International Journal for Parasitology 47 (2017) 529–544

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

Helminth secretomes reflect different lifestyles and parasitized hosts

Yesid Cuesta-Astroz [a,c], Francislon Silva de Oliveira [a,c], Laila Alves Nahum [a,d], Guilherme Oliveira [a,b,*]

[a] Centro de Pesquisas René Rachou (CPqRR), Fundação Oswaldo Cruz (FIOCRUZ), Belo Horizonte, MG 30190-002, Brazil
[b] Instituto Tecnológico Vale, Belém, PA 66055-090, Brazil
[c] Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG 31270-901, Brazil
[d] Faculdade Promove de Tecnologia, Belo Horizonte, MG 30130-180, Brazil

ARTICLE INFO

Article history:
Received 3 November 2016
Received in revised form 20 January 2017
Accepted 24 January 2017
Available online 21 March 2017

Keywords:
Secretome
Protein evolution
Biodiversity
Host-parasite interaction
Computational biology
Helminths

ABSTRACT

Helminths cause a number of medical and agricultural problems and are a major cause of parasitic infections in humans, animals and plants. Comparative analysis of helminth genes and genomes are important to understand the genomic biodiversity and evolution of parasites and their hosts in terms of different selective pressures in their habitats. The interactions between the infective organisms and their hosts are mediated in large part by secreted proteins, known collectively as the "secretome". Proteins secreted by parasites are able to modify a host's environment and modulate their immune system. The composition and function of this set of proteins varies depending on the ecology, lifestyle and environment of an organism. The present study aimed to predict, *in silico*, the secretome in 44 helminth species including Nematoda (31 species) and Platyhelminthes (13 species) and, understand the diversity and evolution of secretomes. Secretomes from plant helminths range from 7.6% (943 proteins) to 13.9% (2,077 proteins) of the filtered proteome with an average of 10.2% (1,412 proteins) and from free-living helminths range from 4.4% (870 proteins) to 13% (3,121 proteins) with an average of 9.8% (2,126 proteins), respectively, and thus are considerably larger secretomes in relation to animal helminth secretomes which range from 4.2% (431 proteins) to 11.8% (2,419 proteins) of the proteomes, with an average of 7.1% (804 proteins). Across 44 secretomes in different helminth species, we found five conserved domains: (i) PF00014 (Kunitz/Bovine pancreatic trypsin inhibitor domain), (ii) PF00046 (Homeobox domain), (iii) PF00188 (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins), (iv) PF00085 (Thioredoxin) and (v) PF07679 (Immunoglobulin I-set domain). Our results detected secreted proteins associated with invasion, infection, adhesion and immunoregulation processes as protease inhibitors and cytokines, among other functions. In summary, this study will contribute towards the understanding of host-parasite interactions and possibly identify new molecular targets for the treatment or diagnosis of helminthiases.

1. Introduction

Helminths (Nematoda and Platyhelminthes) have great medical, veterinary and agricultural relevance and have severe socio-economic impact. According to the World Health Organization (WHO), more than two billion people are infected and many more are at risk of contracting helminthiases, especially in developing countries (World Health Organisation, 2015). Diseases caused by helminths are extremely varied. They include anemia and malnutrition (caused by hookworms such *Ancylostoma ceylanicum*), river blindness (filarial nematode *Onchocerca volvulus*), lymphatic filariasis (filarial nematodes *Brugia malayi, Loa loa* and *Wuchereria bancrofti*), and impaired cognitive development (Hotez et al., 2008; Lustigman et al., 2012). In some cases, helminths can maintain a chronic infection in the host. Recorded cases include patients with more than 30 years of *Schistosoma mansoni* infection (Harris et al., 1984) and a record of 53 years of *Echinococcus granulosus* infection (Spruance, 1974).

Helminths are also responsible for considerable losses in agriculture (Robinson and Dalton, 2009). In plants, helminths cause damage to crops leading to huge economic losses. Some of the most damaging nematodes include plant endoparasites (root-knot nematodes) belonging to the genera *Meloidogyne* spp., *Heterodera* spp. and *Globodera* spp. (Mehta et al., 2008). For example, *Meloidogyne* spp. impact both the quantity and quality of harvest, causing an estimated US$80 billion in damages annually

* Corresponding author at: Instituto Tecnológico Vale, Belém, PA 66055-090, Brazil.
E-mail address: guilherme.oliveira@itv.org (G. Oliveira).

http://dx.doi.org/10.1016/j.ijpara.2017.01.007
0020-7519/

(Wasmuth et al., 2008). Helminth parasites of livestock such as cattle and sheep are the cause of severe economic losses worldwide, and billions of dollars are spent annually on treatment and control of nematodes (Wang et al., 2009).

The rationale in the search for anthelmintic targets before the use of genomics was based on the molecular study of individual genes associated with parasite virulence (Brindley et al., 2009). The genomes of different free living and parasitic helminth species, which affect humans and other organisms of agricultural and veterinary importance, are being sequenced (Brindley et al., 2009). Genomic, transcriptomic and proteomic data from helminths have provided important information for understanding molecular mechanisms involved in metabolism, parasite-host interaction, immune system evasion and molecular evolution, among other topics (Brindley et al., 2009; Lustigman et al., 2012; Nahum et al., 2012; Tsai et al., 2013; Cuesta-Astroz et al., 2014). Therefore, approaches that consider species biology in a global manner have been increasingly deployed (Oliveira and Pierce, 2015).

Secreted proteins play essential roles in organisms from bacteria to mammals (Tjalsma et al., 2000). Such proteins may represent between 8% and 20% of an organism's proteome (Greenbaum et al., 2001). Proteins included in the secretome belong to various functional classes such as cytokines, hormones, digestive enzymes, proteases, toxins, antimicrobial peptides and proteins associated with oxidative stress (Ranganathan and Garg, 2009). Some of those are involved in vital biological processes such as cell adhesion, cell migration, cell to cell communication, differentiation, proliferation, morphogenesis, and regulation of the immune response (Maizels and Yazdanbakhsh, 2003). Parasite secreted proteins not only play a role in the organisms that produce it, but also have been demonstrated to regulate the host immune response and to be the direct cause of pathology (Maizels and Yazdanbakhsh, 2003; Cass et al., 2007; Ferguson et al., 2015; Zhu et al., 2016).

The prediction of secreted proteins depends on the computational identification of specific signals. Protein secretion through the endoplasmic reticulum is associated with a hydrophobic signal peptide at the N-terminus portion, representing the classical secretory pathway (Lonsdale et al., 2016). Some secreted proteins do not contain a signal peptide in their sequence and are referred to as leaderless secretory proteins, being secreted by non-classical secretion pathways (Lonsdale et al., 2016). Non-classical pathways are predicted by tools such as SecretomeP, based on the idea that extracellular proteins share specific features at the sequence level, regardless of the mechanism by which they are secreted (Bendtsen et al., 2004, 2005). The six protein features used in SecretomeP are: number of atoms, number of positively charges residues, low complexity regions (Wootton and Federhen, 1996), sub-cellular localization (Nakai and Horton, 1999), transmembrane helices (Krogh et al., 2001) and pro-peptide prediction (Duckert et al., 2004). Proteins secreted by the non-classical pathways have been identified as glycolytic enzymes, chaperones and translation factors, among others, suggesting that these proteins may be multifunctional ("moonlighting" proteins) (Nombela et al., 2006). Microscopy studies have shown experimental evidence of extracellular vesicles in helminths, specifically in trematodes *Fasciola hepatica, Echinostoma caproni* (Marcilla et al., 2012), *Schistosoma japonicum* (Zhu et al., 2016) and *S. mansoni* (Sotillo et al., 2016). These vesicles are actively released by the parasites and are captured by the host cells, playing an important role in host-parasite interaction (Zhu et al., 2016).

Little is known about the diversity and evolution of secreted proteins in helminths. Processes shaping secretome diversity according to different niches and lifestyles, and specific features allowing parasite survival in different environments, remain open to investigation. The present study aimed at performing a comparative analysis of the predicted secretome across 44 species including free-living and parasitic Nematoda and Platyhelminthes in order address the following issues: (i) the conservation of the secretome among different helminth species; (ii) the diversity of the secreted repertoire on different secretomes; (iii) the correlation between the size and composition of the secretome with the species life style, host and phylogenetic lineage; and (iv) the presence of different protein domain features between secreted and non-secreted proteins.

2. Materials and methods

2.1. Organisms and sequence data

Predicted proteomes were retrieved from WormBase Parasite version 1.0 (http://parasite.wormbase.org) (Harris et al., 2014). The original dataset had 92 helminth proteomes distributed in 82 species. The genomes were filtered according to the scheme mentioned below. The final genome dataset was composed of 44 species (31 Nematoda and 13 Platyhelminthes) (Table 1). *Caenorhabditis elegans* data were retrieved from WormBase (release WS250) (https://www.wormbase.org). Data on *Opisthorchis viverrini* and *Gyrodactylus salaris* were retrieved from internet links provided in the original genome papers (Hahn et al., 2014; Young et al., 2014). Nematode species were divided according to DNA sequence studies that suggested the existence of five clades (Blaxter et al., 1998). For some closely related species, with the exception of *Schistosoma* spp., only one species per genus was included in the final dataset in order to minimize data duplication.

2.2. Data filtering

Sequences were scanned using a script to remove possible error sources and to validate those according to the following criteria: (i) starting with a methionine, (ii) having no internal stop codons, (iii) lacking ambiguous amino acids (aa) not represented in the 20 International Union of Pure and Applied Chemistry (IUPAC) amino acid codes, and (iv) longer than 100 aa, in order to remove possible annotation errors and to minimize the inclusion of non-real protein sequences (Rezende et al., 2012; Mendes et al., 2013). Proteomes with $\geq$65% of sequences retrieved after the filtering processes were included. In the case that more than one genome project was available for the same species, the one with the best values in the filtering process was included in our analyses.

2.3. Secretome prediction

The *in silico* prediction of secreted proteins was performed using different bioinformatics tools and databases (Fig. 1). SignalP 4.1 (Petersen et al., 2011) was used to identify classical secretory proteins. All proteins identified as not having a signal peptide were analyzed with SecretomeP 1.0 g optimized for eukaryotes (Bendtsen et al., 2004) to predict non-classical pathway secreted proteins. To limit false positive results, only records with a neural network (NN) score $\geq$0.9 were considered as secreted proteins. Proteins predicted to be secreted were subsequently scanned for the presence of mitochondrial sequences by TargetP (Emanuelsson et al., 2000) and transmembrane helices by the transmembrane identification based on hidden Markov model (TMHMM) method (Krogh et al., 2001).

This approach has been used for the prediction of soluble secreted proteins in helminths and arthropods described by other authors (Garg and Ranganathan, 2011; Schicht et al., 2013). For consistency, we predicted the secretome datasets using the same

Table 1
Results of the proteome filtering process and the prediction of secreted proteins in 44 helminth species. Total sequences correspond to the number of proteins in the predicted proteome of each species analysed. Valid sequences are the number of sequences after the filtering process; the value as a percentage in relation with the predicted proteome is given in parentheses.

Species	Total sequences	Valid sequences (%)	Total Secretome	Lifestyle
Nematoda				
Clade I				
Trichinella spiralis	16,380	13,617 (83.1)	1,146	AP
Trichuris suis	9,831	8,489 (86.3)	727	AP
Clade III				
Acanthocheilonema viteae	10,397	8,337 (80.1)	507	AP
Ascaris suum	18,542	14,893 (80.3)	1,158	AP
Brugia malayi	17,750	12,116 (68.2)	800	AP
Dirofilaria immitis	12,857	9,163 (71.2)	593	AP
Dracunculus medinensis	9,495	6,559 (69.1)	499	AP
Elaeophora elaphi	9,562	7,636 (79.8)	525	AP
Enterobius vermicularis	12,063	8,842 (73.3)	649	AP
Litomosoides sigmodontis	10,246	8,591 (83.8)	542	AP
Loa loa	15,445	12,204 (79.0)	803	AP
Onchocerca volvulus	12,534	9,314 (74.3)	682	AP
Syphacia muris	10,200	7,768 (76.1)	584	AP
Thelazia callipaeda	9,999	7,729 (77.3)	532	AP
Toxocara canis	16,571	10,668 (65)	858	AP
Wuchereria bancrofti	12,625	8,204 (65)	517	AP
Clade IV				
Bursaphelenchus xylophilus	17,704	14,948 (84.4)	2,077	PP
Globodera pallida	16,403	13,158 (80.2)	1,218	PP
Meloidogyne hapla	14,420	12,391 (86.0)	943	PP
Parastrongyloides trichosuri	14,957	12,913 (86.3)	1,334	AP
Rhabditophanes sp. *KR3021*	13,493	12,168 (90.1)	1,192	AP
Strongyloides ratti	12,430	11,489 (92.4)	973	AP
Clade V				
Ancylostoma ceylanicum	15,892	13,586 (85.4)	1,169	AP
Angiostrongylus costaricensis	9,989	6,956 (69.6)	509	AP
Dictyocaulus viviparus	13,514	11,510 (85.1)	825	AP
Haemonchus contortus	24,747	20,460 (82.6)	2,419	AP
Necator americanus	19,153	13,924 (72.7)	1,231	AP
Nippostrongylus brasiliensis	20,234	14,214 (70.2)	1,469	AP
Oesophagostomum dentatum	25,291	16,790 (66.3)	1,531	AP
Pristionchus pacificus	24,217	20,049 (82.7)	2,388	FL
Caenorhabditis elegans	26,018	24,002 (92.2)	3,121	FL
Platyhelminthes				
Cestoda				
Echinococcus multilocularis	10,189	9,145 (89.7)	524	AP
Hydatigera taeniaeformis	10,907	7,614 (69.8)	468	AP
Hymenolepis microstoma	10,077	9,066 (89.9)	454	AP
Mesocestoides corti	9,056	6,520 (72)	406	AP
Taenia solium	12,481	10,406 (83.3)	538	AP
Trematoda				
Clonorchis sinensis	13,634	11,947 (87.6)	656	AP
Fasciola hepatica	15,739	12,488 (79.3)	992	AP
Schistosoma haematobium	13,073	8,769 (67.0)	379	AP
Schistosoma japonicum	12,743	9,141 (71.7)	476	AP
Schistosoma mansoni	11,828	10,121 (85.5)	431	AP
Opisthorchis viverrini	16,379	12,117 (73.9)	832	AP
Monogenea				
Gyrodactylus salaris	15,488	11,148 (71.9)	653	AP
Turbellaria				
Schmidtea mediterranea	29,850	19,423 (65.0)	870	FL

AP, animal parasite; PP, plant parasite; FL, free-living.

methodology. All programs used in this study were linked using Perl and bash shell scripts. A MySQL database was created to store and retrieve information using queries.

2.4. Functional annotation

Putative secreted proteins were mapped to Gene Ontology (GO) terms and annotated using Blast2GO (Conesa et al., 2005), using default parameters (E-Value-Hit-Filter: 1.0E-6; Annotation cut-off: 55; GO weight: 5; Hsp-Hit Coverage cut-off: 0). We also determined GO terms that were enriched in a group of predicted secretomes compared with a reference group using Blast2GO (Conesa et al., 2005). This was introduced in Blast2GO by integrating Gossip (Blüthgen et al., 2005). Gossip computes Fisher's Exact Test, applying a robust False Discovery Rate (FDR) correction for multiple testing that returns a list of significant GO terms ranked by their corrected *P* values (FDR) (Conesa et al., 2005). In our analysis we applied, as a term filter mode, the corrected *P* value and as a term filter value, 0.05. Enriched GO terms were selected based on the criteria of having a corrected *P* value <0.05.

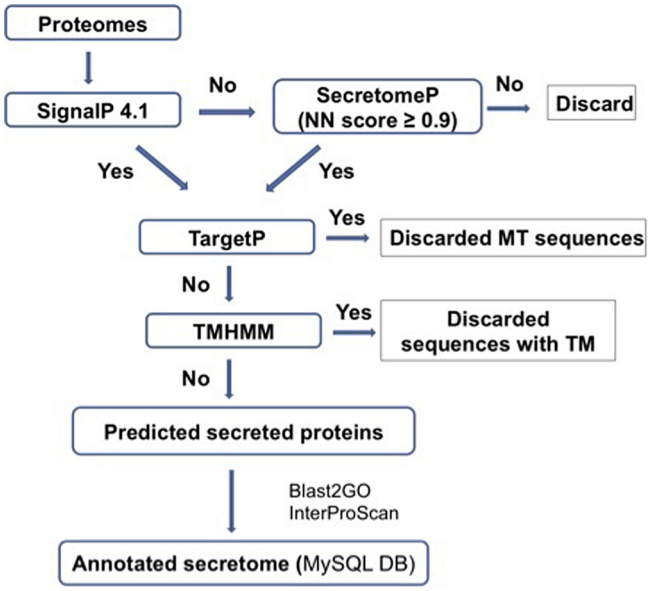

Fig. 1. Workflow for secretome prediction and annotation. SignalP 4.1 (Petersen et al., 2011) was used to identify signal peptide, which is the main feature of classical secretory proteins. Proteins lacking signal peptide were analysed using SecretomeP (Bendtsen et al., 2004), proteins with a neural network (NN) score ≥0.9 were considered secreted proteins by a non-classical pathway. Proteins were also scanned for the presence of mitochondrial sequences by TargetP (Emanuelsson et al., 2000) and transmembrane helices by a TMHMM model (Krogh et al., 2001). Predicted secreted proteins were annotated using Blast2GO (Conesa et al., 2005) and InterProScan (Jones et al., 2014). Final results were deposited in a local MySQL database (DB). MT, mitochondrial; TM, transmembrane.

Additionally, secreted proteins were associated with protein families, domains and functional sites through InterProScan v.5.0.7 (Jones et al., 2014) in a standalone version. InterProScan combines different protein signature recognition methods into one resource integrating the following databases: Coils, FPrintScan, Gene3D, HMM-Panther, HMM-PIR, HMM-Pfam, HMM-Smart, HMM-Tigr, Phobius, ProfileScan, Prosite, PatternScan and Superfamily. For the protein architecture analysis (presence and order in which domains are arranged within the protein sequence), we used the Pfam database (Finn et al., 2016). Pfam is a collection of manually curated families known as Pfam-A and a set of automatically generated families named Pfam-B. Pfam-A domains were considered in the present study. Additional functional information such as GO terms for Pfam domains were retrieved from the InterProScan results. These manual annotations are based on the function of particular domains rather than the function of domain families. These results were deposited in the local MySQL database in order to perform specific queries and retrieve information.

2.5. Ortholog predictions

To detect putative orthologs across predicted secretomes, we performed an OrthoMCL (http://orthomcl.org/orthomcl/) cluster analysis (Fischer et al., 2011) using the default settings (E-value cutoff: 1e-5 and identity: 50%). In this phase, the OrthoMCL maps proteins to groups in OrthoMCL-DB. It performs a BLASTP search against all the proteins in OrthoMCL-DB using a cutoff of 1e-5 and 50% match. Each protein is assigned to the group containing its best hit. If the best matching protein does not have a group, it is assigned to NO_GROUP. We propose a ratio called OG_diversity, which is result of number of orthologous groups/total secretome. This value reflects the diversity of orthologous groups (OGs) per species.

3. Results

3.1. Secretome size

Proteins containing a signal peptide, significant SecretomeP score (NN ≥ 0.9) and lacking a mitochondrial origin signal and transmembrane domains were considered as belonging to the soluble secretome (41,200 proteins), which is divided into proteins secreted by the signal peptide mediated classical pathway (31,192 proteins) and those secreted by the non-classical pathway (10,008 proteins). Sequences belonging to the secretomes in fasta format per species are available at https://figshare.com/s/6410c0479de5e1a6ece1. The secretome constituted on average 7.6% (936 proteins) of the filtered proteomes (valid sequences) in all of the 44 species. In Nematoda 8.4% of the proteins were secreted (1,081 proteins) and in Platyhelminthes 5.5% (590 proteins). The largest number of secreted proteins was observed in *Bursaphelenchus xylophilus* with 13.9% of the proteins (2,077 proteins), and the lowest in *S. mansoni* with 4.2% of the total proteome (431 proteins) (Fig. 2, Table 1).

When we analyzed the secretomes in a phylogenetic context, Clade I Nematoda secretomes had different sizes with 1,146 proteins for *Trichinella spiralis* and 727 proteins for *Trichuris suis*, corresponding to 8.4% and 8.5% of the filtered proteomes, respectively. Secretomes in Clade III covered on average 6.9% (660 proteins), Clade IV 9.8% (1,289 proteins), and Clade V 9.7% (1,629 proteins) of the filtered proteomes. Secretomes from Cestoda contained on average 5.6% (478 proteins), Trematoda 5.6% (627 proteins), Monogenea 5.8% (653 proteins), and Turbellaria 4.4% (870 proteins) of the filtered proteome. On average, in Platyhelminthes the sizes of the secretomes were smaller compared with Nematoda. Animal infecting helminthes also had smaller secretomes (804 proteins) in comparison to plant infecting helminthes (1,412 proteins). Across Platyhelminthes the relative sizes of the secretomes varied from 4.2% (431 proteins) to 7.9% (992 proteins) of the filtered proteome (Fig. 2).

According to the classification used by Krijger et al., (2014), the 44 species were grouped into three classes. Class 1 contained seven species with secretomes comprising less than 500 proteins (six of them are Platyhelminthes). Class 2 comprised 24 species with secretomes ranging from 500 to 1,100 proteins. Class 3 included the remaining 13 species with more than 1,100 proteins (this secretome class contained exclusively nematodes). In relation to the hosts, animal helminth secretomes ranged from 4.2% (431 proteins) to 11.8% (2,419 proteins) of the proteomes, with an average of 7.1% (804 proteins). Secretomes from plant helminths were considerably larger, ranging from 7.6% (943 proteins) to 13.9% (2,077 proteins) with an average of 10.2% (1,412 proteins). The secretome size of free-living helminths ranged from 4.4% (870 proteins) to 13% (3,121 proteins) with an average of 9.8% (2,126 proteins), being also considered a larger secretome compared with animal helminths secretomes.

For each secretome, the protein length distribution was analyzed using the number of proteins in defined length intervals (100–300, 301–500, 501–1,000 and ≥1,001 aa) as fractions of the total secretome. The complete secretome of 44 species (41,200 sequences) presented length distributions as shown in Supplementary Table S1. Fig. 3 presents the length distribution for each species. Most proteins in the secretomes contained between 100 and 300 aa (27,819 proteins) and proteins with more than 1,000 aa were present in smaller numbers across the secretomes (1,165 proteins) (Supplementary Table S1). However, species with a greater presence of these large proteins were found in the free-living helminths, *C. elegans* and *Pristionchus pacificus* (Fig. 3). Clade III species (*Dracunculus medinensis* and *W. bancrofti*) did not have proteins lar-

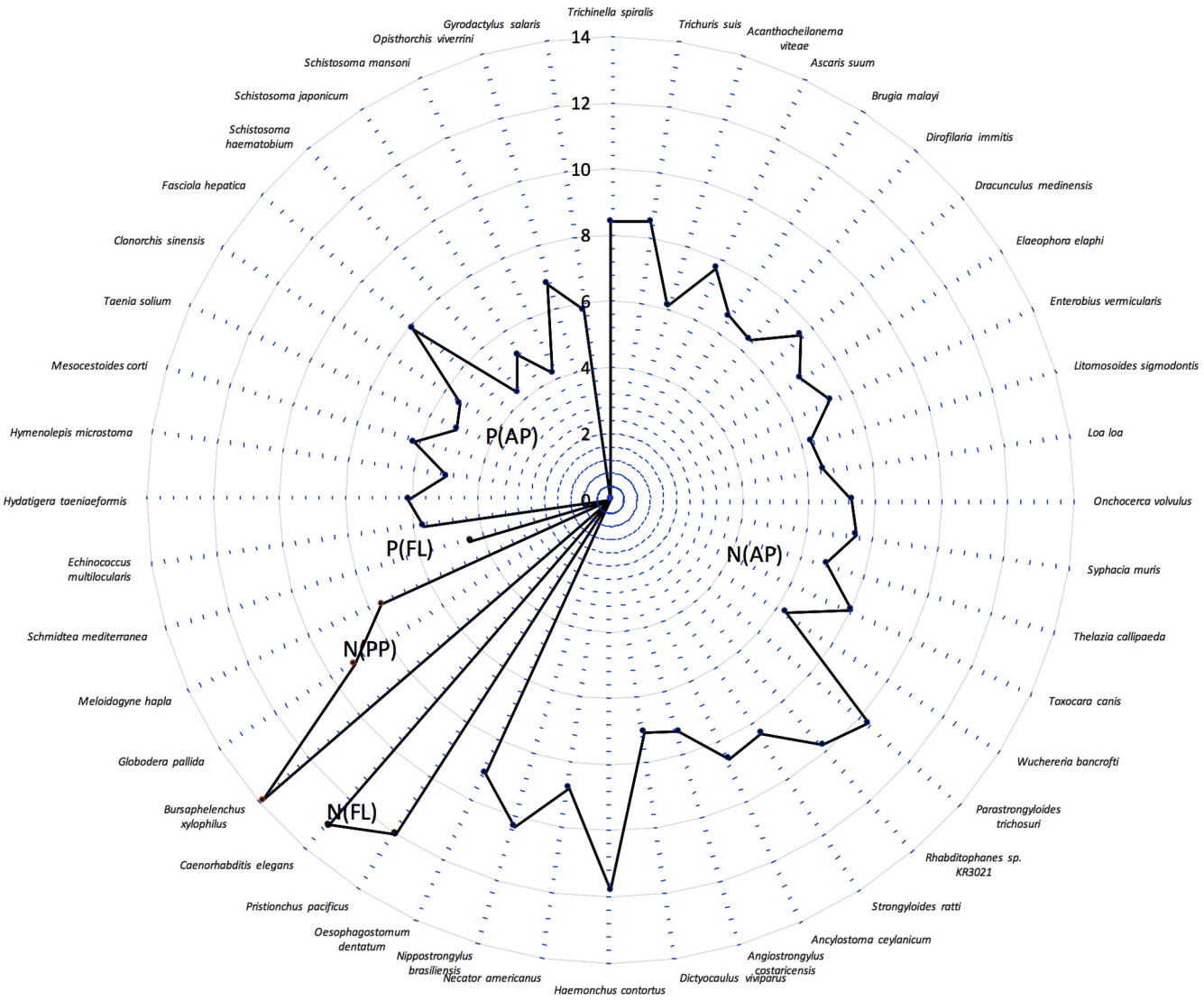

Fig. 2. Relative size distributions of secretomes (%) across 44 helminth species (see Table 1). Radar chart showing the secretome size in terms of the percentage in relation to the filtered proteomes per species of Nematoda (N) or Platyhelminthes (P). Different categories for interpreting comparisons are: AP, animal parasite; PP, plant parasite; FL, free-living parasite.

ger than 1,000 aa. In terms of the number of sequences that are part of the secretome, *C. elegans* had the largest secretome containing 3,121 proteins and the smallest was the *Schistosoma haematobium* secretome with 379 proteins (Fig. 3).

3.2. Protein domain diversity analysis

Out of 41,200 proteins (across 44 species), 20,607 (50%) proteins had at least one Pfam domain assigned. A total of 2,345 domains were identified across the 44 species and the occurrences per species were counted (Supplementary Table S2).

In order to have a global view of domain diversity and distribution, we calculated the most represented (top 25) domains across the secretomes (Table 2) and the domain occurrences per species (Supplementary Table S3). The Shk domain (Pfam PF01549) is a potassium channel inhibitor and was the most represented in the secretomes; 705 times (Table 2). The Shk domain is one of the most recurrent domains across protein architectures. Other well represented domains in secreted proteins were peptidase domains involved in hemoglobin degradation and nutrient uptake, redox processes, and cell to cell communication (Table 2).

In general, antioxidant molecules, proteases, and cell to cell communication domains were predicted in secretomes across helminth species (Supplementary Table S3). These included peroxiredoxin, thioredoxin, protein-disulfide isomerase, trypsins, lectins, cadherins and laminins, among others. Proteases such as metalloproteases degrade extracellular matrix proteins and may be involved in the degradation of plant and animal tissues. However, other domains could mediate the suppression of plant defense by degradation of host proteins involved in pathogen recognition or playing other important roles in defense (Krijger et al., 2014).

3.2.1. Common Pfam domains

Across 44 secretomes in different helminth species, we found five conserved domains: (i) PF00014 (Kunitz/Bovine pancreatic trypsin inhibitor domain), GO:0004867 – serine-type endopeptidase inhibitor activity. This domain prevents or reduces the activity of serine-type endopeptidases. (ii) PF00046 (Homeobox domain), GO:0003677 – DNA binding. This is a protein structural domain that binds DNA or RNA and is thus commonly found in transcription factors (Gehring, 1992). (iii) PF00188 (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins)

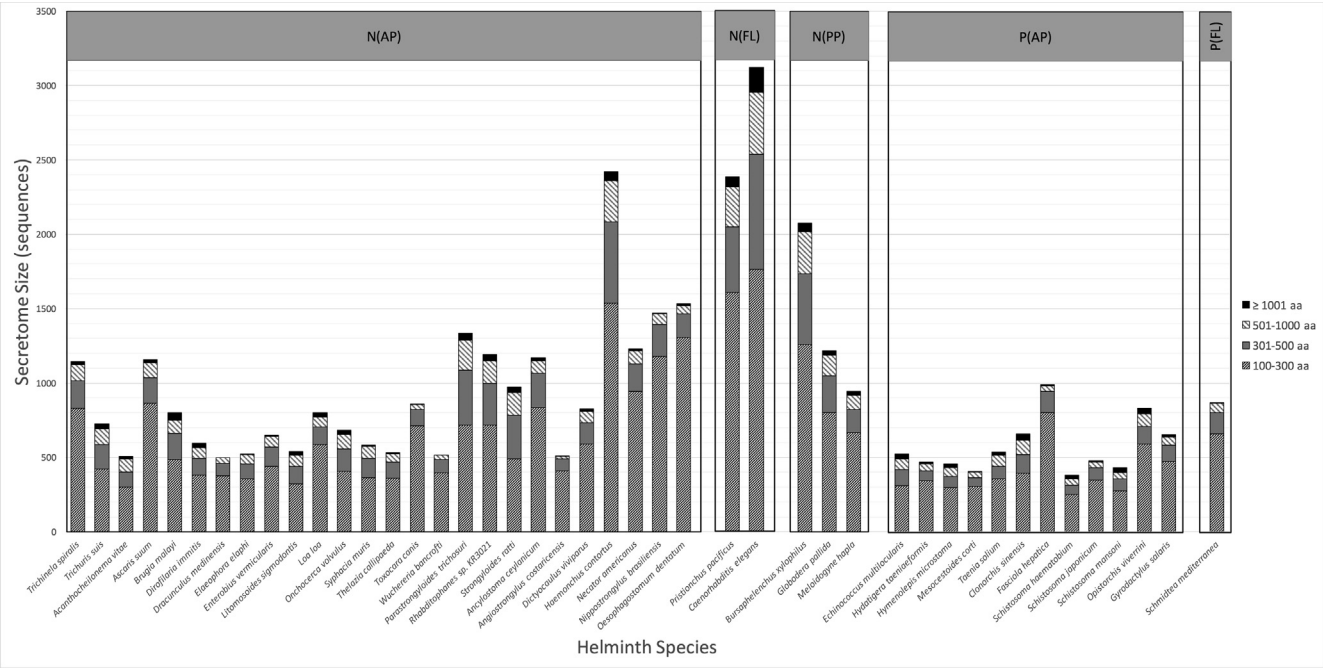

Fig. 3. Secretome size distribution (sequences) and protein size distribution within secretomes (number of amino acids (aa)) across 44 helminth species. Protein size in five length intervals: 101–300 aa, 301–500 aa, 501–1000 aa and ≥1001 aa. The size of each bars corresponds to the absolute size of the secretomes. Gray bar across the top indicates the different categories: N, Nematoda; P, Platyhelminthes; AP, animal parasite; PP, plant parasite; FL, free-living.

with no GO identifier available. Members of this domain (PF00188) are most often secreted and have extracellular functions involved in processes such as regulation of extracellular matrix, ion channel regulation, and in cell to cell adhesion, among other functions. (iv) PF00085 (Thioredoxin), GO:0045454 - cell redox homeostasis, acts as antioxidant by facilitating the reduction of other proteins by cysteine thiol-disulfide exchange. (v) PF07679 (Immunoglobulin I-set domain) - no GO identifier available. This domain is found in several cell adhesion molecules. These common domains are involved in universal functions across secretomes from diverse species.

3.2.2. Species-specific domains

Species-specific domains may represent particular helminth adaptations to specific niches and mechanisms implemented by helminths to establish infection and survival in the host (Garg and Ranganathan, 2012). In terms of exclusive domains, on average 22 domains were found to be species-exclusive. Among the species with the highest amount of exclusive domains were *F. hepatica* (71 domains) and *S. mediterranea* (57 domains). *Onchocerca volvulus* (seven domains) and *Thelazia callipaeda* (seven domains) had the lowest number of exclusive domains. The identification of these species-exclusive domains may enable the *in silico* selection of potential targets for antihelminthic agents. Supplementary Table S4 lists the species-specific domains across 44 species.

The ectoparasite *G. salaris* had the domain PF04203 (sortase) as an exclusive domain. According to the Pfam description, sortase refers to a group of enzymes that modify surface proteins by recognizing and cleaving a carboxyl-terminal signal. These proteins often play important roles in virulence, infection and colonization by pathogens (Proft and Baker, 2009). According to our results, domains involved in pili assembly in bacteria that appeared to be specific to *G. salaris* are PF03743 (bacterial conjugation TrbI-like protein), which influences the kinetics of pilus (Maneewannakul et al., 1992); PF06122 (conjugative relaxosome accessory transposon protein), which is involved in pili formation (Lawley et al., 2002), and PF06586 (TraK protein), which is known to be essential for pilus assembly, but its exact role in this process is unknown (Anthony et al., 1996). Another *G. salaris*-specific domain is PF02839 (carbohydrate binding domain). This short domain is found in many different glycosyl hydrolases and is structurally similar to the C-terminal chitin-binding domains (ChBD) of chitinase A1 and chitinase B (Hashimoto et al., 2000).

In the cyst nematode *Globodera pallida* that infects potatoes, the specific domain PF05630 (necrosis inducing protein NPP1) is related to necrosis inducing proteins from oomycetes, fungi and bacteria (Fellbrich et al., 2002). In *G. pallida*, this protein is involved in early stages of the infection and in callus formation (Gijzen and Nürnberger, 2006; Cotton et al., 2014). A domain associated with plant cell-wall hydrolysis, such as PF00553 (cellulose binding domain), was another domain involved in host-helminth interaction, more precisely in the plant penetration stage (Cotton et al., 2014). An unexpected domain also related to pathogenesis is PF07740 (spider toxin), which is a neurotoxin (Lampe et al., 1993). According our results, this domain was found to be specific to *Taenia solium*, but its function remains unclear.

Particular characteristics of the helminth life cycle are the need to penetrate the host and the chronicity of the interaction between it and the host (Craig and Scott, 2014). For these purposes, helminths have developed several strategies (Craig and Scott, 2014). In *T. suis*, the PF00151 domain (lipase) hydrolyses ester linkages of host triglycerides reflecting a particularity of its environment. This hydrolytic enzyme could also play a role in nematode penetration of the host by tissue disruption (Bahlool et al., 2013). To maintain a chronic infection, cell protection from oxidative damage by reactive oxygen species (ROS) is an important process (Dzik, 2006); a defense response domain PF00199 (catalase) was identified in the plant nematode *Meloidogyne hapla* as a specific domain.

3.2.3. Nematode-specific domains and GO terms enrichment

Only three domains were identified as nematode-specific. The domain PF01683 (EB module) has no known function and is found

Table 2
Top 25 most represented domains found in secreted proteins across 44 helminth species. Also shown are the number of proteins in which every domain appears and the number of occurrences of each domain in all of the secretomes.

Pfam ID	Pfam name	Description	Number of proteins	Domain occurrences	GO terms
PF01549	ShK	ShK is a powerful inhibitor of T lymphocyte voltage-gated potassium channels, in particular Kv1.3. Structural analogues may have use as an immunosuppressants for the treatment of autoimmune diseases	705	1678	–
PF01060	Transthyretn-like family	Apparently nematode-specific protein family.	665	691	GO:0005615 extracellular space
PF01400	Astacin (Peptidase family M12A)	Family of metallopeptidases	527	560	GO:0004222 metalloendopeptidase activity.GO:0006508 proteolysis
PF00188	CAP	CAP protein family (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins (CAP)) are found in a wide range of organisms, including prokaryotes and non-vertebrate eukaryotes.	518	576	–
PF00059	Lectin_C	A C-type lectin (CLEC) is a type of carbohydrate-binding protein domain.	472	574	GO:0030246 carbohydrate binding
PF00112	Peptidase_C1	Cysteine proteases are enzymes that degrade proteins.	389	420	GO:0006508 proteolysis GO:0008234 cysteine-type peptidase activity
PF00014	Kunitz_BPTI	Kunitz domains are the active domains of proteins that inhibit the function of protein degrading enzymes (protease inhibitors).	386	1280	GO:0004867 serine-type endopeptidase inhibitor activity
PF00046	Homeobox	Is a protein structural domain that binds DNA or RNA and is thus commonly found in transcription factors.	329	329	GO:0003677 DNA binding
PF00026	Asp	Aspartic proteases are a family of protease enzymes that use an aspartate residue for catalysis of their peptide substrates.	301	316	GO:0004190 aspartic-type endopeptidase GO:0006508 proteolysis
PF02520	DUF148	A domain of unknown function (DUF) is a protein domain that has no characterised function	281	288	–
PF04155	Ground-like	It has been proposed that the domain containing proteins may bind and modulate the activity of Patched-like membrane molecules, reminiscent of the modulating activities of neuropeptides	263	273	–
PF00085	Thioredoxin	Is a class of small redox, it plays a role in many important biological processes, including redox signaling.	242	410	GO:0045454 cell redox homeostasis
PF00135	COesterase	Carboxyl-esterases have been classified into three categories (A, B and C) on the basis of differential patterns of inhibition by organophosphates.	232	261	–
PF01682	DB	This domain has no known function	223	243	–
PF07679	I-set	Are found in several cell adhesion molecules, including vascular (VCAM), intercellular (ICAM), neural (NCAM) and mucosal addressin (MADCAM) cell adhesion molecules, as well as junction adhesion molecules (JAM).	223	973	–
PF00089	Trypsin	Trypsin (EC 3.4.21.4) is a serine protease from the PA clan superfamily	220	258	GO:0004252 serine-type endopeptidase GO:0006508 proteolysis
PF13499	EF-hand_7	Helix-loop-helix structural domain or motif found in a family of calcium-binding proteins.	183	257	GO:0005509 calcium ion binding
PF00431	CUB	Is a structural motif of approximately 110 residues found almost exclusively in extracellular and plasma membrane-associated proteins.	180	317	–
PF08246	Inhibitor_I29	Cathepsin propeptide inhibitor domain (I29). protease inhibitors are molecules that inhibit the function of proteases.	173	175	–
PF01764	Lipase_3	Triglyceride lipases are lipases that hydrolyse ester linkages of triglycerides.	167	175	GO:0006629 lipid metabolic process
PF00092	VWA	The von Willebrand factor is a large multimeric glycoprotein found in blood plasma.	162	220	–
PF00328	His_Phos_2	A phosphatase is an enzyme that removes a phosphate group from its substrate.	158	168	GO:0003993 acid phosphatase activity
PF00024	PAN_1	The domain is found in diverse proteins, in some they mediate protein–protein interactions, in others they mediate protein-carbohydrate interactions.	155	266	–
PF00069	Pkinase	The protein kinase domain is a structurally conserved protein domain containing the catalytic function of protein kinases	155	171	GO:0004672 protein kinase activity GO:0005524 ATP bindingGO:0006468 protein phosphorylation
PF00090	TSP_1	Thrombospondins (TSP) are secreted proteins with antiangiogenic abilities. Inhibiting the proliferation and migration of endothelial cells by interactions with CD36 expressed on their surface of these cells	148	447	–

GO, Gene Ontology.

associated with the kunitz domain (PF00014). The PF01682 (DB module) domain also has no known function and is found associated with Ig (PF00047) and fn3 (PF00041) domains, as well as with some lipases (PF00657). These domains reflect specific functions or biological processes and are accessory domains that work in synergy with other domains (Finn et al., 2016). PF01060 (transthyretin-like family) was another nematode-specific domain with unknown function.

We found interesting GO terms such as nematode larval development (GO:0002119; FDR = 3,89E-123) enriched in nematodes. This is a nematode-specific term and is related to the developmental progression of the nematode larva over time, from its formation

to the mature form. Our results indicated that secreted proteins involved in the molting process and molting signaling were included in these GO terms. Metallopeptidase activity (GO:0008237; FDR = 1,75E-20) was also a nematode-specific term that is related to proteins associated with tissue migration and hemoglobin degradation (Dzik, 2006). Defense response (GO:0006952; FDR = 4,45E-11), which is related to proteins that act in response to the presence of pathogens and proteins associated with antimicrobial peptide activity, is included in this term (Binns et al., 2009). Some defense response proteins have the PF15291 domain (Dermcidin, antibiotic peptide) and in *C. elegans* participate in protection against pathogenic Gram-positive bacteria (Amaral et al., 2012).

3.2.4. Platyhelminth-specific domains and GO terms enrichment

There were no Platyhelminthe-specific domains. However, we found enriched GO terms for this taxon. Here, we highlight some of those. Homophilic cell adhesion (GO:0007156; FDR = 2,93E-11) is related to proteins involved in cell to cell communication such as cadherins and laminins (Binns et al., 2009). These proteins are important for host recognition by the parasite and interactions between the host and parasite. (Rowe et al., 2009; Leontovyč et al., 2016). The platelet activation ontology term (GO:0030168; FDR = 3,99E-02) refers to a series of events triggered by the exposure of platelets to subendothelial tissue (Binns et al., 2009). These events could be related to helminth invasion and migration (Dzik, 2006). Leukocyte activation (GO:0045321; FDR = 2,92E-03) participates in the changes in morphology and behavior of leukocytes resulting from exposure to a specific antigen, cellular ligand or soluble factor (Binns et al., 2009). This term fits nicely with the proposed effect of secreted proteins and is clearly involved in helminth-host interaction (Dzik, 2006). Negative regulation of cellular communication (GO:0010648; FDR = 3,18E-02) is related to any process that decreases the frequency, rate or extent of cellular communication such as signaling, cell to cell attachment, extracellular matrix interaction, or between a cell and any other aspect of its environment (Binns et al., 2009).

Within Platyhelminthes we identified a trematode-specific domain PF08034 (trematode eggshell synthesis protein). This domain is present in the eggshell protein vitelline protein B1 (vpB1) (Pomaznoy et al., 2013). vpB1 is produced by mature vitelline cells to form the hard protective trematode eggshell and is crucial for eggshell synthesis in trematodes (Robinson et al., 2009). There were no Cestoda-specific domains. Monogenea and Turbellaria domains are species-specific for *G. salaris* and *S. mediterranea*, respectively.

3.2.5. Free-living species-specific domains and GO terms enrichment

There were no specific domains for free-living helminths, but enriched GO terms were identified and point to specific functions and biological processes associated with the free-living lifestyle. Response to gamma radiation (GO:0010332; FDR = 4,69E-03) was identified and is related to any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a gamma radiation stimulus (Binns et al., 2009). Another term related to radiation exposure was response to UV (GO:0009411; FDR = 2,31E-02), which is associated with UV stimulus. Some terms linked to microorganism defense were identified, such as defense response to Gram-negative bacteria (GO:0050829; FDR = 2,61E-02) and defense response to Gram-positive bacteria (GO:0050830; FDR = 2,71E-02). The term GO:0009410 (FDR = 3,09E-02) is associated with xenobiotic compound stimulus. Proteins involved in the redox process and antioxidants are also essential for the survival of free-living species (Dzik, 2006). Another enriched term was chitin catabolic process (GO:0006032; FDR = 4,17E-02) that results in the breakdown of chitin, which is an abundant protein in the free-living environment (Beier and Bertilsson, 2013).

In free-living nematodes, we found a specific glycoside hydrolase: PF00857 (isochorismatase family), PF05089 (glycoside hydrolase family 89) that includes enzymes with N-acetylglucosaminidase EC 3.2.1.50 activity, and PF12972 (glycoside hydrolase family 89). We identified 22 OGs in free-living nematodes (*C. elegans* and *P. pacificus*).

Schmidtea mediterranea, the only representative of free-living Platyhelminthes analyzed in this work, had 57 species-specific Pfam domains, being the helminth with the second largest number of specific domains. An interesting domain, PF03173 (putative carbohydrate binding domain), was found to be *S. mediterranea*-specific. This domain is involved in chitin degradation which is one of the most abundant polysaccharides on Earth (Tews et al., 1996). Its exclusivity could be related to specific aspects of the free -living life style of *S. mediterranea* (Tews et al., 1996; Beier and Bertilsson, 2013).

3.2.6. Plant-infecting helminth-specific domains and GO terms enrichment

Cell wall-degrading enzymes have no counterpart in most animals. Some examples of these key proteins are cellulases, xylanases, pectate lyases, and other members of the glycosyl hydrolase family (Dieterich and Sommer, 2009). We identified Pfam domains related to these important proteins in the plant-helminth interaction.

PF00295 (glycosyl hydrolases family 28) is an *M. hapla*-specific domain. This domain appears in plant bacterial pathogens such as *Erwinia carotovora* or *Ralstonia solanacearum* (*Pseudomonas solanacearum*), and fungal pathogens such as *Aspergillus niger*, and is involved in maceration and soft-rotting of plant tissue (Finn et al., 2016). Specific glycosyl hydrolases were identified in *B. xylophilus* such as PF02015 (glycosyl hydrolase family 45) that contains enzymes with only one known activity: endoglucanase (EC 3.2.1.4) and PF00722 (glycosyl hydrolase family 16), which contains enzymes with a number of known activities: lichenase (EC 3.2.1.73), xyloglucan xyloglucosyltransferase (EC 2.4.1.207), agarase (EC 3.2.1.81), kappa-carrageenase (EC 3.2.1.83).

Specific glycosyl hydrolase families were identified in *G. pallida,* such as PF04616 (Glycosyl hydrolase family 43) that includes enzymes with the following activities: beta-xylosidase (EC 3.2.1.37), alpha-L-arabinofuranosidase (EC 3.2.1.55), arabinanase (EC 3.2.1.99), and xylanase (EC 3.2.1.8).

GO terms were enriched for plant helminths, reflecting some specific and vital processes such as pectate lyase activity (GO:0030570; FDR = 1,14E-23). PF03211 (pectate lyase) is the domain related to this GO term. This activity is responsible for the maceration and soft rotting of plant tissue and has been implicated in plant disease (Marín-Rodríguez et al., 2002). Cellulase activity (GO:0008810; FDR = 8,32E-17), another plant-specific enriched term, is related to colonization of the plant by helminths and penetration in plant tissues (Ma et al., 2011). Defense response to bacterium (GO:0042742; FDR = 2,96E-13) is related to reactions triggered in response to the presence of bacteria that act to protect the cell or organism such as the effect of antibacterial peptide activity (Cotton et al., 2012).

3.2.7. Pfam domain comparisons in secreted and non-secreted proteins

Comparisons between secreted and non-secreted protein domains across 44 helminth species allowed profiling of the secretome fingerprints. Five thousand, four hundred and twenty-nine domains were identified in non-secreted and 2,345 in secreted proteins. Fifty-six domains were secretome-exclusive (Supplementary Table S5) and 2,289 were shared between secreted and non-

secreted proteins. Secretome-specific domains are involved in processes such as recognition, binding, degradation and uptake of extracellular complex nutrients, signal transduction, and adhesion.

3.3. Complex and repetitive secreted proteins and GO terms enrichment

Out of 20,607 secreted proteins with Pfam domain annotations, 98.9% have a simple domain organization (Suh and Hutter, 2012), containing ≤3 different domains, and the following GO terms were enriched in these proteins: protein disulfide isomerase activity (GO:0003756; FDR = 3,28E-11), cell redox homeostasis (GO:0045454; FDR = 1,50E-10), metalloendopeptidase activity (GO:0004222; FDR = 4,41E-10), serine-type endopeptidase inhibitor activity (GO:0004867; FDR = 1,03E-08), cysteine-type endopeptidase activity (GO:0004197; FDR = 3,87E-05), and serine-type carboxypeptidase activity (GO:0004185; FDR = 4,96E-05).

Our results also showed that 1.1% of the secreted proteins contained ≥4 different domains, which according to Suh and Hutter (2012) are named "complex secreted proteins". In these proteins, we found enriched GO terms such as cell adhesion mediated by integrin (GO:0033627; FDR = 4,99E-24), basement membrane organization (GO:0071711; FDR = 3,37E-22), positive regulation of endopeptidase activity (GO:0010950; FDR = 9,31E-18), positive regulation of locomotion (GO:0040017; FDR = 2,59E-17), response to misfolded protein (GO:0051788; FDR = 1,91E-18), and regulation of cell proliferation (GO:0042127; FDR = 1,17E-10). Among these terms, there are proteins involved in protein–protein interaction including laminin, integrins, proteins highly enriched in EGF domains and thrombospondin repeats (Dzik, 2006). Table 3 indicates the number of complex proteins per species according to the classification related to the number of different domains present in a protein. *Schistosoma mansoni* had the highest percentage (3.4%) of proteins with four or more different domains. *Toxocara canis* did not have complex secreted proteins in the secretome.

The majority of potentially secreted proteins contained a small number of domains. Out of 4,451 protein architectures, 2,276 (51.1%) had a single domain (unidomain), which was represented in 14,444 proteins (70,1%). One hundred and sixty-eight (3.7%) architectures contained more than 10 domains, which were represented in 237 (1.1%) proteins. Only 62 architectures (1.4%) contained more than 20 domains, which were represented in 103 (0.5%) proteins. Compared with non-secreted proteins, out of 25,708 architectures, 5,439 (21.1%) were unidomain, represented in 202,862 (43.2%) proteins. Seven hundred and nineteen (2.8%) architectures contain more than 10 domains that were equivalent to 1,159 (0.24%) proteins and 103 (0.40%) contained more than 20 domains equivalent to 354 (0.07%) proteins (Supplementary Table S6). In general, secreted proteins were simpler in terms of the domain architecture than non-secreted proteins. However, large proteins containing more than 10 and 20 domains were over-represented in the secreted proteins compared with non-secreted. This is explained by the presence in the secretome of proteins that have highly repetitive domains and are involved in cell to cell communication, protein binding and adhesion, among other functions.

Repetitive secreted proteins contain multiple copies of one or two different domains (Suh and Hutter, 2012). The top 50 highly repetitive secreted proteins contained 83 proteins. Mainly nematode proteins were part of this classification (Supplementary Table S7). *Caenorhabditis elegans* was the species with more proteins (19 in total). Among the top 50 architectures of highly repetitive proteins, most had a signal peptide, which means that they are proteins secreted via the classical pathway (Petersen et al., 2011). Only two proteins had no signal peptide. The top 50 highly

Table 3
Protein domain organization in secreted proteins. Simple domain organization: ≤3 different domains. Complex secreted proteins: ≥4 different domains. The value as a percentage in relation to the secreted proteins with Pfam domain annotations per species is shown in parentheses.

Species	Proteins with ≤3 different domains (%)	Proteins with ≥4 different domains (%)
Nematoda		
Clade I		
Trichinella spiralis	406 (98.5)	6 (1.5)
Trichuris suis	403 (98.3)	7 (1.7)
Clade III		
Acanthocheilonema viteae	307 (97.5)	8 (2.5)
Ascaris suum	527 (99.2)	4 (0.8%)
Brugia malayi	456 (97.2)	13 (2.8)
Dirofilaria immitis	310 (97.1)	9 (2.9)
Dracunculus medinensis	279 (99.6)	1 (0.4)
Elaeophora elaphi	306 (99.6)	1 (0.4)
Enterobius vermicularis	347 (99.1)	3 (0.9)
Litomosoides sigmodontis	302 (97.4)	8 (2.6)
Loa loa	364 (97.8)	8 (2,2)
Onchocerca volvulus	352 (98.0)	7 (2)
Syphacia muris	320 (99.6)	1 (0.4)
Thelazia callipaeda	268 (98.9)	3 (1.1)
Toxocara canis	405 (100)	0
Wuchereria bancrofti	266 (99.6)	1 (0.4)
Clade IV		
Bursaphelenchus xylophilus	999 (98.9)	11 (1.1)
Globodera pallida	488 (99.1)	4 (0.9)
Meloidogyne hapla	359 (98.6)	5 (1.4)
Parastrongyloides trichosuri	759 (99.8)	1 (0.2)
Rhabditophanes sp. *KR3021*	585 (98.8)	7 (1.2)
Strongyloides ratti	537 (99.2)	4 (0.8)
Clade V		
Ancylostoma ceylanicum	732 (99.7)	2 (0.3)
Angiostrongylus costaricensis	254 (99.6)	1 (0.4)
Dictyocaulus viviparus	433 (99.0)	4 (1)
Haemonchus contortus	1,227 (98.9)	13 (1)
Necator americanus	601 (99.5)	3 (0.5)
Nippostrongylus brasiliensis	594 (99.6)	2 (0.4)
Oesophagostomum dentatum	763 (99.6)	3 (0.4)
Pristionchus pacificus	926 (99.3)	6 (0.7)
Caenorhabditis elegans	1,798 (98)	35 (2)
Platyhelminthes		
Cestoda		
Echinococcus multilocularis	253 (98.4)	4 (1.6)
Hydatigera taeniaeformis	229 (99.5)	1 (0.5)
Hymenolepis microstoma	212 (97.2)	6 (2.8)
Mesocestoides corti	175 (99.4)	1 (0.6)
Taenia solium	246 (99.6)	1 (0.4)
Trematoda		
Clonorchis sinensis	319 (98.7)	4 (1.3)
Fasciola hepática	449 (99.3)	3 (0.7)
Schistosoma haematobium	200 (98.5)	3 (1.5)
Schistosoma japonicum	246 (99.6)	1 (0.4)
Schistosoma mansoni	227 (96.6)	8 (3.4)
Opisthorchis viverrini	296 (98)	6 (2)
Monogenea		
Gyrodactylus salaris	265 (98.8)	3 (1.2)
Turbellaria		
Schmidtea mediterranea	592 (99.5)	3 (0.5)

repetitive proteins contained 21 different domains (Supplementary Table S7). Domains involved in protein–protein interactions, present in cystein-rich proteins (particularly characteristic of secreted proteins) and endopeptidase inhibitor activity, among other functions, were characteristic of the repetitive proteins. Domains of unknown function were also present. Many of those were found exclusively in nematodes.

3.4. Orthologous classification of secreted proteins

OrthoMCL was used to arrange the secreted proteins into clusters and to identify groups of the most conserved proteins among secretomes. A total of 26,870 proteins (65%) out of 41,200 secreted proteins were identified across 44 species and were sorted into 7,419 orthologous clusters. Two thousand and sixty-six proteins matched other sequences in the OrthoMCL online database, but had the "NO_GROUP" (designation by OrthoMCL DB) (Table 4). Additionally, 12,265 proteins had no OrthoMCL database hits and were considered unique. Certain clusters contained a large number of proteins. The most abundant OG was OG5_186610 with 169 sequences. The proteins included in this OG belonged to the CAP protein family (PF00188, cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins) (Supplementary Table S8).

Table 4 shows OrthoMCL distributions of secretome proteins by species. Although the *S. mansoni* and *S. haematobium* secretomes were smaller, a higher proportion of the *S. mansoni* secretome (425/431 proteins) and *S. haematobium* secretome (364/379) had OrthoMCL group assignments. These proportions were comparable with *C. elegans* (3,099/3,121). In terms of OG_diversity, *S. haematobium* had the largest number (0.75) of OGs according to total secretome. The secretome with the least diversity of OGs was *T. spiralis* (0.29). We did not find an OG shared by all 44 helminths species.

We found OGs shared between Nematoda and Platyhelminthes or specific to each phylum (Supplementary Table S9). In Platyhelminthes, we found three specific OGs (OG5_184997, OG5_245819, OG5_246846) which are related to the protein trematode eggshell. In Nematoda we found nine specific OGs (OG5_143235, OG5_145143, OG5_165053, OG5_177002, OG5_192328, OG5_217124, OG5_220058, OG5_221445, OG5_221498) associated with nematode cuticle collagen protein and eight (OG5_192636, OG5_192710, OG5_214854, OG5_218309, OG5_218926, OG5_218927, OG5_218928, OG5_221603) with nematode fatty acid retinoid binding protein, among other specific OGs. These Nematoda proteins mentioned above have no counterparts in their plant or animal hosts and thus represent potential targets for new nematicides (Finn et al., 2016).

We also compared OGs between free-living and non-free-living, plant helminths and non-plant helminths, secreted and non-secreted proteins, and obtained species-specific OGs (Supplementary Table S9). Regarding these three comparisons, in free-living helminths, we identified a specific OG that is involved in chitin degradation (OG5_146351). A plant helminth-specific OG was identified (OG5_132299), which is related to pectate lyase and celullase activity. OGs related to nuclear proteins such as helicases (OG5_126701), reverse transcriptase (OG5_126627) and deoxyribonucleases (OG5_126665), for example, were observed exclusively in the non-secreted group in comparison with secreted proteins. Secretome-specific OGs were found to be involved in degradation processes (OG5_134198), adhesion and protein binding (OG5_139086, OG5_241782) among other typical secreted proteins.

3.5. Dynamics of domain architecture

The emergence of proteins with new and species-specific domains or domain combinations could be one of the main mechanisms of secretome evolution and diversity (Barrera et al., 2014). Therefore, distinct domain architectures can give rise to new functions and new molecular interaction alternatives (Barrera et al., 2014).

On average, each secreted protein had 1.73 domains, a proportion that did not vary significantly among taxonomic groups. *Brugia malayi* (2.89) had the highest number of domains per protein and *F. hepatica* (1.27) the lowest. The number of domains in non-secreted proteins was 1.65 per protein on average. Architecture diversity in a species is the number of protein architectures observed, divided by the number of proteins with Pfam domains. If the ratio is close to 1 there is a greater diversity of protein architectures because it means that every protein represents specific domain architecture. According to our results, Platyhelminthes have the highest architecture diversity in secreted proteins (Fig. 4).

Overall, 4,451 unique domain architectures were identified across 44 secretomes, with 2,830 of those (63.5%) appearing exclusively in a single secretome and only two 'core' domain architectures (0.04%) that were present in all secretomes. Non-secreted proteins displayed a richer domain architecture with 25,709 unique domain architectures identified; 13,934 (54.1%) of those appeared exclusively in a single secretome and 158 'core' domain architectures (0.61%). The 'core' domain architectures of the secretomes were composed of only one domain, corresponding to proteins that were related to essential and conserved functions such as the Homeobox domain (PF00046) secreted by non-classical pathway and the cysteine-rich secretory family (PF00188, SP), a classical pathway secreted protein.

The average number of exclusive domain architectures per species was 64, ranging from 24 to 157. *Fasciola hepatica* had the highest number of exclusive domain architectures (41.8%). The lowest number of exclusive domain architectures was found in *T. callipaeda* (11%). Free-living helminths had values above 26% with the highest value in *S. mediterranea* (32%). Plant parasitic helminths and the ectoparasite *G. salaris* also had high values of exclusive domain architectures. The highest numbers of domain architectures were found in *C. elegans* (594) and *Mesocestoides corti* (143) had the lowest value.

4. Discussion

The identification of secreted proteins may provide a catalog of potential new immunomodulators, the development of new diagnostic tests, potential new drug targets and treatments (Geary et al., 2012). The need for new methods for diagnosis and control is well recognized (Liang et al., 2003; Melman et al., 2009). For this reason, the study of helminth secretomes may provide new avenues for the development of control measures. The secretomes of helminths also provide the repertoire of proteins that show the imprint of adaptation to several habitats (Krijger et al., 2014).

The helminth expressed secretome is dynamic and adjusts to the developmental stage of the worm, the milieu to which each life cycle stage is exposed and the state of the host immune system, among other conditions (McSorley and Maizels, 2012). For example, it has been shown for *B. malayi* microfilariae that the composition of the secreted proteins depends on the local environment of the microfilarie in the human host (Moreno and Geary, 2008). On the other hand, it has been demonstrated that not only secreted proteins are involved, but also carbohydrates (Thomas et al., 2003; Jenkins et al., 2005) and lipid mediators play important roles in modulation of the host immune system by worms (Van der Kleij et al., 2002; Brattig et al., 2009).

The quality of genome assemblies and gene predictions might have an effect on our results which should be experimentally validated. However, these predictions provide strong clues as to

Table 4
OrthoMCL results. The number of proteins observed in each OrthoMCL category. Hits (%): OrthoMCL database hits, the value as a percentage in relation with the secretome size is shown in parentheses. OGs: ortholog groups. OG_diversity: number of orthologous groups divided by the secretome (OGs/secretome); if the ratio is close to 1, there is a greater diversity of orthologous groups per species.

Species	Secretome size (seqs)	Hits (%)	OGs	No_Group	No hits	OG_diversity
Nematoda						
Clade I						
Trichinella spiralis	1,146	464 (40.4)	342	8	682	0.29
Trichuris suis	727	474 (65.2)	365	14	253	0.50
Clade III						
Acanthocheilonema viteae	507	451 (88.9)	362	57	56	0.71
Ascaris suum	1,158	772 (66.6)	618	52	386	0.53
Brugia malayi	800	729 (91.1)	473	142	71	0.59
Dirofilaria immitis	593	500 (84.3)	378	98	93	0.63
Dracunculus medinensis	499	356 (71.3)	274	29	143	0.54
Elaeophora elaphi	525	470 (89.5)	361	92	55	0.68
Enterobius vermicularis	649	447 (68.8)	374	28	202	0.57
Litomosoides sigmodontis	542	488 (90)	377	91	54	0.69
Loa loa	803	609 (75.8)	441	114	194	0.54
Onchocerca volvulus	682	566 (82.9)	408	96	116	0.59
Syphacia muris	584	431 (73.8)	359	26	153	0.61
Thelazia callipaeda	532	416 (78.2)	336	59	116	0.63
Toxocara canis	858	558 (65)	454	52	300	0.52
Wuchereria bancrofti	517	461 (89.1)	348	94	56	0.67
Clade IV						
Bursaphelenchus xylophilus	2,077	1,197 (57.6)	759	35	880	0.36
Globodera pallida	1,218	580 (47.6)	430	22	638	0.35
Meloidogyne hapla	943	450 (47.7)	378	13	493	0.40
Parastrongyloides trichosuri	1,334	889 (66.6)	546	28	445	0.40
Rhabditophanes sp. *KR3021*	1,192	739 (61.9)	545	33	453	0.45
Strongyloides ratti	973	680 (69.8)	480	27	293	0.49
Clade V						
Ancylostoma ceylanicum	1,169	951 (81.3)	672	24	218	0.57
Angiostrongylus costaricensis	509	362 (71.1)	318	10	147	0.62
Dictyocaulus viviparus	825	639 (77.4)	570	16	186	0.69
Haemonchus contortus	2,419	1,756 (72.6)	934	39	663	0.38
Necator americanus	1,231	905 (73.5)	735	26	326	0.59
Nippostrongylus brasiliensis	1,469	928 (63.1)	721	27	541	0.49
Oesophagostomum dentatum	1,531	1,115 (72.8)	799	20	416	0.52
Pristionchus pacificus	2,388	1,274 (53.3)	849	58	1,114	0.35
Caenorhabditis elegans	3,121	3,099 (99.3)	1,958	315	22	0.62
Platyhelminthes						
Cestoda						
Echinococcus multilocularis	524	319 (60.8)	268	16	205	0.51
Hydatigera taeniaeformis	468	276 (58.9)	241	13	192	0.51
Hymenolepis microstoma	454	267 (58.8)	205	13	187	0.45
Mesocestoides corti	406	223 (54.9)	191	5	183	0.47
Taenia solium	538	303 (56.3)	256	10	235	0.47
Trematoda						
Clonorchis sinensis	656	427 (65)	322	34	229	0.49
Fasciola hepática	992	756 (76.2)	647	28	236	0.65
Schistosoma haematobium	379	364 (96)	286	44	15	0.75
Schistosoma japonicum	476	408 (85.7)	351	35	68	0.73
Schistosoma mansoni	431	425 (98.6)	270	72	6	0.62
Opisthorchis viverrini	832	388 (46.6)	286	23	444	0.34
Monogenea						
Gyrodactylus salaris	653	307 (47)	263	10	346	0.40
Turbellaria						
Schmidtea mediterranea	870	716 (82.3)	461	18	154	0.52

which proteins to expect and the adaptations that evolved in organisms with distinct life styles.

We found that platyhelminthes have smaller secretomes, which is in agreement with previous studies (Garg and Ranganathan, 2012; Tsai et al., 2013; Gomez et al., 2015). Small secretomes could be related to an environment or niche with compounds or nutrients that are easier to obtain and metabolize (Krijger et al., 2014). The *C. elegans* genome encodes a large proportion of secreted proteins compared with other invertebrate and vertebrate organisms (Suh and Hutter, 2012). According to our results, 3,121 proteins were predicted as secreted. This represents 13% of the filtered proteome, being the largest secretome in this study. *Caenorhabditis elegans* possesses an elaborated set of secreted proteins, illustrating that genetic complexity does not necessarily correlate with anatomical complexity (Suh and Hutter, 2012).

The higher number of secreted proteins in the free-living nematodes *C. elegans* and *P. pacificus* permits the use of a wider variety of substrates present in soil or plant debris, which are likely more difficult to degrade than those available in animal hosts (Krijger et al., 2014). The plant helminth *B. xylophilus* also had a larger secretome than other animal helminths. This finding may be explained, assuming that animal hosts represent a nutritionally simpler envi-

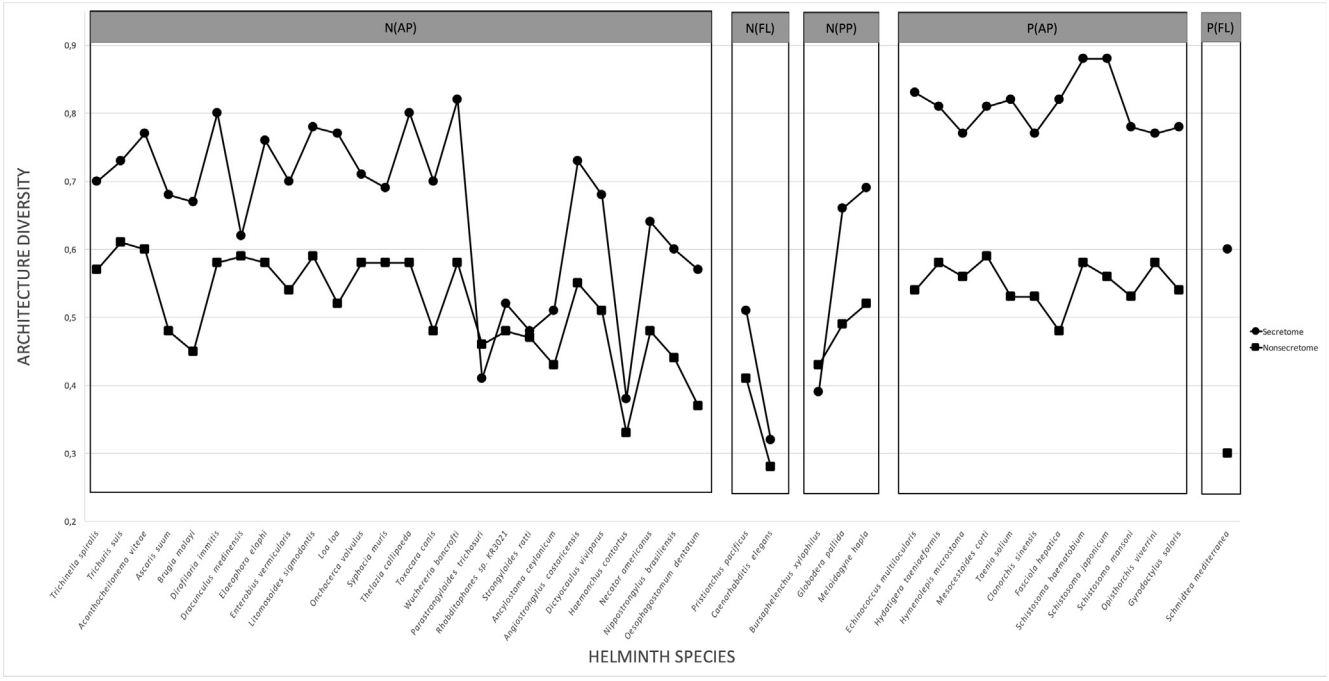

Fig. 4. Protein architecture diversity distribution across secreted and non-secreted proteins in 44 helminth species. Architecture diversity is the number of protein architectures observed divided by the number of proteins with Pfam domains in each species. If the ratio is close to 1, there is a greater diversity of protein architectures. Round marker, secreted proteins; square marker, non-secreted proteins. Gray bar across the top indicates the different categories: N, Nematoda; P, Platyhelminthes; AP, animal parasite; PP, plant parasite; FL, free living.

ronment than plant hosts (Krijger et al., 2014). Competition with microorganisms in the environment and other plant helminths are possibly more severe than in animal hosts, which may affect the number of secreted proteins involved in counteracting competitors (Krijger et al., 2014). Therefore, the secretome size could be related to parasitic life style and environment rather than the proteome size. *Schistosoma haematobium* has the smallest secretome across 44 helminth species. The small size of animal helminth secretomes may reflect an adaptation or a selective advantage in order to evade recognition by the immune system (Krijger et al., 2014).

However, according to our results *H. contortus*, a pathogenic nematode of ruminants, had a large secretome, almost equal to free-living nematodes and plant helminthes. The *H. contortus* secretome is particularly rich in peptidases linked to key roles in host invasion, locomotion, migration into stomach tissue (during the histotropic phase), degradation of blood, and other proteins as evasive mechanisms (Schwarz et al., 2013).

The study of protein domains has provided solutions to some human diseases. One example is the case of Kunitz domains (protease inhibitor domain) that are stable as standalone peptides, able to recognize specific protein structures. These properties have led to attempts to develop biopharmaceutical drugs targeting this domain (Lehmann, 2008). The first of these drugs to be marketed was a kallikrein inhibitor called ecallantide, which was used for the treatment of angioedema (Lehmann, 2008).

Protein domains are independent, compact and stable protein structural units that fold independently of other units, thus with potentially different biological functions (Barrera et al., 2014) and may catalyze different reactions (Barrera et al., 2014). The study of domain diversity in secretomes is a fast and effective way to characterize protein diversity and may provide clues to the different life styles and environments in which these organisms live. The architecture of secreted proteins is, therefore, relevant to the understanding of the interaction between the helminth and the environment. In addition, the prediction of domain architectures enables determination of the overall protein function and diversity, and has been used to transfer genomic annotations in newly sequenced genomes (Barrera et al., 2014).

Few domains were found in common across the secretomes, possibly because helminths have evolved separately as parasitic worms, leading to specific adaptations for their particular niche (Zarowiecki and Berriman, 2015). Only five domains were common across 44 secretomes; these domains were universal and are involved in biological functions such as peptidases, inhibitors, antioxidants and cell adhesion. All of these are important processes in interaction with the host (Dzik, 2006). Of the five domains, the Kunitz domain (PF00014) was the most frequent and is present in proteins containing 117 distinct architectures, suggesting that this domain is relevant in very diverse activities, one of those being protease inhibition (Rawlings et al., 2004).

According to our results, species-specific domains are involved in particular functions related to life style or ecological niche in helminths. For example, in *G. salaris*, a salmon ectoparasite, it was possible to identify domains related to pilus formation in bacteria. These domains could be participating in the attachment of the ectoparasite to the host and have important roles in virulence, infection and colonization by pathogens (Proft and Baker, 2009). However, the presence of these bacterial domains in the secretome of *G. salaris* could be also related to the ectoparasitic life style, because *Gyrodactilus* spp. feed on host mucus and epithelial cells (Cable and Harris, 2002). On the other hand, Bird and colleagues (2009) postulated that pilus formation domains have been acquired from bacteria by ancestral nematodes via horizontal gene transfer, and such events would be relevant for the establishment of the parasitic life style (Bird et al., 2009).

Exclusive plant domains such as those found in cellulases, pectate lyases and cell wall degrading enzymes are key adaptations towards plant parasitism, most probably achieved by horizontal gene transfer from a rhizobial bacteria (Dieterich and Sommer,

2009). In the potato parasite helminth *G. pallida*, we found the exclusive domain necrosis inducing protein (NPP1) (PF05630) with the architecture signal peptide (SP) and PF05630, that is involved in early stages of infection and is present in the potato fungus *Phytopthora infestans*, suggesting that plant parasitism has evolved from fungal associations (Dieterich and Sommer, 2009).

Specific glycoside hydrolases (GH) found in *B. xylophilus* suggest that they play important roles in fungal cell wall degradation (Shinya et al., 2013). GH 16 is involved in fungal cell wall degradation with endo-beta-1,3-glucanase activity and beta 1,3 glucan is one of the main components of the cell wall (Adams, 2004). This protein was only present in *B. xylophilus*, reflecting the difference in their food sources. Some secreted GH in *B. xylophilus* were acquired from other organisms by horizontal gene transfer as supported by phylogenetic evidence, which showed that they were gained from ascomycete fungi and bacteria (Kikuchi et al., 2011; Shinya et al., 2013). Another interesting GH protein identified was GH 18 (chitinase) that is involved in insect cuticle degradation and has an antifungal role (Staats et al., 2014). This protein was present in some helminths such as *B. malayi, B. xylophilus* and *O. volvulus* that have insect vectors.

The most represented domains across helminth secretomes were ranked in a top 25 classification. The ShK (PF01549) domain was the most common domain. It is suggested that this domain is important in parasitic interactions (Heizer et al., 2013). ShK proteins can inhibit calcium-dependent lymphocyte activation (Tudor et al., 1996). This suggests a direct immunomodulatory role for ShK homologs in helminths and its potential biopharmaceutical applications. Kunitz_BPTI (PF00014) and Inhibitor_I29 (PF08246) are protease inhibitors. It has been suggested that these are involved in protecting helminths from host molecules, in particular those derived from the gastrointestinal tract, such as a broad diversity of peptidases (Heizer et al., 2013). In this way gastrointestinal helminths can safely navigate and survive within host digestive tract.

CAP (PF00188) domains were also among the most prevalent domains across the helminth secretomes analyzed. They play roles in larval migration and evasion of the host immune response (Sotillo et al., 2014). CAP domains have also been found associated with proteins with immunodulatory activity (Cantacessi et al., 2009) and have been studied in some parasitic nematode species such as the hookworm *Ancylostoma caninum* (Hawdon et al., 1999), and the murine strongyloid nematode, *Heligmosomoides polygyrus* (Moreno et al., 2011).

The trypsin (PF00089) domain is involved in the breakdown of proteins. Trypsin domains were up-regulated in the parasitic stages of the nematodes *Cooperia oncophora* and *Ostertagia ostertagi* (Heizer et al., 2013). Proteins associated with this domain play a role in the feeding process (Goyal et al., 2005). These secreted proteases may also participate in countering the host immune responses by hydrolyzing antibodies or in parasite establishment in the host (Heizer et al., 2013).

The lectin_C (PF00059) domain is related to extracellular metazoan proteins with diverse functions. In general, it is involved in calcium-dependent carbohydrate binding. This domain has been linked to proteins involved in the host-parasite interface which may assist in evading the host immune response (Loukas and Maizels, 2000). Some nematode C-type lectins have been observed in the parasite's epicuticle (Heizer et al., 2013).

Hosts use oxidative stress as a means of combating parasites (Schirmer et al., 1987). It was hypothesized that parasites would have a very well developed redox system to defend themselves against ROS attacks (Zarowiecki and Berriman, 2015). Domains involved in the detoxification process such as transthyretin-like family (PF01060) and thioredoxin (PF00085) were among the most represented in the present work. Another interesting process in host-helminth interaction was lipid catabolism due to the lipid-rich environment in which helminths reside (Bansal et al., 2005). Lipase_3 (PF01764) and coesterase (PF00135) are involved in host-parasite interactions and according our results are placed among the most represented domains across the helminth secretomes. Domains involved in cell adhesion or protein–protein interaction such as: I-set (PF07679), CUB (PF00431), PAN_1 (PF00024), and TSP_1 (PF00090) were placed in this top 25 ranking.

Among the most represented predicted protein families across secretomes, we found proteases which are involved in blood coagulation, protein metabolism, immune reactions and tissue remodeling (Dzik, 2006). The role of proteases in extracellular matrix degradation seems to be exclusive to parasites. Thus, free-living helminths do not secrete extracellular matrix degradation proteases (Lackey et al., 1989). The specific action and release of peptidases after host infection is an important function in the transition from a free-living organism to a parasite (Hawdon et al., 1995; Gamble and Mansfield, 1996). Metallo- and cysteine proteases (PF01400 and PF00112, respectively) are involved in the tissue/cell invasion process and nutrient uptake. In this study astacin (PF01400) is the most represented domain in nematodes. This could support the idea that these proteins had an expansion in nematodes as reported by Park et al., (2010). Furthermore, studies suggest that cysteine proteases appeared early in evolution, degrading intra- and extracellular proteins (Sajid and McKerrow, 2002). On the other hand, aspartic proteases correspond to a major enzymatic class of parasitic helminths and play a key role in the ability to degrade hemoglobin (Brinkworth et al., 2001). Such proteins are the targets of protective antibodies against the human parasite *Necator americanus* (Pearson et al., 2009).

Domain repetitions in multidomain proteins are important for the overall domain function (Messih et al., 2012). Domain repetition is a predominant mechanism for protein diversity and evolution (Barrera et al., 2014). For example, the glutamate receptor interacting protein (GRIP) contains seven PDZ domains, two of which interact with a α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) receptor, but only in the presence of the adjacent copies (Messih et al., 2012).

Our results showed that domains such as PAN, TSP-1, VWA, I-set and EGF are usually present as tandem copies, sometimes in combination with other adhesive domains. The presence of these repeats suggests that they play a common functional role during the invasion process (host interaction) (Suh and Hutter, 2012; Mendes et al., 2013). Adhesive proteins play important roles in ligand binding, cell to cell and cell-extracellular matrix interactions (Bork and Rohde, 1991).

In the present study, we provide access to the full list of core domains, exclusive domains and domain architectures of the secreted proteins across 44 helminths. This information can be a useful resource for researchers interested in comparative studies of secretomes across different helminths in order to know more about the protein evolution in these crucial proteins and their interaction with the host, and to understand protein function and evolution. The number and order of the domains will determine the function, as is the case of multimodular enzymes. The rationale is that proteins containing the same domain composition could be similarly annotated (Barrera et al., 2014).

Our results showed that secreted proteins have higher architecture diversity compared with non-secreted proteins. The only exceptions were the secretomes of *B. xylophilus* and *Parastrongyloides trichosuri*. This metric was calculated in order to shed light on the diversity of secreted proteins in terms of protein domains and architectures, once domain and architecture diversity point toward the existence of different mechanisms to attach to or interact with host components and the environment. On average, Platy-

helminthes had the smaller secretomes across 44 helminth species, albeit with more diverse architectures.

We believe this is the first proteome-wide comparative study of predicted secretomes in helminths using species with different life styles and habitats. Our approach identified an enriched list of annotated secreted proteins across the predicted proteomes of 44 species. The secretome consisted on average of 7.6% of the filtered proteomes and most of the proteins ranged in size from 100 to 300 aa, with *C. elegans* having the largest secretome with 3,121 proteins. The most common domain across the secretomes was the Shk domain. Only five protein domains were conserved across the 44 secretomes and reveal mechanisms that appear to be conserved among plant, animal and free-living helminths. Our findings also indicate that the secretome composition is not conserved across species and the differences suggest possible unique adaptations to specific niches. The data provided by our integrative approach will be relevant for studies on the evolution of these species to occupy different habitats, the evolution of protein families and diversification in the host-parasite molecular interactions. Finally, the knowledge base contains underlying information for the targeted development of new parasite control measures based on vaccines and new diagnostic tools.

Acknowledgments

The authors acknowledge the use of computing resources of the Bioinformatics Platform at Oswaldo Cruz Foundation-Minas, Brazil. This work was supported by the National Institutes of Health (NIH)/Fogarty International Center, USA (TW007012 and 1P50AI098507-01) to GO, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil (REDE 21/2015 and 070/13) to GO, FAPEMIG (RED-00014-14 and PPM-00189-13) to GO and Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq, Brazil (304138/2014-2) to GO. GO is a CNPq fellow (309312/2012-4). FSO is a Fundação de Amparo à Pesquisa do Estado de Minas-FAPEMIG, Brazil fellow and YCA is a CAPES fellow.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijpara.2017.01.007.

References

Adams, D.J., 2004. Fungal cell wall chitinases and glucanases. Microbiology 150, 2029–2035.
Amaral, A.C., Silva, O.N., Mundim, N.C.C.R., De Carvalho, M.J.A., Migliolo, L., Leite, J.R. S.A., Prates, M.V., Bocca, A.L., Franco, O.L., Felipe, M.S.S., 2012. Predicting antimicrobial peptides from eukaryotic genomes: *In silico* strategies to develop antibiotics. Peptides 37, 301–308.
Anthony, K.G., Kathir, P., Moore, D., Ippen-Ihler, K., Frost, L.S., 1996. Analysis of the traLEKBP sequence and the TraP protein from three F-like plasmids: F, R100-1 and ColB2. J. Bacteriol. 178, 3194–3200.
Bahlool, Q.Z.M., Skovgaard, A., Kania, P.W., Buchmann, K., 2013. Effects of excretory/secretory products from *Anisakis simplex* (Nematoda) on immune gene expression in rainbow trout (*Oncorhynchus mykiss*). Fish Shellfish Immunol. 35, 734–739.
Bansal, D., Bhatti, H.S., Sehgal, R., 2005. Role of cholesterol in parasitic infections. Lipids Health Dis. 4, 10.
Barrera, A., Alastruey-Izquierdo, A., Martín, M.J., Cuesta, I., Vizcaíno, J.A., 2014. Analysis of the protein domain and domain architecture content in fungi and its application in the search of new antifungal targets. PLoS Comput. Biol. 10, e1003733.
Beier, S., Bertilsson, S., 2013. Bacterial chitin degradation-mechanisms and ecophysiological strategies. Front. Microbiol. 4, 149.
Bendtsen, J.D., Jensen, L.J., Blom, N., Von Heijne, G., Brunak, S., 2004. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel. 17, 349–356.
Bendtsen, J.D., Nielsen, H., Widdick, D., Palmer, T., Brunak, S., 2005. Prediction of twin-arginine signal peptides. BMC Bioinformatics 6, 167.
Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., Apweiler, R., 2009. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 25, 3045–3046.
Bird, D.M., Williamson, V.M., Abad, P., McCarter, J., Danchin, E.G.J., Castagnone-Sereno, P., Opperman, C.H., 2009. The genomes of root-knot nematodes. Annu. Rev. Phytopathol. 47, 333–351.
Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., Vida, J.T., Thomas, W.K., 1998. A molecular evolutionary framework for the phylum Nematoda. Nature 392, 71–75.
Blüthgen, N., Brand, K., Cajavec, B., Swat, M., Herzel, H., Beule, D., 2005. Biological profiling of gene groups utilizing Gene Ontology. Genome Inform. 16, 106–115.
Bork, P., Rohde, K., 1991. More von Willebrand factor type A domains? Sequence similarities with malaria thrombospondin-related anonymous protein, dihydropyridine-sensitive calcium channel and inter-alpha-trypsin inhibitor. Biochem. J., 908–910
Brattig, N.W., Schwohl, A., Hoerauf, A., Büttner, D.W., 2009. Identification of the lipid mediator prostaglandin E2 in tissue immune cells of humans infected with the filaria *Onchocerca volvulus*. Acta Trop. 112, 231–235.
Brindley, P.J., Mitreva, M., Ghedin, E., Lustigman, S., 2009. Helminth genomics: The implications for human health. PLoS Negl. Trop. Dis. 3, e538.
Brinkworth, R.I., Prociv, P., Loukas, A., Brindley, P.J., 2001. Hemoglobin-degrading, aspartic proteases of blood-feeding parasites. Substrate specificity revealed by homology models. J. Biol. Chem. 276, 38844–38851.
Cable, J., Harris, P.D., 2002. Gyrodactylid developmental biology: Historical review, current status and future trends. Int. J. Parasitol. 32, 255–280.
Cantacessi, C., Campbell, B.E., Visser, A., Geldhof, P., Nolan, M.J., Nisbet, A.J., Matthews, J.B., Loukas, A., Hofmann, A., Otranto, D., Sternberg, P.W., Gasser, R.B., 2009. A portrait of the "SCP/TAPS" proteins of eukaryotes - Developing a framework for fundamental research and biotechnological outcomes. Biotechnol. Adv. 27, 376–388.
Cass, C.L., Johnson, J.R., Califf, L.L., Xu, T., Hernandez, H.J., Stadecker, M.J., Yates, J.R., Williams, D.L., 2007. Proteomic analysis of *Schistosoma mansoni* egg secretions. Mol. Biochem. Parasitol. 155, 84–93.
Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.
Cotton, S., Donnelly, S., Robinson, M.W., Dalton, J.P., Thivierge, K., 2012. Defense peptides secreted by helminth pathogens: antimicrobial and/or immunomodulator molecules? Front. Immunol. 3, 269.
Cotton, J.A., Lilley, C.J., Jones, L.M., Kikuchi, T., Reid, A.J., Thorpe, P., Tsai, I.J., Beasley, H., Blok, V., Cock, P.J., Eves-van den Akker, S., Holroyd, N., Hunt, M., Mantelin, S., Naghra, H., Pain, A., Palomares-Rius, J.E., Zarowiecki, M., Berriman, M., Jones, J.T., Urwin, P.E., 2014. The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. Genome Biol. 15, R43.
Craig, J.M., Scott, A.L., 2014. Helminths in the lungs. Parasite Immunol. 36, 463–474.
Cuesta-Astroz, Y., Scholte, L.L.S., Pais, F.S.M., Oliveira, G., Nahum, L.A., 2014. Evolutionary analysis of the cystatin family in three *Schistosoma* species. Front. Genet. 5, 206.
Dieterich, C., Sommer, R.J., 2009. How to become a parasite – Lessons from the genomes of nematodes. Trends Genet. 25, 203–209.
Duckert, P., Brunak, S., Blom, N., 2004. Prediction of proprotein convertase cleavage sites. Protein Eng. Des. Sel. 17, 107–112.
Dzik, J.M., 2006. Molecules released by helminth parasites involved in host colonization. Acta Biochim. Pol. 53, 33–64.
Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016.
Fellbrich, G., Romanski, A., Varet, A., Blume, B., Brunner, F., Engelhardt, S., Felix, G., Kemmerling, B., Krzymowska, M., Nürnberger, T., 2002. NPP1, a *Phytophthora*-associated trigger of plant defense in parsley and *Arabidopsis*. Plant J. 32, 375–390.
Ferguson, B.J., Newland, S.A., Gibbs, S.E., Tourlomousis, P., Fernandes dos Santos, P., Patel, M.N., Hall, S.W., Walczak, H., Schramm, G., Haas, H., Dunne, D.W., Cooke, A., Zaccone, P., 2015. The *Schistosoma mansoni* T2 ribonuclease omega-1 modulates inflammasome-dependent IL-1β secretion in macrophages. Int. J. Parasitol. 45, 809–813.
Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285.
Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., Shanmugam, D., Roos, D.S., Stoeckert, C.J., 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr. Protoc. Bioinforma. Chapter 6 (Unit-6), 1219.
Gamble, H.R., Mansfield, L.S., 1996. Characterization of excretory-secretory products from larval stages of *Haemonchus contortus* cultured in vitro. Vet. Parasitol. 62, 291–305.
Garg, G., Ranganathan, S., 2011. *In silico* secretome analysis approach for next generation sequencing transcriptomic data. BMC Genomics 12 (Suppl. 3), S14.
Garg, G., Ranganathan, S., 2012. Helminth secretome database (HSD): a collection of helminth excretory/secretory proteins predicted from expressed sequence tags (ESTs). BMC Genomics 13 (Suppl. 7), S8.

Geary, J., Satti, M., Moreno, Y., Madrill, N., Whitten, D., Headley, S.A., Agnew, D., Geary, T., Mackenzie, C., 2012. First analysis of the secretome of the canine heartworm, *Dirofilaria immitis*. Parasit. Vectors 5, 140.
Gehring, W.J., 1992. The homeobox in perspective. Trends Biochem. Sci. 17, 277–280.
Gijzen, M., Nürnberger, T., 2006. Nep1-like proteins from plant pathogens: recruitment and diversification of the NPP1 domain across taxa. Phytochemistry 67, 1800–1807.
Gomez, S., Adalid-Peralta, L., Palafox-Fonseca, H., Cantu-Robles, V.A., Soberón, X., Sciutto, E., Fragoso, G., Bobes, R.J., Laclette, J.P., del Yauner, L.P., Ochoa-Leyva, A., 2015. Genome analysis of Excretory/Secretory proteins in Taenia solium reveals their Abundance of Antigenic Regions (AAR). Sci. Rep. 5, 9683.
Goyal, K., Pinelli, C., Maslen, S.L., Rastogi, R.K., Stephens, E., Tunnacliffe, A., 2005. Dehydration-regulated processing of late embryogenesis abundant protein in a desiccation-tolerant nematode. FEBS Lett. 579, 4093–4098.
Greenbaum, D., Luscombe, N.M., Jansen, R., Qian, J., Gerstein, M., 2001. Interrelating different types of genomic data, from proteome to secretome: 'oming in on function. Genome Res. 11, 1463–1468.
Hahn, C., Fromm, B., Bachmann, L., 2014. Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparasitic neodermata. Genome Biol. Evol. 6, 1105–1117.
Harris, A.R.C., Russell, R.J., Charters, A.D., Hospital, R.P., Box, X.G.P.O., 1984. A review of schistosomiasis in immigrants in Western Australia, demonstrating the unusual longevity of *Schistosoma mansoni*. Trans. R. Soc. Trop. Med. Hyg. 78, 385–388.
Harris, T.W., Baran, J., Bieri, T., Cabunoc, A., Chan, J., Chen, W.J., Davis, P., Done, J., Grove, C., Howe, K., Kishore, R., Lee, R., Li, Y., Muller, H.M., Nakamura, C., Ozersky, P., Paulini, M., Raciti, D., Schindelman, G., Tuli, M.A., Van Auken, K., Wang, D., Wang, X., Williams, G., Wong, J.D., Yook, K., Schedl, T., Hodgkin, J., Berriman, M., Kersey, P., Spieth, J., Stein, L., Sternberg, P.W., 2014. WormBase 2014: New views of curated biology. Nucleic Acids Res. 42 (Database issue), D789–D793.
Hashimoto, M., Ikegami, T., Seino, S., Ohuchi, N., Fukada, H., Sugiyama, J., Shirakawa, M., Watanabe, T., 2000. Expression and characterization of the chitin-binding domain of chitinase A1 from *Bacillus circulans* WL-12. J. Bacteriol. 182, 3045–3054.
Hawdon, J.M., Jones, B.F., Perregaux, M.A., Hotez, P.J., 1995. *Ancylostoma caninum*: metalloprotease release coincides with activation of infective larvae in vitro. Exp. Parasitol. 80, 205–211.
Hawdon, J.M., Narasimhan, S., Hotez, P.J., 1999. *Ancylostoma* secreted protein 2: cloning and characterization of a second member of a family of nematode secreted proteins from *Ancylostoma caninum*. Mol. Biochem. Parasitol. 99, 149–165.
Heizer, E., Zarlenga, D.S., Rosa, B., Gao, X., Gasser, R.B., De Graef, J., Geldhof, P., Mitreva, M., 2013. Transcriptome analyses reveal protein and domain families that delineate stage-related development in the economically important parasitic nematodes, *Ostertagia ostertagi* and *Cooperia oncophora*. BMC Genomics 14, 118.
Hotez, P.J., Brindley, P.J., Bethony, J.M., King, C.H., Pearce, E.J., Jacobson, J., 2008. Helminth infections: the great neglected tropical diseases. J. Clin. Invest. 118, 1311–1321.
Jenkins, S.J., Hewitson, J.P., Ferret-Bernard, S., Mountford, A.P., 2005. Schistosome larvae stimulate macrophage cytokine production through TLR4-dependent and -independent pathways. Int. Immunol. 17, 1409–1418.
Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30, 1236–1240.
Kikuchi, T., Cotton, J.A., Dalzell, J.J., Hasegawa, K., Kanzaki, N., McVeigh, P., Takanashi, T., Tsai, I.J., Assefa, S.A., Cock, P.J.A., Otto, T.D., Hunt, M., Reid, A.J., Sanchez-Flores, A., Tsuchihara, K., Yokoi, T., Larsson, M.C., Miwa, J., Maule, A.G., Sahashi, N., Jones, J.T., Berriman, M., 2011. Genomic Insights into the Origin of Parasitism in the Emerging Plant Pathogen *Bursaphelenchus xylophilus*. PLoS Pathog. 7, e1002219.
Krijger, J.-J., Thon, M.R., Deising, H.B., Wirsel, S.G.R., 2014. Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. BMC Genomics 15, 722.
Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.
Lackey, A., James, E.R., Sakanari, J.A., Resnick, S.D., Brown, M., Bianco, A.E., McKerrow, J.H., 1989. Extracellular proteases of *Onchocerca*. Exp. Parasitol. 68, 176–185.
Lampe, R.A., Defeo, P.A., Davison, M.D., Young, J., Herman, J.L., Spreen, R.C., Horn, M. B., Mangano, T.J., Keith, R.A., 1993. Isolation and pharmacological characterization of omega-grammotoxin SIA, a novel peptide inhibitor of neuronal voltage-sensitive calcium channel responses. Mol. Pharmacol. 44, 451–460.
Lawley, T.D., Gilmour, M.W., Gunton, J.E., Standeven, L.J., Taylor, D.E., 2002. Functional and mutational analysis of conjugative transfer region 1 (Tra1) from the IncHI1 plasmid R27. J. Bacteriol. 184, 2173–2180.
Lehmann, A., 2008. Ecallantide (DX-88), a plasma kallikrein inhibitor for the treatment of hereditary angioedema and the prevention of blood loss in on-pump cardiothoracic surgery. Expert Opin. Biol. Ther. 8, 1187–1199.
Leontovyč, R., Young, N.D., Korhonen, P.K., Hall, R.S., Tan, P., Mikeš, L., Kašný, M., Horák, P., Gasser, R.B., 2016. Comparative Transcriptomic Exploration Reveals Unique Molecular Adaptations of Neuropathogenic Trichobilharzia to Invade and Parasitize Its Avian Definitive Host. PLoS Negl. Trop. Dis. 10, e0004406.
Liang, Y.S., Dai, J.R., Zhu, Y.C., Coles, G.C., Doenhoff, M.J., 2003. Genetic analysis of praziquantel resistance in *Schistosoma mansoni*. Southeast Asian. J. Trop. Med. Public Health 34, 274–280.
Lonsdale, A., Davis, M.J., Doblin, M.S., Bacic, A., 2016. Better Than Nothing? Limitations of the Prediction Tool SecretomeP in the Search for Leaderless Secretory Proteins (LSPs) in Plants. Front. Plant Sci. 7, 1451.
Loukas, A., Maizels, R.M., 2000. Helminth C-type lectins and host-parasite interactions. Parasitol. Today. 16, 333–339.
Lustigman, S., Prichard, R.K., Gazzinelli, A., Grant, W.N., Boatin, B.A., McCarthy, J.S., Basáñez, M.G., 2012. A research agenda for helminth diseases of humans: The problem of helminthiases. PLoS Negl. Trop. Dis. 6, e1582.
Ma, H.B., Lu, Q., Liang, J., Zhang, X.Y., 2011. Functional analysis of the cellulose gene of the pine wood nematode, *Bursaphelenchus xylophilus*, using RNA interference. Genet. Mol. Res. 10, 1931–1941.
Maizels, R.M., Yazdanbakhsh, M., 2003. Immune regulation by helminth parasites: cellular and molecular mechanisms. Nat. Rev. Immunol. 3, 733–744.
Maneewannakul, S., Maneewannakul, K., Ippen-Ihler, K., 1992. Characterization, localization, and sequence of F transfer region products: the pilus assembly gene product TraW and a new product, TrbI. J. Bacteriol. 174, 5567–5574.
Marcilla, A., Trelis, M., Cortés, A., Sotillo, J., Cantalapiedra, F., Minguez, M.T., Valero, M.L., Sánchez del Pino, M.M., Muñoz-Antoli, C., Toledo, R., Bernal, D., 2012. Extracellular vesicles from parasitic helminths contain specific excretory/secretory proteins and are internalized in intestinal host cells. PLoS One 7, e45974.
Marín-Rodríguez, M.C., Orchard, J., Seymour, G.B., 2002. Pectate lyases, cell wall degradation and fruit softening. J. Exp. Bot. 53, 2115–2119.
McSorley, H.J., Maizels, R.M., 2012. Helminth infections and host immune regulation. Clin. Microbiol. Rev. 25, 585–608.
Mehta, A., Brasileiro, A.C.M., Souza, D.S.L., Romano, E., Campos, M.A., Grossi-De-Sá, M.F., Silva, M.S., Franco, O.L., Fragoso, R.R., Bevitori, R., Rocha, T.L., 2008. Plant-pathogen interactions: What is proteomics telling us? FEBS J. 275, 3731–3746.
Melman, S.D., Steinauer, M.L., Cunningham, C., Kubatko, L.S., Mwangi, I.N., Wynn, N. B., Mutuku, M.W., Karanja, D.M.S., Colley, D.G., Black, C.L., Secor, W.E., Mkoji, G. M., Loker, E.S., 2009. Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. PLoS Negl. Trop. Dis. 3, e504.
Mendes, T.A.O., Lobo, F.P., Rodrigues, T.S., Rodrigues-Luiz, G.F., daRocha, W.D., Fujiwara, R.T., Teixeira, S.M.R., Bartholomeu, D.C., 2013. Repeat-enriched proteins are related to host cell invasion and immune evasion in parasitic protozoa. Mol. Biol. Evol. 30, 951–963.
Messih, M.A., Chitale, M., Bajic, V.B., Kihara, D., Gao, X., 2012. Protein domain recurrence and order can enhance prediction of protein functions. Bioinformatics 28, 444–450.
Moreno, Y., Geary, T.G., 2008. Stage- and gender-specific proteomic analysis of *Brugia malayi* excretory-secretory products. PLoS Negl. Trop. Dis. 2, e326.
Moreno, Y., Gros, P.P., Tam, M., Segura, M., Valanparambil, R., Geary, T.G., Stevenson, M.M., 2011. Proteomic analysis of excretory-secretory products of *Heligmosomoides polygyrus* assessed with Next-Generation sequencing transcriptomic information. PLoS Negl. Trop. Dis. 5, e1370.
Nahum, L.A., Mourão, M.M., Oliveira, G., 2012. New frontiers in *Schistosoma* genomics and transcriptomics. J. Parasitol. Res. 2012, 849132.
Nakai, K., Horton, P., 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. 24, 34–36.
Nombela, C., Gil, C., Chaffin, W.L., 2006. Non-conventional protein secretion in yeast. Trends Microbiol. 14, 15–21.
Oliveira, G., Pierce, R.J., 2015. How has the genomics era impacted schistosomiasis drug discovery? Future Med. Chem. 7, 685–687.
Park, J.-O., Pan, J., Möhrlen, F., Schupp, M.-O., Johnsen, R., Baillie, D.L., Zapf, R., Moerman, D.G., Hutter, H., 2010. Characterization of the astacin family of metalloproteases in *C. elegans*. BMC Dev. Biol. 10, 14.
Pearson, M.S., Bethony, J.M., Pickering, D.A., de Oliveira, L.M., Jariwala, A., Santiago, H., Miles, A.P., Zhan, B., Jiang, D., Ranjit, N., Mulvenna, J., Tribolet, L., Plieskatt, J., Smith, T., Bottazzi, M.E., Jones, K., Keegan, B., Hotez, P.J., Loukas, A., 2009. An enzymatically inactivated hemoglobinase from *Necator americanus* induces neutralizing antibodies against multiple hookworm species and protects dogs against heterologous hookworm infection. FASEB J. 23, 3007–3019.
Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.
Pomaznoy, M., Tatkov, S., Katokhin, A., Afonnikov, D., Babenko, V., Furman, D., Brusentsov, I., Belavin, P., Najakshin, A., Guselnikov, S., Vasiliev, G., Sivkov, A., Prokhortchouk, E., Skryabin, K., Mordvinov, V., 2013. Adult Opisthorchis felineus major protein fractions deduced from transcripts: comparison with liver flukes *Opisthorchis viverrini* and *Clonorchis sinensis*. Exp. Parasitol. 135, 297–306.
Proft, T., Baker, E.N., 2009. Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. Cell. Mol. Life Sci. 66, 613–635.
Ranganathan, S., Garg, G., 2009. Secretome: clues into pathogen infection and clinical applications. Genome Med. 1, 113.
Rawlings, N.D., Tolle, D.P., Barrett, A.J., 2004. Evolutionary families of peptidase inhibitors. Biochem. J. 378, 705–716.
Rezende, A.M., Folador, E.L., de Resende, D.M., Ruiz, J.C., 2012. Computational prediction of protein-protein interactions in Leishmania predicted proteomes. PLoS One 7, e51304.

Robinson, M.W., Dalton, J.P., 2009. Zoonotic helminth infections with particular emphasis on fasciolosis and other trematodiases. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364, 2763–2776.

Robinson, M.W., Menon, R., Donnelly, S.M., Dalton, J.P., Ranganathan, S., 2009. An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host. Mol. Cell. Proteomics 8, 1891–1907.

Rowe, J.A., Claessens, A., Corrigan, R.A., Arman, M., 2009. Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. Expert Rev. Mol. Med. 11, e16.

Sajid, M., McKerrow, J.H., 2002. Cysteine proteases of parasitic organisms. Mol. Biochem. Parasitol. 120, 1–21.

Schicht, S., Qi, W., Poveda, L., Strube, C., 2013. The predicted secretome and transmembranome of the poultry red mite *Dermanyssus gallinae*. Parasit. Vectors 6, 259.

Schirmer, R.H., Schöllhammer, T., Eisenbrand, G., Krauth-Siegel, R.L., 1987. Oxidative stress as a defense mechanism against parasitic infections. Free Radic. Res. Commun. 3, 3–12.

Schwarz, E.M., Korhonen, P.K., Campbell, B.E., Young, N.D., Jex, A.R., Jabbar, A., Hall, R.S., Mondal, A., Howe, A.C., Pell, J., Hofmann, A., Boag, P.R., Zhu, X.-Q., Gregory, T., Loukas, A., Williams, B.A., Antoshechkin, I., Brown, C., Sternberg, P.W., Gasser, R.B., 2013. The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. Genome Biol. 14, R89.

Shinya, R., Morisaka, H., Kikuchi, T., Takeuchi, Y., Ueda, M., Futai, K., 2013. Secretome analysis of the pine wood nematode *Bursaphelenchus xylophilus* reveals the tangled roots of parasitism and its potential for molecular mimicry. PLoS One 8, e67377.

Sotillo, J., Sanchez-Flores, A., Cantacessi, C., Harcus, Y., Pickering, D., Bouchery, T., Camberis, M., Tang, S.-C., Giacomin, P., Mulvenna, J., Mitreva, M., Berriman, M., LeGros, G., Maizels, R.M., Loukas, A., 2014. Secreted proteomes of different developmental stages of the gastrointestinal nematode *Nippostrongylus brasiliensis*. Mol. Cell. Proteomics 13, 2736–2751.

Sotillo, J., Pearson, M., Potriquet, J., Becker, L., Pickering, D., Mulvenna, J., Loukas, A., 2016. Extracellular vesicles secreted by *Schistosoma mansoni* contain protein vaccine candidates. Int. J. Parasitol. 46, 1–5.

Spruance, S.L., 1974. Latent period of 53 years in a case of hydatid cyst disease. Arch. Intern. Med. 134, 741–742.

Staats, C.C., Junges, A., Guedes, R.L.M., Thompson, C.E., de Morais, G.L., Boldo, J.T., de Almeida, L.G.P., Andreis, F.C., Gerber, A.L., Sbaraini, N., da Paixão, R.L., Broetto, de A., Landell, M., Santi, L., Beys-da-Silva, W.O., Silveira, C.P., Serrano, T.R., de Oliveira, E.S., Kmetzsch, L., Vainstein, M.H., de Vasconcelos, A.T.R., Schrank, A., 2014. Comparative genome analysis of entomopathogenic fungi reveals a complex set of secreted proteins. BMC Genomics 15, 822.

Suh, J., Hutter, H., 2012. A survey of putative secreted and transmembrane proteins encoded in the *C. elegans* genome. BMC Genomics 13, 333.

Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K.S., Vorgias, C.E., 1996. Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. Nat. Struct. Biol. 3, 638–648.

Thomas, P.G., Carter, M.R., Atochina, O., Da'Dara, A.A., Piskorska, D., McGuire, E., Harn, D.A., 2003. Maturation of dendritic cell 2 phenotype by a helminth glycan uses a Toll-like receptor 4-dependent mechanism. J. Immunol. 171, 5837–5841.

Tjalsma, H., Bolhuis, A., Jongbloed, J.D., Bron, S., van Dijl, J.M., 2000. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiol. Mol. Biol. Rev. 64, 515–547.

Tsai, I.J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K. L., Tracey, A., Bobes, R.J., Fragoso, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H.M., Cai, J., Camicia, F., Clark, R., Cucher, M., De Silva, N., Day, T.A., Deplazes, P., Estrada, K., Fernández, C., Holland, P.W.H., Hou, J., Hu, S., Huckvale, T., Hung, S.S., Kamenetzky, L., Keane, J.A., Kiss, F., Koziol, U., Lambert, O., Liu, K., Luo, X., Luo, Y., Macchiaroli, N., Nichol, S., Paps, J., Parkinson, J., Pouchkina-Stantcheva, N., Riddiford, N., Rosenzvit, M., Salinas, G., Wasmuth, J.D., Zamanian, M., Zheng, Y. *Taenia solium* Genome Consortium, Cai, X., Soberón, X., Olson, P.D., Laclette, J.P., Brehm, K., Berriman, M., 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496, 57–63.

Tudor, J.E., Pallaghy, P.K., Pennington, M.W., Norton, Raymond S., 1996. Solution structure of ShK toxin, a novel potassium channel inhibitor from a sea anemone. Nat. Struct. Biol. 3, 317–320.

Van der Kleij, D., Latz, E., Brouwers, J.F.H.M., Kruize, Y.C.M., Schmitz, M., Kurt-Jones, E.A., Espevik, T., De Jong, E.C., Kapsenberg, M.L., Golenbock, D.T., Tielens, A.G.M., Yazdanbakhsh, M., 2002. A novel host-parasite lipid cross-talk. Schistosomal lyso-phosphatidylserine activates toll-like receptor 2 and affects immune polarization. J. Biol. Chem. 277, 48122–48129.

Wang, Z., Martin, J., Abubucker, S., Yin, Y., Gasser, R.B., Mitreva, M., 2009. Systematic analysis of insertions and deletions specific to nematode proteins and their proposed functional and evolutionary relevance. BMC Evol. Biol. 9, 23.

Wasmuth, J., Schmid, R., Hedley, A., Blaxter, M., 2008. On the extent and origins of genic novelty in the phylum Nematoda. PLoS Negl. Trop. Dis. 2, e258.

Wootton, J.C., Federhen, S., 1996. Analysis of compositionally biased regions in sequence databases. Methods Enzymol. 266, 554–571.

World Health Organisation, 2015. Investing to Overcome the Global Impact of Neglected Tropical Diseases. Third WHO Report on Neglected Tropical Diseases. World Health Organization, Geneva, Switzerland. Available at: <http://apps.who.int/iris/bitstream/10665/152781/1/9789241564861_eng.pdf?ua=1>.

Young, N.D., Nagarajan, N., Lin, S.J., Korhonen, P.K., Jex, A.R., Hall, R.S., Safavi-Hemami, H., Kaewkong, W., Bertrand, D., Gao, S., Seet, Q., Wongkham, S., Teh, B. T., Wongkham, C., Intapan, P.M., Maleewong, W., Yang, X., Hu, M., Wang, Z., Hofmann, A., Sternberg, P.W., Tan, P., Wang, J., Gasser, R.B., 2014. The *Opisthorchis viverrini* genome provides insights into life in the bile duct. Nat. Commun. 5, 4378.

Zarowiecki, M., Berriman, M., 2015. What helminth genomes have taught us about parasite evolution. Parasitology 142 (Suppl.), S85–S97.

Zhu, L., Liu, J., Dao, J., Lu, K., Li, H., Gu, H., Liu, J., Feng, X., Cheng, G., 2016. Molecular characterization of S. japonicum exosome-like vesicles reveals their regulatory roles in parasite-host interactions. Sci. Rep. 6, 25885.