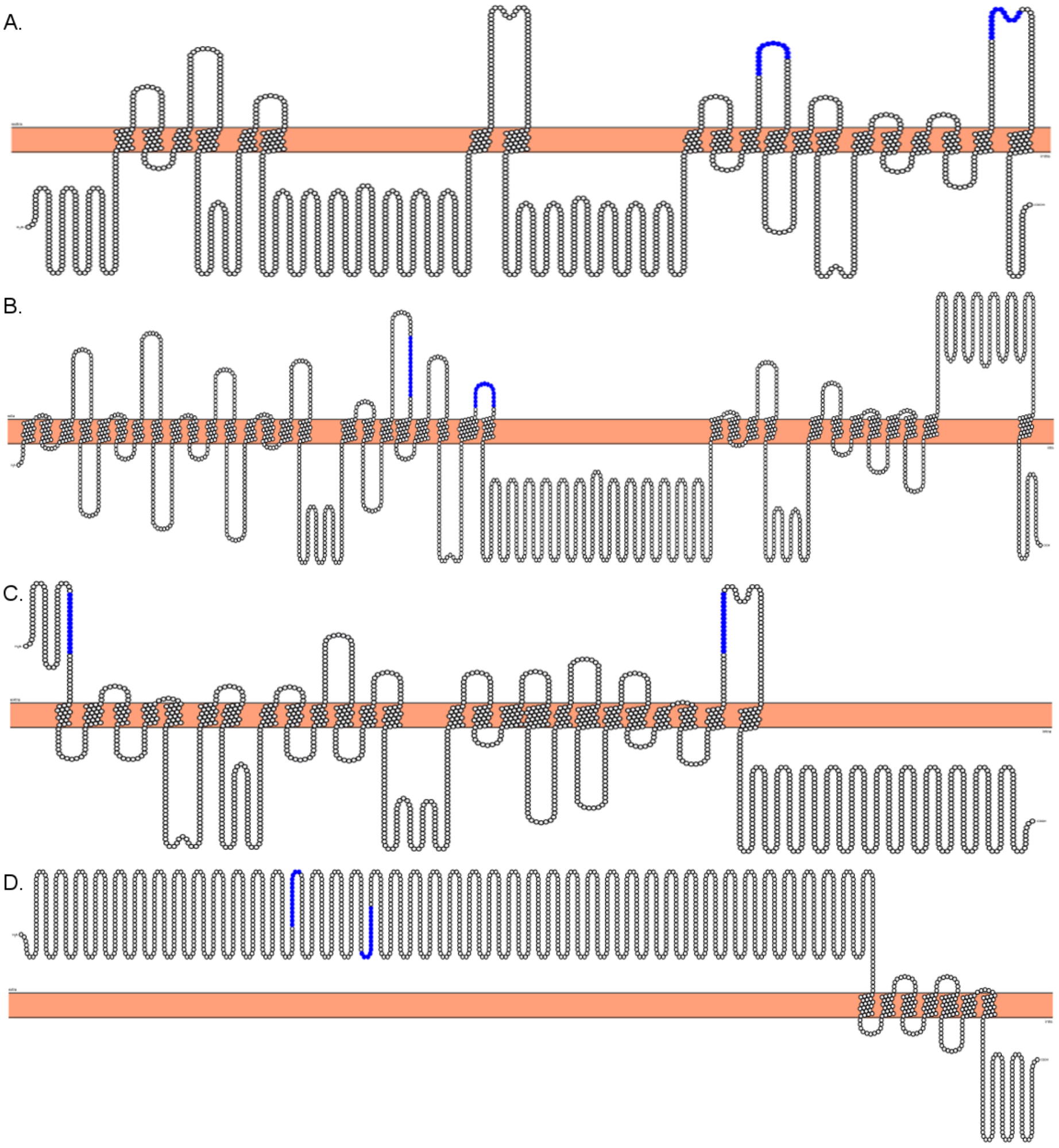
A.

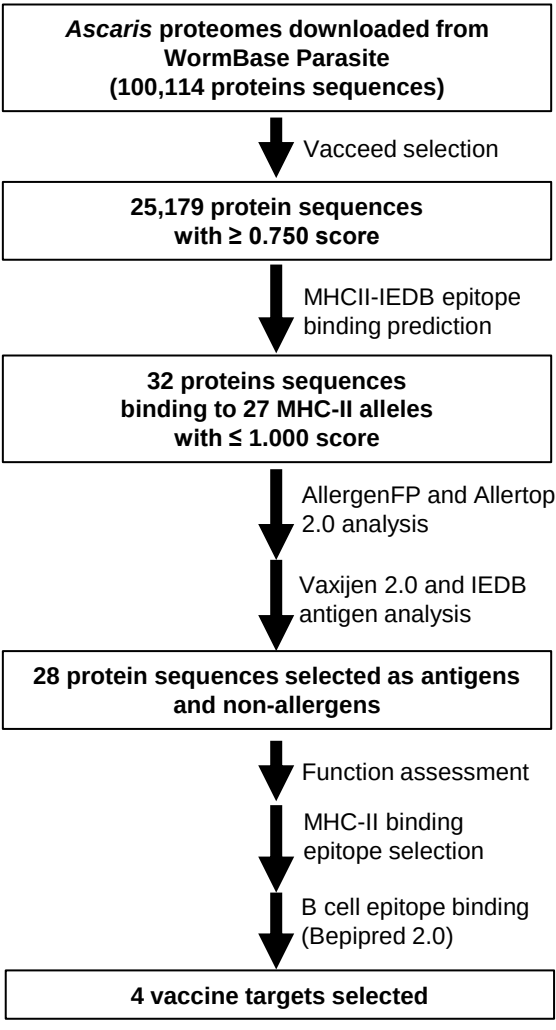


B.

	All tools	No DeepLoc	No Phobius	No SignalP	No TargetP	No TMHMM	No Wolf Psort
All tools	1	0.99	1	1	1	0.97	0.98
No DeepLoc		1	0.99	0.99	0.99	0.96	0.98
No Phobius			1	1	0.99	0.96	0.98
No SignalP				1	0.99	0.97	0.98
No TargetP					1	0.97	0.98
No TMHMM						1	0.95
No Wolf Psort							1




A.
B.
C.
D.


Ascaris proteomes downloaded from WormBase Parasite (100,114 proteins sequences)
Vacceed selection
25,179 protein sequences with ≥ 0.750 score
MHCII-IEDB epitope binding prediction
32 proteins sequences binding to 27 MHC-II alleles with ≤ 1.000 score
AllergenFP and Allertop 2.0 analysis
Vaxijen 2.0 and IEDB antigen analysis
28 protein sequences selected as antigens and non-allergens
Function assessment
MHC-II binding epitope selection
B cell epitope binding (Bepipred 2.0)
4 vaccine targets selected

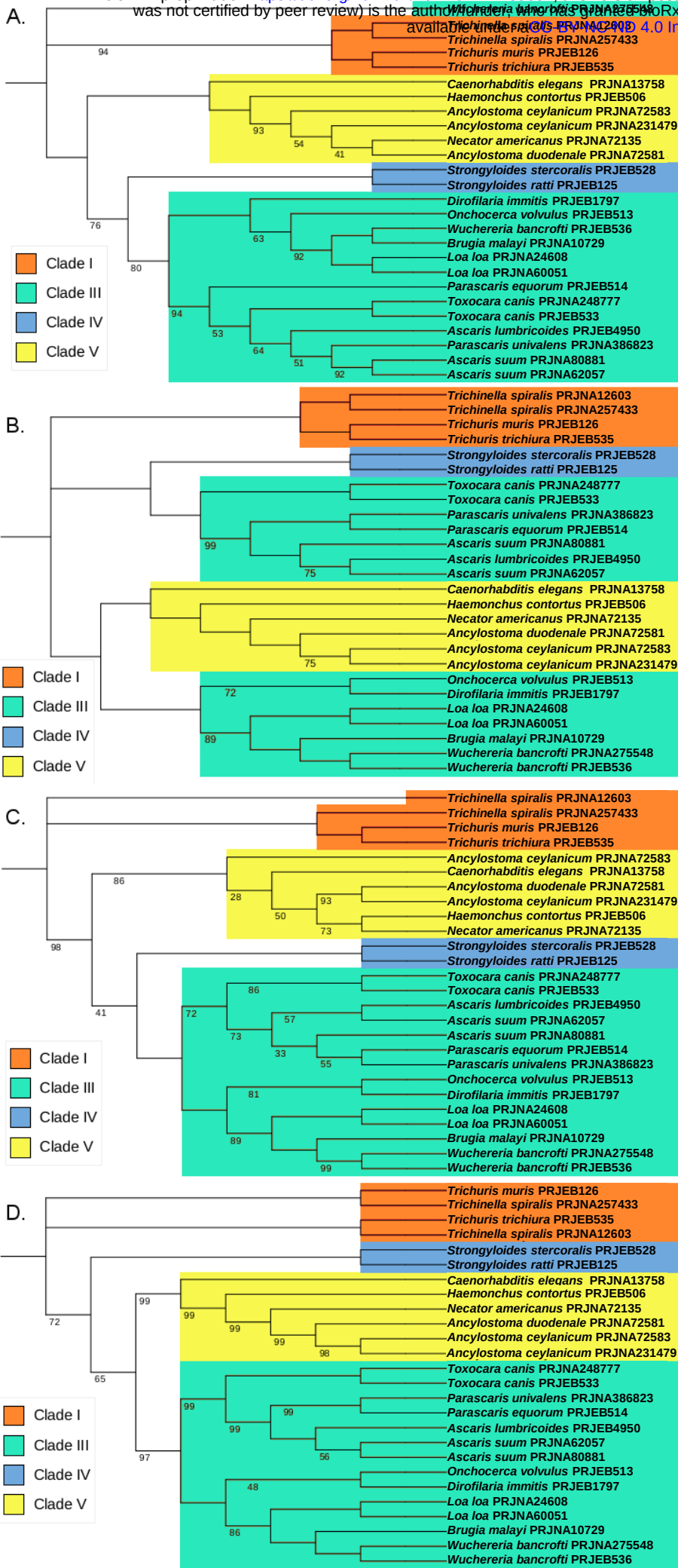


Table 1 – Proteins and respective epitopes identified as potential vaccination targets against *Ascaris lumbricoides* and *Ascaris suum*. Amino acid, aa; *Ascaris suum, A. suum*; *Ascaris lumbricoides, A lumbricoides*.

Protein function	**Predicted Protein Family**	**Species**	**WormBase Protein transcript identifier** [a]	**Selected CD4+ Th cell binding epitopes (aa)**	**Selected B-cell epitopes (aa)**
Membrane transporter	Piezo Family	*A. suum*	*GS_03113*	NCLKYFANFFFYRFG	LLSVHLKNDDDSIEST
			AgR007_g063_t01		
			AgR007_g063_t02	SLFLRPMRVALALLN	VDPSFDPVIPKEEVI
			AgR007_g063_t03		
			AgR007_g063_t04		
			AgR007_g063_t06		
		A. lumbricoides	*ALUE_0000666901*		
	Voltage-dependent calcium channel, T-type, alpha-1 subunit	*A. suum*	*GS_24322*	LRLLRALRPLRVINR	DATGVDMQPVENYN
			AgB13X_g094_t03		
			AgB13X_g094_t04	FKNFGMAFLTLFRIA	SIPPKSVER
			AgB13X_g094_t05		

		A. lumbricoides	*ALUE_0000418301*		
	Voltage-dependent calcium channel, L-type, alpha-1 subunit	*A. suum*	*GS_04697*	NNNFHTFPAAILVLF	ALNDETHIHRNNN
			AgR007_g282_t01		
			AgR007_g282_t02	ERSLLCLTLSNPLRK	SNEEDRGPVYNA
			AgR007_g282_t14		
		A. lumbricoides	*ALUE_0001482301*		
Cell adhesion	No predicted Family. Protocadherin-like proteins	*A. suum*	*GS_06422*	HTFRRFITAISLLDR	LSQSDHHILPRFANFVD
			AgB13X_g096_t02		DRTESLRSVTIQLLCSLR
		A. lumbricoides	*ALUE_0000418601*	NQEGVVHILSRKIFD	
					RQQFTLTFPYFSDGKFK

a. https://parasite.wormbase.org/index.html