Accepted Manuscript

The barber's pole worm CAP protein superfamily - A basis for fundamental discovery and biotechnology advances

Namitha Mohandas, Neil D. Young, Abdul Jabbar, Pasi K. Korhonen, Anson V. Koehler, Parisa Amani, Ross S. Hall, Paul W. Sternberg, Aaron R. Jex, Andreas Hofmann, Robin B. Gasser

PII: S0734-9750(15)30016-1
DOI: doi: 10.1016/j.biotechadv.2015.07.003
Reference: JBA 6955

To appear in: *Biotechnology Advances*

Received date: 26 May 2015
Revised date: 2 July 2015
Accepted date: 11 July 2015

Please cite this article as: Mohandas Namitha, Young Neil D., Jabbar Abdul, Korhonen Pasi K., Koehler Anson V., Amani Parisa, Hall Ross S., Sternberg Paul W., Jex Aaron R., Hofmann Andreas, Gasser Robin B., The barber's pole worm CAP protein superfamily - A basis for fundamental discovery and biotechnology advances, *Biotechnology Advances* (2015), doi: 10.1016/j.biotechadv.2015.07.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Biotechnology Advances
26.05.2015 (6,700 words)

The barber's pole worm CAP protein superfamily - a basis for fundamental discovery and biotechnology advances

Namitha Mohandas[a], Neil D. Young[a], Abdul Jabbar[a], Pasi K. Korhonen[a], Anson V. Koehler[a], Parisa Amani[c], Ross S. Hall[a], Paul W. Sternberg[b], Aaron R. Jex[a], Andreas Hofmann[c,*], Robin B. Gasser[a,*]

[a] *The University of Melbourne, Pathogen Genomics and Genetics Program, Parkville, Victoria, Australia*
[b] *California Institute of Technology, Division of Biology, Pasadena, California, USA*
[c] *Eskitis Institute, Griffith University, Brisbane, Australia*

* Corresponding author:
robinbg@unimelb.edu.au (R. B. Gasser) or a.hofmann@griffith.edu.au (A. Hofmann)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ABSTRACT

Parasitic worm proteins that belong to the cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1 (CAP) superfamily are proposed to play key roles in the infection process and the modulation of immune responses in host animals. However, there is limited information on these proteins for most socio-economically important worms. Here, we review the CAP protein superfamily of *Haemonchus contortus* (barber's pole worm), a highly significant parasitic roundworm (order Strongylida) of small ruminants. To do this, we mined genome and transcriptomic datasets, predicted and curated full-length amino acid sequences (n = 45), undertook systematic phylogenetic analyses of these data and investigated transcription throughout the life cycle of *H. contortus*. We inferred functions for selected *C. elegans* orthologs (including *vap-1*, *vap-2*, *scl-5* and *lon-1*) based on genetic networking and by integrating data and published information, and were able to infer that a subset of orthologs and their interaction partners play pivotal roles in growth and development via the insulin-like and/or the TGF-beta signaling pathways. The identification of the important and conserved growth regulator LON-1 led us to appraise the three-dimensional structure of this CAP protein by comparative modelling. This model revealed the presence of different topological moieties on the canonical fold of the CAP domain, which coincide with an overall charge separation as indicated by the electrostatic surface potential map. These observations suggest the existence of separate sites for effector binding and receptor interactions, and thus support the proposal that these worm molecules act in similar ways as venoms act as ligands for chemokine receptors or G protein-coupled receptor effectors. In conclusion, this review should guide future molecular studies of these molecules, and could support the development of novel interventions against haemonchosis.

Keywords:
Nematodes
Haemonchus contortus
Genome
Transcriptome
CAP (SCP/TAPS) proteins
Genetic networking
Transcription profiles
Structure models

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1. Introduction

Haemonchus contortus, also known as the barber's pole worm, is an economically important parasitic worm (helminth) of livestock worldwide (Veglia, 1915; Sutherland and Scott, 2010). This roundworm (= nematode) infects millions of sheep, goats and other ruminants worldwide, and causes disease, deaths and very substantial production losses. *H. contortus* has a life cycle that takes about three weeks (Veglia, 1915): eggs are excreted in host faeces; in the environment, single first-stage larvae (L1s) develops within eggs to then hatch (in one day) and moult through to the second- (L2s) and third-stage (L3s) larvae; the infective L3s are ingested by the host while grazing, and these develop to fourth-stage larvae (L4s) and then dioecious adults in the abomasum (stomach). The last two stages feed on blood from capillaries and cause haemonchosis (manifested in anaemia, haemorrhagic gastritis, oedema and other complications), often leading to death in severely affected animals (Sutherland and Scott, 2010).

Although a parasite-derived vaccine (Barbervax; http://barbervax.com.au/) has been released, haemonchosis control has relied heavily on the use of anthelmintic drugs. As anthelmintic resistance has become very widespread in *H. contortus* and related worms (Wolstenholme et al., 2004; Kaplan and Vidyashankar, 2012; Wolstenholme and Martin, 2014), there is an ongoing need for the development of new compounds (cf. von Samson-Himmelstjerna et al., 2005; Kaminsky et al., 2008; Little et al., 2011). Identifying new intervention targets has been challenging, particularly because of a limited understanding of the molecular biology of *H. contortus* and how it interacts with its host (Nikolaou and Gasser, 2006). Excretory/secretory (ES) molecules from *H. contortus* likely play key roles in the host interplay and the infection process, a statement supported by research revealing abundant expression of transthyretin-like (TTL) protein, peptidase inhibitor, fatty acid retinoid binding protein, lectin and sperm-coating protein (SCP)-like protein genes in the blood-feeding (haematophagous) stages of *H. contortus* (see Laing et al., 2013; Schwarz et al., 2013). In addition, *c*ysteine-rich secretory proteins, *a*ntigen 5 and *p*athogenesis-related 1 (CAP) superfamily members (Gibbs et al., 2008), typified by SCP-like extracellular domains (InterProScan codes: IPR014044 and IPR001283), are also prominent in ES products. Although the two CAP proteins *Hc*24 and *Hc*40 have been identified amongst ES molecules of *H. contortus* (see Schallig et al., 1997; Rehman and Jasmer, 1998; Yatsuda et al., 2003), there appear to be many more related molecules in this worm, according to recent evidence (Schwarz et al., 2013). This information supports the proposal of a diversified and specific involvement of CAP proteins during *H. contortus* infection and in host interplay. Despite the major socioeconomic impact of haemonchosis globally and some promise for CAP proteins as drug or vaccine targets (Cantacessi et al., 2009, 2012), there is still limited detailed information on this group of proteins for *Haemonchus*.

Although challenging in the past, detailed global molecular explorations have now become feasible because of the accessibility of high throughput nucleic acid sequencing technologies and advanced informatics (Mardis et al., 2008, 2013; Koboldt et al., 2013; Cantacessi et al., 2012, 2015). Using an Illumina sequencing approach, we recently characterized a draft genome and transcriptomes for *H. contortus* from Australia (NCBI BioProject accession no. PRJNA193158; Schwarz et al, 2013), providing an important resource to explore the systems biology of this nematode. In a preliminary assessment, we predicted a sizeable set of CAP protein-encoding genes in this genome (see Schwarz et al., 2013). In the present article, we curate full-length open reading frames (ORFs) and predict the full complement of CAP proteins, establish their genetic/phylogenetic relationships, study the transcription profiles of their genes in all key developmental stages and model these proteins *in silico*. This review should guide future molecular explorations of these

ACCEPTED MANUSCRIPT

CAP proteins and could support the design of novel intervention approaches for haemonchosis as biotechnological outcomes.

2. Methodology

We used the draft genome and transcriptomes of all key developmental stages/sexes of *H. contortus* from Australia (NCBI BioProject accession no. PRJNA193158; WormBase (www.wormbase.org) (Schwarz et al., 2013). This genome is 320 Mb and predicted to encode 23,610 proteins (Schwarz et al., 2013). The CAP protein-encoding genes were inferred using an integrative approach of evidence-based gene prediction (assembled transcripts and RNA-seq data) and *ab initio* gene prediction (MAKER and GeneMark). We employed the MAKER2 pipeline (Holt and Yandell, 2011) to predict CAP protein-encoding genes in the *H. contortus* genome assembly. Then, GeneMark-ES (Ter-Hovhannisyan et al., 2008) was used to predict the open reading frames (ORFs) of assembled transcripts. All 167,013 assembled transcripts of *H. contortus* were subjected to an homology-based search against sequences available in the public databases NCBI and InterProScan (Zdobnov et al., 2001; Quevillon et al., 2005) using tBLASTx (Altschul al., 1990) analysis (e-value cut-off: 10^{-5}). All *H. contortus* transcripts containing regions of homology (i.e. at least one CAP domain) with known CAP protein-encoding genes were extracted and mapped to *H. contortus* genomic scaffolds using BLAT (Kent, 2002). Gene predictions (Maker2 GFF file), BLAT-transcript alignments (PSL files) and RNA-seq reads mapped to genomic scaffolds (BAM files) were visually inspected and curated using the Integrative Genomics Viewer (IGV; Robinson et al., 2011). Then, we integrated the gene prediction information from multiple sources to obtain a consensus of each gene and corresponding full-length transcript using ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/). Then, protein domain searches were conducted using InterProScan (Zdobnov et al., 2001; Hunter et al., 2009), employing default parameters.

Decontaminated and quality filtered paired-end RNA-seq reads were mapped to the set of full-length transcripts using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2010). The numbers of the mapped reads per individual transcript were extracted using SAMtools (Li et al., 2009). The resultant read counts per transcript per life cycle stage were used as input data for edgeR (Robinson et al., 2010). Using an edgeR dispersion factor of zero, the genes were considered differentially transcribed if the logarithmic fold-change (FC) compared with the normalised read count data was ≥ 2 and the *P*-value was $\leq 10^{-5}$. Transcription levels were compared in a pairwise manner between different developmental stages of *H. contortus*. Genes were recorded as differentially transcribed using edgeR-calculated common and gene-wise dispersion factors, if the FC compared with the normalised read count data was ≥ 2 and the false discovery rate (FDR) was ≤ 0.05. A heat map (representing mapped reads) was produced using the heatmap.2 function in the gplots package in R (http://www.R-project.org/).

Amino acid (aa) sequences encoded by individual CAP protein genes were translated from full-length transcripts and then aligned using MUSCLE v.3.8 (Edgar, 2004). Because of substantial length differences, the predicted CAP protein sequences were divided into distinct groups based on manual alignments and patterns within CAP domains. For each group, conserved amino acid residues and PROSITE patterns (cf. Gibbs et al., 2008) were identified. These three sets of data were then individually subjected to phylogenetic analysis using Bayesian inference (BI) employing MrBayes v.3.2.2 (Ronquist and Huelsenbeck, 2003); the optimal model of sequence evolution was assessed utilising a mixed amino acid substitution model and four rate categories approximating a Γ distribution, four chains and

ACCEPTED MANUSCRIPT

1,000,000 generations, sampling every 100th generation; the first 25% of the generations sampled were removed from the analysis as burn-in.

We used the program GeneMania (Warde-Farley et al., 2010) to display predicted or experimentally validated genetic interactions of *C. elegans* orthologs encoding CAP proteins of *H. contortus*. Those results were compiled and the networks saved in scalable vector graphics (svg) format. Images were labeled and formatted using Inkscape image editor (http://www.inkscape.org/en/).

Annotation and fold appraisal was guided by results from automated model predictions performed with I-TASSER (Yang et al., 2015). Using PSIPRED (Bryson et al., 2005) secondary structure predictions for inferred *H. contortus* CAP proteins, structure-based amino acid sequence alignments were made using SBAL (Wang et al., 2012; Wang and Hofmann, 2015), and the amino acid sequence of SteCRISP was added using the crystal structure deposited in PDB entry 1rc9 (Guo et al., 2005). In preparation for comparative modelling of *Hc*-CAP-15, SteCRISP was identified as the most suitable template (*P*-value: $2x10^{-6}$) by surveying the PDB with pGenThreader (Bryson et al., 2005) for the protein with highest homology. A three-dimensional model was constructed by comparative modelling using the program MODELLER (Sali and Blundell, 1993) in two stages employing the structure of SteCRISP as a template and the structure-based amino acid sequence alignment of *Hc*-CAP-15 and SteCRISP. First, 20 independent models were computed without restraint, and the model with the lowest objective function was selected and visually inspected. The model topology around all free cysteine residues was assessed in relation to potential formation of disulphide bonds. In a second run, geometric restraints were applied to enforce the formation of disulphide bonds (Cys149-Cys159 and Cys151-Cys158). As before, from 20 independently calculated models, the one with the lowest objective function was selected.

3. The full complement of CAP proteins and their relationships

Homology-based searches against sequences in NCBI and InterProScan using tBLASTx analysis identified 1,155 transcripts, which mapped to 261 genomic scaffolds. Manual curation identified 110 genomic scaffolds that are predicted to encode proteins of the CAP superfamily (Gibbs et al., 2008). Subsequent mapping of assembled transcripts to these genomic scaffolds revealed 49 full-length and 61 partial transcripts. Of the 49 amino acid sequences encoding full-length transcripts, 45 of them contained one to three CAP domains and were predicted to represent 33 single-domain, 11 double-domain and one triple-domain CAP proteins.

The features of these 45 predicted proteins (including length, InterProscan domains and signal peptides) are summarised in Table 1. Specifically, the single-domain proteins (184-300 aa) are inferred to be considerably shorter than the double-domain proteins (380-462 aa) and differed in sequence from one another by 0-90% and 1-79%, respectively, upon pairwise comparison; the triple-domain protein identified is 700 amino acids long.

Based on the InterProScan survey, all 45 predicted proteins belong to the cysteine-rich secreted protein-related family (PTHR10334) with the allergen V5/Tpx-1 family (PR00837) and venom-allergen 5 (PR00838) signatures; 39 of these 45 proteins had the PR-1-like (SSF55797), SCP (SM00198), pathogenesis-related (G3DSA: 3.40.33.10) and CAP (PF00188) signatures (Table 1). In addition, 11 single-domain and seven double-domain proteins had an allergen V5/Tpx-1 related, conserved site CRISP-1 (PS01009). The allergen V5/Tpx-1 related, conserved site CRISP-2 (PS01010) was present in two single-domain but absent from all 11 double-domain proteins (Table 1). Moreover, 27 single-domain proteins

ACCEPTED MANUSCRIPT

and eight double-domain proteins had a signal peptide, which was absent from the other nine predicted proteins (Table 1).

Based on Prosite patterns (Table 1), we initially assigned the 45 proteins to five distinct groups: *Group 1* represented all proteins that possessed conserved CAP3 and CAP4 motifs and variable CAP1 and CAP2 motifs; *group 2* represented those that did not have a CAP4 motif and were variable in sequence within all the CAP1, CAP2 and CAP3 motifs; *group 3* contained proteins that were variable in sequence within the CAP1, CAP2 and CAP4 motifs but had the conserved CAP3 motif. *Group 4* represented all 11 double-domain proteins characterised by the conserved CAP3 motif at the C-terminal of the double-domain proteins and the variable CAP1, CAP2 and CAP4 motifs. *Hc*-CAP-44 did not have the CAP4 motif and *Hc*-CAP-41 only consisted of a partial CAP2 motif. Notably, the N-terminal moiety of the double-domain proteins was variable in sequence within all CAP motifs. The only triple-domain protein *Hc*-CAP-45 was classified as *group 5*.

A phylogenetic analysis of amino acid sequence data displayed the relationships of CAP proteins within individual groups (Fig. 1A). Specifically, within *group 1*, 12 predicted proteins with close homology to *Na*-ASP-2 and *Hc*24 (Yatsuda et al., 2003; Asojo et al., 2005a) grouped separately (pp = 0.59-1.00) to the exclusion of CAP proteins with homology to *C. elegans* LON-1 protein (*Hc*-CAP-15 and -16) and GLIPR-1 (*Hc*-CAP-17) (cf. Shaye and Greenwald, 2011). Within *group 2*, three predicted proteins (*Hc*-CAP-22, -23 and -24) that shared homology to *Ac*-ASP-7 (Datu et al., 2008; Osman et al., 2012) formed a separate clade to the exclusion of *Hc*-CAP-25. *Group 4* double-domain proteins also formed separate clusters according to homology amongst published CAP proteins; *Hc*-CAP-34 and *Hc*-CAP-35, which had homology to *Hc*40 (Rehman and Jasmer, 1998; Yatsuda et al., 2003) formed a separate clade to double-domain proteins which shared sequence homology to ASP-1 (i.e. *Hc*-CAP-36 to *Hc*-CAP-41) and VAP-2 (i.e. *Hc*-CAP-42 and *Hc*-CAP-43). With reference to the phylogenetic trees, we then proceeded to explore transcription profiles for all 45 individual genes in all life stages of *H. contortus*.

4. Transcription profiles

Significant differences in transcription were recorded among many of the 45 genes encoding CAP proteins (i.e. 33 single-, 11 double- and 1 triple-domain proteins) (Fig. 1B), of which 7 single-domain and 7 double-domain proteins were up-regulated in egg and L1-L3 stages, respectively, with respect to other developmental stages (Fig. 1B). Conversely, transcripts encoding 26 single-, 4 double- and one triple domain CAP proteins were significantly up-regulated in haematophagous stages of *H. contortus* (Fig. 1B), initially suggesting that these molecules play key roles in the infection process and/or in maintaining the parasite-host relationship. Although no direct comparison was undertaken here, this up-regulation appears to be distinct from hookworms *A. caninum*, *A. ceylanicum* and *N. americanus* (see Goud et al., 2005; Datu et al., 2008; Osman 2012; Schwarz et al., 2015), which usually undergo percutaneous, pulmonary and/or somatic migration, and establish in the small intestine rather than the abomasum. This latter difference could explain the apparent expansions of CAP protein genes in *N. americanus* (n = 128; Tang et al., 2014) and *A. ceylanicum* (n = 432; Schwarz et al., 2015) compared with *H. contortus.* The relatively high proportion (28%) of CAP molecules of *H. contortus* with homologues in *C. elegans* contrasts with the situation for *N. americanus* and *A. ceylanicum* (see Tang et al., 2014; Schwarz et al., 2015), in which only six of 137 (4.3%) CAP proteins have orthologues in the free-living nematode. In addition, there are five paralogs encoding related proteins (*Hc*-CAP-22, -23, -24, -31 and -32) in *H. contortus,* which further suggests that they relate to parasite-specific functions.

ACCEPTED MANUSCRIPT

Gender-enriched transcription was also explored in female and male haematophagous stages of *H. contortus* (Fig. 1B). Significant differences in transcription were recorded for genes encoding 29 single- and seven double-domain CAP proteins, four of which were up-regulated in female haematophagous stages (i.e. L4 and adult) (see Fig. 1B). Conversely, transcripts encoding seven single-, two double-domain and one triple-domain CAP proteins were significantly up-regulated in male haematophagous stages ($p < 0.05$), thus suggesting key roles in reproduction, interaction between the sexes and/or the parasite-host relationship. Most of the gender-enriched molecules encoded by *Hc-cap-7*, *-10*, *-22*, *-23*, *-29*, *-31*, *-32* and *-38* were common to L4 and adult stages, but absent from egg and L1-L3 stages (Fig. 1B).

5. Functional annotation based on publicly available information and genetic interactions in *C. elegans*

Of all 45 predicted CAP proteins of *H. contortus*, 13 had *C. elegans* homologs (Supplementary Table 1). Specifically, eight of the 11 double-domain proteins of *H. contortus* had homologs in *C. elegans*, including *Ce*-VAP-1 (six homologs: *Hc*-CAP-36 to *Hc*-CAP-41), *Ce*-VAP-2 (two homologs: *Hc*-CAP-42 and *Hc*-CAP-43) (Supplementary Table 1). Of the 33 single-domain proteins of *H. contortus*, five had homologs in *C. elegans* or human, namely *Ce*-SCL-5 (*Hc*-CAP-14) and *Ce*-SCL-19 (*Hc*-CAP-13), *Ce*-LON-1 (two isoforms; *Hc*-CAP-15 and *Hc*-CAP-16) and *Ce*-F57B7.2 (*Hc*-CAP-17) (Supplementary Table 1).

5.1. Single-domain proteins

Based on information available for *C. elegans* (WormBase), we propose that the LON-1 homologs, *Hc*-CAP-15 and -16, regulate body length and polyploidisation in hypodermal cells in adult and larval *H. contortus*. In *C. elegans*, the protein encoded by *lon-1* is a target of TGF-beta signalling and is expressed in hypodermal tissues; the *lon-1* gene is epistatic to *dbl-1* and negatively regulated by Sma/Mab pathway signalling (Maduzia et al., 2002; Morita et al., 2002; Tuck et al., 2014). The disruption of *lon-1* function in *C. elegans* by dsRNAi results in long worms, indicating that it is a negative regulator of body length (Morita et al., 2002). *Ce-lon-1* encodes a 312 amino acid protein with a sequence motif (GHYVQVVW) that is conserved across many metazoan organisms including *H. contortus* (e.g., Kasahara et al., 1989; Lu et al., 1993; Murphy et al., 1995; Kjeldsen et al., 1996; Morita et al., 2002; Delannoy-Normand et al., 2010). Expression studies have also shown that *Ce*-LON-1 is repressed by *Ce*-DBL-1, suggesting that *Ce*-LON-1 is a downstream component of the TGF-beta growth regulation pathway (Morita et al., 2002). LON-1 is expressed mainly in adult and larval stages (as is the case in *H. contortus*) and has dose-dependent effects on body length linked to hypodermal ploidy (cf. Morita et al., 2002).

The *lon-1* gene has numerous, mostly proven interactions with genes including *Sma-2*, *-3*, *-4* (complex), *-6*, *-9*, *-10*, *-12*, *-13*, *-14*, *-16*, *-17*, *-18* and *-19* encoding various Smad proteins (transcription factors), and *dbl-1* (see above), *daf-4*, *kin-29* (encoding a kinase involved in regulating the expression of chemosensory receptors and entry into the dauer pathway), *rnt-1*, *crm-1*, *lgg-1* and *che-2* (encodes a protein with G-protein beta-like WD-40 repeats that affects chemotaxis, longevity and dauer formation; expressed in male tail rays, some head neurons and cilia of sensory neurons; e.g., Vowels and Thomas, 1992; Bargmann et al., 1993; Fujiwara et al., 1999; Swoboda et al., 2000; Pujol et al., 2001; Tonkin et al., 2002), most of which are involved (up or down-stream) in TGF-beta signalling. Given its relative conservation between *C. elegans* and *H. contortus*, we propose that this

ACCEPTED MANUSCRIPT

complement of genes/proteins regulates body size and specification of sensory structures in the male tail, spicule formation, male tail patterning and morphogenesis, oocyte and germline quality and/or reproductive aging, and are expressed in the pharynx, intestine and/or hypodermis of *H. contortus*.

Hc-cap-17 is a close ortholog of *Ce*-F57B7.2 (isoform b) orphan gene of *C. elegans* (73% identity) and the human gene GLIPR (accession number: AAA82731; 29% identity) (Murphy et al., 1995). The latter gene encodes a member of the Glioma pathogenesis related-1 (GLIPR1) protein subfamily, also referred to as *r*elated to *t*estis *s*pecific, *v*espid and *p*athogenesis related-1 (RTVP1) proteins, being amongst the best characterized CAP subfamily members in mammals (Gibbs et al., 2008). In contrast to the germline-enriched and sex-biased expression profile (cluster A) of *Ce*-F57B7.2 (isoform b) in *C. elegans* (see Reinke et al., 2004), *Hc-cap-17* had a high and constitutive transcription in all stages of *H. contortus* studied. Therefore, it seems likely that *Hc-cap-17* plays a role in growth and development throughout the life cycle, although this gene or its product could be involved in reproductive processes. Interestingly, evidence from humans shows that GLIPR1L1 is predominantly expressed in the testis, with trace amounts in the bladder (Ren et al., 2006), although the precise cellular localization of GLIPR1L1 is presently unknown. The mouse *glipr1l1* gene also has a testis-enriched expression profile; the encoded protein, which undergoes post-translational N-linked glycosylation during spermatogenesis, localizes to the connecting piece of elongated spermatids and sperm (Gibbs et al., 2010). After sperm capacitation, mouse GLIPR1L1 has been shown to localize to the anterior area of the sperm head and has a role in the binding of sperm to the zona pellucida surrounding the oocyte (Gibbs et al., 2010). Taken together, this information suggests that, possibly together with other CAP superfamily members and other proteins, GLIPR1L1 is involved in the binding of sperm to the oocyte complex, which strengthens the proposed role of CAP proteins in cellular adhesion (cf. Gibbs et al., 2008).

Ce-scl-5, the *C. elegans* ortholog of *Hc-cap-14*, encodes an extracellular CAP protein; the *Ce-scl-5* gene is predicted to interact directly with *Ce-vap-1* and *Ce-vap-2* as well as a transcription factor enhancer encoded by the orphan gene F47B8.2 (Lee et al., 2008). Although its function is presently unknown, abundant transcription of *Hc-cap-14* in parasitic stages of *H. contortus* indicates a role in parasitism or the host-parasite interplay. Another gene, *Hc-cap-13*, is a *Ce-scl-19* ortholog encoding a cysteine-rich secretory protein Limulus clotting factor C (LCCL) domain containing 1 (HGNC: CRISPLD1) protein, with no known interaction with any other gene in *C. elegans*. Based on high levels of expression of CAP orthologs in various embryonic tissues undergoing growth and development (in ear, eye, palate and lips) of mammals (human and mouse) (Robertson et al., 1997; Delrieu et al., 2002; Ahsan et al., 2005; Bhattacharya et al., 2005; Gibbs et al., 2008; Chiquet et al., 2011), we propose that *Hc*-CAP-13 is involved in growth and development in *H. contortus*. This statement is supported by high levels of transcription in the egg, L2 and L3 stages and possibly in eggs within the female adult (Fig. 1).

5.2. Double-domain proteins

Six CAP proteins encoded by *Hc-cap-36 to Hc-cap-41* are homologs of *Ce*-VAP-1 (Supplementary Table 1), which is a secreted protein similar to the venom allergen-like (VAL) proteins found in a number of invertebrates, including parasitic nematodes, and a human homolog of cysteine-rich secretory protein 3 isoform 2 precursor (Shaye and Greenwald, 2011). The genes *Hc-cap-36* and *Hc-cap-37* usually exhibited higher transcription in the free-living stages of *H. contortus* (particularly L1 and L2) compared with parasitic (haematophagous) stages in which females displayed higher transcription

ACCEPTED MANUSCRIPT

than males. Although the transcription of *Hc-cap-38* was low in free-living stages, male haematophagous stages had higher transcription than females. For *Hc-cap-41*, highest transcription was seen in the free-living stages, particularly L1 and L2, compared with very low levels in the parasitic stages. The transcription profiles for *Hc-cap-39* and *Hc-cap-40* were similar, with highest levels in L2 and L3 and lowest levels in haematophagous (parasitic) stages. The significant expansion of VAP-1 orthologs in *H. contortus* suggests diversified functional roles of these six molecules in the parasite, which appears to be reflected in considerable differences among their transcription profiles. Based on functional data for *C. elegans*, we hypothesise that at least some of these *Hc-cap* genes are expressed specifically within amphid sheath cells and that they are involved in motility. In *C. elegans*, a *Ce-vap-1* reporter fusion is expressed specifically within such cells, and knock-down of *Ce-vap-1* results in an Unc phenotype (WBPhenotype:0000643); worms exhibit altered motility (Cronin et al., 2005). In addition, we propose that gender-enriched transcription profiles for *Hc-cap-36* and *Hc-cap-37* (females) or *Hc-cap-38* (males) in haematophagous stages relate also to reproductive and/or developmental processes. A recent molecular investigation (Lozano-Torres et al., 2014) has shown that the silencing of a gene of *Globodera rostochiensis* (potato cyst nematode) with ~35% homology to *Hc-cap-38* suppressed the activation of plant host defences, indicating a role of at least some CAP protein gene homologs in immune modulation.

Using genetic network analysis (cf. Supplementary Fig. 2), *Ce-vap-1* is inferred to interact with 17 *Ce-scl* paralogs (*Ce*-scl-1, *-2*, *-3*, *-5*, *-6*, *-7*, *-8*, *-9*, *-10*, *-11*, *-12*, *-13*, *-14*, *-15*, *-17*, *-18*, *-20*) (Lee et al., 2008). Various experiments have also shown that *Ce-scl-1* interacts (directly) with *daf-2*, *daf-16* and *age-1* (Larsen, 2003; Libina et al., 2003; Ookuma et al., 2003; Patterson, 2003; Liu et al., 2004; Zhong and Sternberg, 2006; Harada et al., 2007). Although *Ce-vap-1* and *Ce-vap-2* do not directly interact with each other, all of the 17 *Ce-scl* paralogs are predicted to interact with *Ce-vap-2* (Lee et al., 2008), which itself interacts specifically with *egl-9* (Gort et al., 2008). *Ce-scl-20* encodes an ortholog of human cysteine-rich secretory protein (HGNC:GLIPR1) (Shaye and Greenwald, 2011), which is a *p53* target gene with pro-apoptotic activity mediated by an increased production of reactive oxygen species (ROS) (Pinkston-Gosse and Kenyon, 2007; Li et al., 2008). *Ce-scl-18* and *-19* encode orthologs of human cysteine-rich secretory protein LCCL domain containing 1 (HGNC:CRISPLD1), whose function is presently unknown, but which might be involved in the parasite-host interplay and sustaining parasitism.

The two *Ce-vap-2* homologs of *H. contortus*, *Hc-cap-42 to Hc-cap-43*, encode an ortholog of human cysteine-rich secretory protein 3 (HGNC:CRISP3), which is involved in receptor-mediated endocytosis (Balklava et al., 2007). These receptors tightly bind extracellular macromolecules (ligands) that they specifically recognize; the area of the plasma-membrane with the receptor-ligand complex then undergoes endocytosis and forms a transport vesicle containing the receptor-ligand complex. Although involved in the specific uptake of particular substances (e.g., low density lipoproteins [LDLs] or iron) required by cells, endocytosis has also been linked to the transduction of signals from the periphery of cells to their nuclei (Ceresa and Schmid, 2000). The up-regulation of transcription for *Hc-cap-42 to Hc-cap-43* apparent in the L2 stage might indicate a substantial increase in nutrient uptake in the free-living environment prior to ensheathment, after which the L3 stage (consistent with *C. elegans* dauer) can no longer feed and absorb nutrients from its environment, but rather is reliant on accumulated nutrient resources. The increased transcription of both *Hc-cap-42* to *Hc-cap-43* in the female adult stage of *H. contortus* suggests an increased nutrient uptake of progeny within gravid females. This proposal is consistent with evidence for *C. elegans* showing that *Ce-vap-2* is involved in the uptake of yolk in oocytes (Grant and Hirsh, 1999; Gort et al., 2008). Experiments in *C.*

ACCEPTED MANUSCRIPT

elegans have demonstrated that knock-down of *Ce-vap-2* results in an Rme phenotype (endocytic transport defective or oocyte physiology variant; WBPhenotype:0001425), leading to a lack of yolk uptake in oocytes.

In addition, *Ce-vap-2* interacts genetically with *egl-9* (experimental evidence; Gort et al., 2008), which encodes a proline hydroxylase. Functional genomic studies have shown that *Ce-vap-2* suppresses *egl-9*, leading to egg laying (Egl) defect (http://www.wormbase.org/species/all/phenotype/WBPhenotype:0000006#07153--10; Fraser et al., 2000; Maeda et al., 2001; Kamath and Ahringer, 2003; Simmer et al., 2003; Rual et al., 2004; Sonnichens et al., 2005). The *Ce*-EGL-9 protein is known to function in a conserved hypoxia-sensing pathway to negatively regulate *Ce*-HIF-1 (hypoxia inducible factor) by hydroxylating prolyl *Ce*-HIF-1 residues (Epstein et al., 2001); EGL-9 activity is negatively regulated by its physical association with *Ce*-CYSL-1, a protein with similarity to cysteine synthases, which acts to transduce information about the environmental O_2 levels through hydrogen sulfide (H_2S) signaling (Ma et al., 2012). *Ce*-EGL-9 belongs to a superfamily of proteins, such as AlkB and leprecan, which are implicated in oxidative detoxification of alkylated bases and hydroxylation of proteins (Aravind and Koonin, 2001). *Ce*-EGL-9 is expressed in musculature, neurons and hypodermis, and is needed for muscle function for egg laying (Trent et al., 1983). Taken together, we propose that *Hc-vap-2* is centrally involved in aspects of reproduction, particularly egg-yolk uptake as well as egg laying, in intimate association with a *Ce-egl-9* ortholog and/or other complementary genes.

The independent interaction of *C. elegans* orthologs *Ce-vap-1* and *Ce-vap-2* with *Ce-scl-1*, which also interacts with parts of the insulin-like signaling pathway, including *Ce-age-1*, *Ce-daf-2* and *Ce-daf-16* (Supplementary Fig. 2) suggest also a critical and integrated involvement also in the regulation of nematode growth and development, the transition from the free-living L3 (analogous to dauer in *C. elegans*) to the parasitic stage in *H. contortus* and/or the maintenance of a balanced parasite-host relationship.

6. Structure models

In total, 35 (78%) of the 45 predicted CAP protein sequences contained an N-terminal signal peptide (Table 1), whereas the remainder did not (Table 1). These results support the existence of two main "types" of eukaryotic CAP proteins. Both secreted and non-secreted CAP proteins are characterized structurally by the presence of a conserved SCP-domain (Osman et al., 2012). In the present study, the two criteria used to classify CAP proteins were the conservation of CAP motifs and the occurrence of multiple CAP domains (see Fig. 2). Of the 45 protein sequences predicted here, the 33 single-domain proteins were classified into *group 1* (n = 16; CAP4 and CAP 3 motifs conserved), *group 2* (n = 4; CAP3 motif absent) and *group 3* (n = 13; CAP3 motif variable) proteins. Furthermore, 11 double-domain proteins represent *group 4* and one triple-domain protein was classified as *group 5*. Notably, all multi-domain CAP proteins of *H. contortus* have four intact CAP motifs and are thus inferred to adopt a CAP protein fold. This is different from the finding for hookworms, where the C-terminal moiety of some multi-domain proteins does not adopt such a fold (A. Hofmann, unpublished data). Criteria used previously for the categorization of structural groups (Osman et al., 2012) were not applicable to the prediction of CAP proteins of the barber's pole worm, since all canonical cysteine residues are highly conserved (see Supplementary Fig. 1).

Generally, the SCP-fold, harbouring the four CAP motifs, comprises an α-β-α sandwich (see Fig. 3) that might be extended at the N-, and more frequently at the C-terminal ends. Notable examples of such C-terminal extension domains include a C-type lectin, an ion channel regulator as well as the LCCL (Coch-5b2 and Lgl1) domain (Gibbs et al., 2008).

ACCEPTED MANUSCRIPT

The combination of a conserved fold with extension domains that carry particular functions, as observed for the abundant superfamily of CAP proteins, proposes a "vehicle-payload" model that ascribes a "delivery" function to the conserved SCP-fold (Osman et al., 2012). Such proteins might also be involved in enzymatic activities, recognition or binding (Borloo et al., 2013; Mason et al., 2014b).

Although, at first glance, the functional roles of conserved domains might be inferred for an individual protein of interest, three-dimensional models obtained by conventional computational modelling may not accurately reflect the structures of these molecules on an atomic level, particularly for proteins with conserved folds but rather low amino acid sequence identity (cf. Qian et al., 2011; Asojo et al., 2011; Mason et al., 2014a). Consequently, future studies investigating particular *Hc*-CAP proteins should include the experimental determination of three-dimensional structures.

The identification of the two homologs, *Hc*-CAP-15 and *Hc*-CAP-16, of *Ce*-LON-1, an important growth regulator of many metazoans, led us to construct a homology model for *Hc*-CAP-15 (*group 1*) for further appraisal. *Hc*-CAP-15 is identical in amino acid sequence to *Hc*-CAP-16, but their genes differ in sequence by 1.7% and are located on distinct genomic scaffolds. The distinctive mapping of transcripts to these two loci suggests that these genes are transcribed and expressed in different tissues or cells and have a slightly different regulation. Although the transcription profiles of these two genes are very similar across the life cycle of *H. contortus*, *Hc*-CAP-15 appears to be transcribed at higher levels in the L3 stage than *Hc*-CAP-16 (Fig. 1), suggesting a specific up-regulation in particular tissues. These two proteins were subjected to comparative modelling with SteCRISP as a template. An appraisal of the predicted three-dimensional structure (see Fig. 3) revealed that the canonical disulphide bridges are indeed conserved. Furthermore, immediately upstream of the CAP1 motif, two additional disulphide bridges are possible, but do not belong to the canonical set. These four additional cysteine residues are a LON-1 specific feature and are conserved in LON-1 sequences including that of *Ce*-LON-1. It is noted that the choice of the disulphide bridge formation, as shown in Fig. 3 (Cys149-Cys159, Cys151-Cys158), is arbitrary and another conformation (Cys149-Cys158, Cys151-Cys159) might also be possible.

The electrostatic surface potential map of *Hc*-CAP-15 shows an obvious charge separation (Fig. 3) that accompanies the separation into (at least) two topological moieties, suggesting that separate receptor docking and effector sites are highly likely on this molecule, thus resembling the structure-function relationships generally observed with chemokines (Saini et al., 2011). Such observations have been made earlier with other CAP proteins of nematodes (Asojo et al., 2005b, Osman et al., 2012), and support the notion that these parasite molecules may act as agonists or ligands for GPCRs such as chemokine receptors, similar to snake and bee toxins that also belong to the CAP superfamily (e.g. Yamazaki et al., 2003).

Interestingly, in the present study, we found no evidence of neutrophil inhibitory factor (NIF) homologs. NIFs can be quite abundant in ES products from some parasitic worms (Hewitson et al., 2009). In hookworms, for example, a NIF homolog of *Ancylostoma caninum*, SCP-1, has been reported and has been shown to bind host integrin CR3 (CD11b/CD18), leading to an inhibition of neutrophil function, including oxidative burst (Moyle et al., 1994; Rieu et al., 1996). The unexpected absence of a CAP protein with homology to NIF from *H. contortus* might suggest that other molecules assume a role akin to NIF (cf. Anbu and Joshi, 2008) in the parasitic stage of *H. contortus*, or that NIFs representing CAP proteins are most abundant in parasites that undergo extensive tissue migration or live within tissues.

ACCEPTED MANUSCRIPT

7. Concluding remarks

Surprisingly, most (82%) of the curated 45 CAP protein sequences inferred here had not been detected previously by proteomic analysis (Yatsuda et al., 2003). The relatively large number of transcripts encoding various CAP proteins in *H. contortus* compares with findings for *Ancylostoma caninum, A. ceylanicum* and *N. americanus* (cf. Hawdon et al., 1996, 1999; Datu et al., 2008; Tang et al., 2014; Schwarz et al., 2015), but contrasts with the small numbers of homologous genes inferred from transcriptomic and genomic sequence data sets for some other nematodes studied to date, including *Ascaris suum*, *Toxocara canis* and *Trichuris suis* (see Jex et al., 2011, 2014; Zhu et al., 2015). The reason(s) for this apparent difference is unclear, but could relate to differences in developmental and reproductive biology as well as varying modes of host invasion and immune modulation or evasion among nematode species (cf. Cantacessi et al., 2009). In the future, expanded data sets for other nematodes as well as improved genome assemblies (cf. Tsai et al., 2010) should allow the accurate identification and curation of CAP protein-coding genes, and enable detailed comparative studies. Such work should allow for in-depth and accurate analyses of protein features, thus assisting future functional and structural analyses of CAP protein superfamily members of nematodes.

The availability of the entire genomic and transcriptomic sequence data sets (Schwarz et al., 2013; Laing et al., 2013) and advances in functional genomics (Lok, 2012) should open up opportunities for fundamental investigations of CAP proteins and their genes in different developmental stages of *H contortus*. The present transcription profiles for genes encoding CAP proteins suggest that these molecules play diverse but critical roles in development and reproduction at different life cycle stages of this parasite.

In the future, assessing the levels of transcription and expression of CAP protein genes in the tissues of different developmental stages might improve our understanding of the function of these molecules. Moreover, the recent characterization of the gene silencing machinery for *H. contortus* (see Schwarz et al., 2013) could open up interesting opportunities for functional genomic studies of CAP protein genes in the larval stages of this parasite *in vitro*. For instance, virus-based transduction (Hagen et al., 2014; Hagen et al., 2015a,b) might be useful for the delivery of microRNA-adapted small hairpin RNAs (shRNAmirs) into *H. contortus* to achieve effective gene silencing. This area deserves attention and has the potential to overcome the inefficiency of RNAi approaches employed to date on *H. contortus* (reviewed by Knox et al., 2007; Lok, 2012). The development of an effective gene-silencing system for *H. contortus* could lead to a major and rapid advance of our understanding of orphan CAP proteins and their genes, and could help establish their actual involvement in biological and developmental pathways in this important parasite as well as parasite-host interactions at the immuno-molecular level. Such fundamental insights could pave the way to the design of new interventions, such as vaccines, against haemonchosis and other nematodiases of animals and humans. In our opinion, this focus has particular merit, given the recent success in immunizing animals against challenge infection with strongylid nematodes with selected CAP proteins (Hewitson et al., 2015; Vlamick et al., 2015).

ACCEPTED MANUSCRIPT

Acknowledgements

Research funding from the National Health and Medical Research Council and Australian Research Council is gratefully acknowledged; support from the Victorian Life Sciences Computation Initiative (VLSCI) is also acknowledged (RBG). We acknowledge the contributions of all staff at WormBase (www.wormbase.org). P.W.S. thanks the Howard Hughes Medical Institute (HHMI) and the National Institutes of Health (NIH). The authors thank Christopher Lee for some assistance with bioinformatics and the collection of some articles.

References

Ahsan M, Ohta K, Kuriyama S, Tanaka H. Novel soluble molecule, Akhirin, is expressed in the embryonic chick eyes and exhibits heterophilic cell-adhesion activity. Dev Dyn 2005;233:95-104.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ Basic Local Alignment Search Tool. J Mol Biol 1990;215:403-10.

Ali F, Brown A, Stanssens P, Timothy LM, Soule HR, Pritchard DI. Vaccination with neutrophil inhibitory factor reduces the fecundity of the hookworm *Ancylostoma ceylanicum*. Parasite Immunol 2001;23:237-49.

Aravind L, Koonin EV. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol 2001;2: research0007.1–research0007.8.

Asojo OA, Goud G, Dhar K, Loukas A, Zhan B, Deumic V, Liu S, Borgstahl GE, Hotez PJ. X-ray structure of *Na*-ASP-2, a pathogenesis-related-1 protein from the nematode parasite, *Necator americanus*, and a vaccine antigen for human hookworm infection. J Mol Biol 2005a;346:801-14.

Asojo OA, Loukas A, Inan M, Barent R, Huang J, Plantz B, Swanson A, Gouthro M, Meagher MM, Hotez PJ. Crystallization and preliminary X-ray analysis of Na-ASP-1, a multi-domain pathogenesis-related-1 protein from the human hookworm parasite *Necator americanus*. Acta Crystallogr Sect F Struct Biol Cryst Commun 2005b;61:391-4.

Asojo OA. Structure of a two-CAP-domain protein from the human hookworm parasite *Necator americanus*. Acta Crystallogr D Biol Crystallogr 2011;67:455-62.

Balklava Z, Pant S, Fares H, Grant BD. Genome-wide analysis identifies a general requirement for polarity proteins in endocytic traffic. Nat Cell Biol 2007;9:1066-73.

Bargmann CI, Hartweig EA, Horvitz HR. Odorant-selective genes and neurons mediate olfaction in *C. elegans*. Cell 1993;74:515-27.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783-95.

Bhattacharya SK, Annangudi SP, Salomon RG, Kuchtey RW, Peachey NS, Crabb JW. Cochlin deposits in the trabecular meshwork of the glaucomatous DBA/2J mouse. Exp Eye Res 2005;80:741-4.

Borloo J, Geldhof P, Peelaers I, Van Meulder F, Ameloot P, Callewaert N, Vercruysse J, Claerebout E, Strelkov SV, Weeks SD. Structure of *Ostertagia ostertagi* ASP-1: insights into disulfide-mediated cyclization and dimerization. Acta Crystallogr D Biol Crystallogr 2013;69:493-503.

Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT. Protein structure prediction servers at University College London. Nucl Acids Res 2005;33:W36-8.

Cantacessi C, Campbell BE, Visser A, Geldhof P, Nolan MJ, Nisbet AJ, Matthews JB, Loukas A, Hofmann A, Otranto D, Sternberg PW, Gasser RB. A portrait of the "SCP/TAPS" proteins of eukaryotes – developing a framework for fundamental research and biotechnological outcomes. Biotechnol Adv 2009;27:376-88.

ACCEPTED MANUSCRIPT

Cantacessi C, Campbell BE, Gasser RB. Key strongylid nematodes of animals – Impact of next-generation transcriptomics on systems biology and biotechnology. Biotechnol Adv 2012;30:469-88.

Cantacessi C, Hofmann A, Campbell BE, Gasser RB. Impact of next-generation technologies on exploring socioeconomically important parasites and developing new interventions. Methods Mol Biol 2015;1247:437-74.

Ceresa BP, Schmid SL. Regulation of signal transduction by endocytosis. Curr Opin Cell Biol 2000;12:204-10.

Chiquet BT, Henry R, Burt A, Mulliken JB, Stal S, Blanton SH, Hecht JT. Nonsyndromic cleft lip and palate: CRISPLD genes and the folate gene pathway connection. Birth Defects Res. A. Clin. Mol Teratol 2011;91:44-9.

Cronin CJ, Mendel JE, Mukhtar S, Kim YM, Stirbl RC, Bruck J, Sternberg PW. An automated system for measuring parameters of nematode sinusoidal movement. BMC Genet 2005;6:5.

Datu BJ, Gasser RB, Nagaraj SH, Ong EK, O'Donoghue P, McInnes R, Ranganathan S, Loukas A. Transcriptional changes in the hookworm, *Ancylostoma caninum*, during the transition from a free-living to a parasitic larva. PLoS Negl Trop Dis 2008;2:e130.

De Lano W. The PyMOL Molecular Graphics System (http://www.pymol.org/), 2002.

Delannoy-Normand A, Cortet J, Cabaret J, Neveu C. A suite of genes expressed during transition to parasitic lifestyle in the trichostrongylid nematode *Haemonchus contortus* encode potentially secreted proteins conserved in *Teladorsagia circumcincta*. Vet Parasitol 2010;174:106-14.

Delrieu I, Waller CC, Mota MM, Grainger M, Langhorne J, Holder AA. PSLAP, a protein with multiple adhesive motifs, is expressed in *Plasmodium falciparum* gametocytes. Mol Biochem Parasitol 2002;121, 11-20.

Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 2007;35:W522-25.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32;1792-7.

Epstein ACR, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, Mukherji M, Metzen E, Wilson MI, Dhanda A, YM T, Masson N, Hamilton DL, Jaakkola P, Barstead R, Hodgkin J, Maxwell PH, Pugh CW, Schofield CJ, Ratcliffe PJ. *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell 2001;107:43-54.

Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. Nature 2000;408:325-30.

Fujiwara M, Ishihara T, Katsura I. A novel WD40 protein, CHE-2, acts cell-autonomously in the formation of *C. elegans* sensory cilia. Development 1999;126:4839-48.

Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins – roles in reproduction, Cancer, and Immune Defence. Endocr Rev 2008;29, 865-97.

Gibbs GM, Lo JC, Nixon B, Jamsai D, O'Connor AE, Rijal S, Sanchez-Partida LG, Hearn MT, Bianco DM, O'Bryan MK. Glioma pathogenesis-related 1-like 1 is testis enriched, dynamically modified, and redistributed during male germ cell maturation and has a potential role in sperm-oocyte binding. Endocrinology 2010;151:2331-42.

Gort EH, Van Haaften G, Verlaan I, Groot AJ, Plasterk RH, Shvarts A, Suijkerbuijk KP, van Laar T, van der Wall E, Raman V, van PJD, Tijsterman M, Vooijs M. The TWIST1 oncogene is a direct target of hypoxia-inducible factor-2alpha. Oncogene 2008;27:1501-10.

Goud GN, Bottazzi ME, Zhan B, Mendez S, Deumic V, Plieskatt J, Liu S, Wang Y, Bueno L, Fujiwara R, Samuel A, Ahn SY, Solanki M, Asojo OA, Wang J, Bethony JM, Loukas A, Roy M, Hotez PJ. Expression of the *Necator americanus* hookworm larval antigen *Na*-ASP-2 in *Pichia pastoris* and purification of the recombinant protein for use in human clinical trials. Vaccine 2005;23:4754-64.

Grant B, Hirsh D. Receptor-mediated endocytosis in the *Caenorhabditis elegans* oocyte. Mol Biol Cell 1999;10:4311-26.

ACCEPTED MANUSCRIPT

Guo M., Teng M., Niu L., Liu Q., Huang Q, Hao Q. Crystal structure of the cysteine-rich secretory protein stecrisp reveals that the cysteine-rich domain has a K+ channel inhibitor-like fold. J Biol Chem 2005;280:12405-12.

Hagen J, Young ND, Every AL, Pagel CN, Schnoeller C, Scheerlinck JP, Gasser RB, Kalinna BH. Omega-1 knockdown in *Schistosoma mansoni* eggs by lentivirus transduction reduces granuloma size in vivo. Nat. Commun 2014;5:5375.

Hagen J, Scheerlinck JP, Young ND, Gasser RB, Kalinna BH 2015a. Prospects for vector-based gene silencing to explore immunobiological features of *Schistosoma mansoni*. Adv Parasitol 2015a;88:85-122.

Hagen J, Scheerlinck JP, Gasser RB. Knocking down schistosomes - promise for lentiviral transduction in parasites. Trends Parasitol. 2015b; pii: S1471-4922(15)00062-8.

Harada H, Kurauchi M, Hayashi R, Eki T. Shortened lifespan of nematode *Caenorhabditis elegans* after prolonged exposure to heavy metals and detergents. Ecotoxicol Environ Saf 2007;66:378-83.

Hawdon JM, Jones BF, Hoffman DR, Hotez PJ. Cloning and characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. J Biol Chem 1996;271:6672-8.

Hawdon JM, Narasimhan S, Hotez PJ. *Ancylostoma* secreted protein 2: cloning and characterization of a second member of a family of nematode secreted proteins from *Ancylostoma caninum*. Mol Biochem Parasitol 1999;99:149-65.

Hewitson JP, Grainger JR, Maizels RM. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. Mol Biochem Parasitol 2009;167:1-11.

Hewitson JP, Filbey KJ, Esser-von Bieren J, Camberis M, Schwartz C, Murray J, Reynolds LA, Blair N, Robertson E, Harcus Y, Boon L, Huang SC, Yang L, Tu Y, Miller MJ, Voehringer D, Le Gros G, Harris N, Maizels RM. Concerted activity of IgG1 antibodies and IL-4/IL-25-dependent effector cells trap helminth larvae in the tissues following vaccination with defined secreted antigens, providing sterile immunity to challenge infection. PLoS Pathog 2015;11:e1004676.

Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 2011;12:491.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. InterPro: the integrative protein signature database. Nucleic Acids Res 2009;37:211-15.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative Genomics Viewer. Nat Biotechnol 2011;29:24-6.

Jex AR, Liu S, Li B, Young ND, Hall RS, Li Y, Yang L, Zeng N, Xu X, Xiong Z, Chen F, Wu X, Zhang G, Fang X, Kang Y, Anderson GA, Harris TW, Campbell BE, Vlaminck J, Wang T, Cantacessi C, Schwarz EM, Ranganathan S, Geldhof P, Nejsum P, Sternberg PW, Yang H, Wang J, Wang J, Gasser RB. *Ascaris suum* draft genome. Nature 2011;479:529-33.

Jex AR, Nejsum P, Schwarz EM, Hu L, Young ND, Hall RS, Korhonen PK, Liao S, Thamsborg S, Xia J, Xu P, Wang S, Scheerlinck JP, Hofmann A, Sternberg PW, Wang J, Gasser RB. Genome and transcriptome of the porcine whipworm *Trichuris suis*. Nat Genet 2014;46:701-6.

Kamath RS, Ahringer J. Genome-wide RNAi screening in *Caenorhabditis elegans*. Methods 2003;30:313-21.

Kaminsky R, Ducray P, Jung M, Clover R, Rufener L, Bouvier J, Weber SS, Wenger A, Wieland-Berghausen S, Goebel T, Gauvry N, Pautrat F, Skripsky T, Froelich O, Komoin-Oka C, Westlund B, Sluder A, Mäser P. A new class of anthelmintics effective against drug-resistant nematodes. Nature 2008;452:176-80.

Kaplan RM, Vidyashankar AN. An inconvenient truth: global worming and anthelmintic resistance. Parasitology 2012;186:70-8.

ACCEPTED MANUSCRIPT

Kasahara M, Gutknecht J, Brew K, Spurr N, Goodfellow PN. Cloning and mapping of a testis-specific gene with sequence similarity to a sperm-coating glycoprotein gene. Genomics 1989;5:527-34.

Kent WJ. BLAT-the BLAST-like alignment tool. Genome Res 2002;12:656-64.

Kjeldsen L, Cowland JB, Johnsen AH, Borregaard N. SGP28, a novel matrix glycoprotein in specific granules of human neutrophils with similarity to a human testis-specific gene product and a rodent sperm-coating glycoprotein. FEBS Lett 1996;380:246-50.

Knox DP, Geldhof P, Visser A, Britton C. RNA interference in parasitic nematodes of animals: a reality check? Trends Parasitol 2007;23:105-7.

Knox DP. Proteases in blood-feeding nematodes and their potential as vaccine candidates. Adv Exp Med Biol 2011;712:155-76.

Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell 2013;155:27-38.

Laing R, Kikuchi T, Martinelli A, Tsai IJ, Beech RN, Redman E, Holroyd N, Bartley DJ, Beasley H, Britton C, Curran D, Devaney E, Gilabert A, Hunt M, Jackson F, Johnston SL, Kryukov I, Li K, Morrison AA, Reid AJ, Sargison N, Saunders GI, Wasmuth JD, Wolstenholme A, Berriman M, Gilleard JS, Cotton JA. The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. Genome Biol 2013;14:R88.

Larsen PL. Direct and indirect transcriptional targets of DAF-16. Sci Aging Knowledge Environ 2003;PE9.

Lee I, Lehner B, Crombie C, Wong W, Fraser AG, Marcotte EM. A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. Nat Genet 2008;40:181-8.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589-95.

Li L, Abdel Fattah E, Cao G, Ren C, Yang G, Goltsov AA, Chinault AC, Cai WW, Timme TL, Thompson TC. Glioma pathogenesis-related protein 1 exerts tumor suppressor activities through proapoptotic reactive oxygen species-c-Jun-NH2 kinase signaling. Cancer Res 2008;68:434-43.

Libina N, Berman JR, Kenyon C. Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. Cell 2003;115:489-502.

Little PR, Hodge A, Maeder SJ, Wirtherle NC, Nicholas DR, Cox GG, Conder GA. Efficacy of a combined oral formulation of derquantel-abamectin against the adult and larval stages of nematodes in sheep, including anthelmintic-resistant strains. Vet Parasitol 2011;181:180-93.

Liu T, Zimmerman KK, Patterson GI. Regulation of signaling genes by TGFbeta during entry into dauer diapause in *C. elegans*. BMC Dev Biol 2004;4:11.

Lok JB. Nucleic acid transfection and transgenesis in parasitic nematodes. Parasitology 2012;139:574-88.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res 2005;33:6494-506.

Lozano-Torres JL, Wilbers RH, Warmerdam S, Finkers-Tomczak A, Diaz-Granados A, van Schaik CC, Helder J, Bakker J, Goverse A, Schots A, Smant G. Apoplastic venom allergen-like proteins of cyst nematodes modulate the activation of basal plant innate immunity by cell surface receptors. PLoS Pathog 2014;10:e1004569.

Lu G, Villalba M, Coscia MR, Hoffman DR, King TP. Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps and yellow jackets. J. Immunol 1993;150:2823-30.

Ma DK, Vozdek R, Bhatla N, Horvitz HR. CYSL-1 interacts with the O_2-sensing hydroxylase EGL-9 to promote H2S-modulated hypoxia-induced behavioral plasticity in *C. elegans*. Neuron 2012;73:925-40.

Maduzia LL, Gumienny TL, Zimmerman CM, Wang H, Shetgiri P, Krishna S, Roberts AF, Padgett RW. *lon-1* regulates *Caenorhabditis elegans* body size downstream of the dbl-1 TGF beta signaling pathway. Dev Biol 2002;246:418-28.

ACCEPTED MANUSCRIPT

Maeda I, Kohara Y, Yamamoto M, Sugimoto A. Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. Curr Biol 2001;11:171-6.

Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet 2008;24:133-41.

Mardis ER. Next-generation sequencing platforms. Annu Rev Anal Chem 2013;6:287-303.

Mason L, Amani P, Cross M, Baker J, Bailey U, Jones MK, Gasser RB, Hofmann A. The relevance of structural biology in studying molecules involved in parasite-host interactions - potential for designing new interventions. Aust J Chem 2014a;67:1732-40.

Mason L, Tribolet L, Simon A, von Gnielinski N, Nienaber L, Taylor P, Willis C, Jones MK, Sternberg PW, Gasser RB, Loukas A, Hofmann A. Probing the equatorial groove of the hookworm protein and vaccine candidate antigen, *Na*-ASP-2. Int J Biochem Cell Biol 2014b;50:146-55.

Morita K, Flemming AJ, Sugihara Y, Mochii M, Suzuki Y, Yoshida S, Wood WB, Kohara Y, Leroi AM, Ueno N. A *Caenorhabditis elegans* TGF-beta, DBL-1, controls the expression of LON-1, a PR-related protein that regulates polyploidization and body length. EMBO J 2002;21:1063-73.

Moyle M, Foster DL, McGrath DE, Brown SM, Laroche Y, De Meutter J, Stanssens P, Bogowitz CA, Fried VA, Ely JA, Soule HR, Vlasuk GP. A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. J Biol Chem 1994; 269:10008-15.

Murphy EV, Zhang Y, Zhu W, Biggs J. The human glioma pathogenesis-related protein is structurally related to plant pathogenesis-related proteins and its gene is expressed specifically in brain tumors. Gene 1995;159:131-5.

Nikolaou S, Gasser RB. Prospects for exploring molecular developmental processes in *Haemonchus contortus*. Int. J. Parasitol 2006;36:859-68.

Ookuma S, Fukuda M, Nishida E. Identification of a DAF-16 transcriptional target gene, scl-1, that regulates longevity and stress resistance in *Caenorhabditis elegans*. Curr Biol 2003;13:427-31.

Osman A, Wang, CK, Winter, A, Loukas, A, Tribolet, L, Gasser RB, Hofmann A. Hookworm SCP/TAPS protein structure—A key to understanding host–parasite interactions and developing new interventions. Biotechnol Adv 2012;30:652-7.

Patterson GI. Aging: new targets, new functions. Curr Biol 2003;13:279-81.

Pinkston-Gosse J, Kenyon C. DAF-16/FOXO targets genes that regulate tumor growth in *Caenorhabditis elegans*. Nat Genet 2007;39:1403-9.

Preston S, Jabbar A, Nowell C, Joachim A, Ruttkowski B, Baell J, Cardno T, Korhonen PK, Piedrafita D, Ansell BR, Jex AR, Hofmann A, Gasser RB. Low cost whole-organism screening of compounds for anthelmintic activity. Int J Parasitol 2015;45:333-43.

Pujol N, Link EM, Liu LX, Kurz CL, Alloing G, Tan M-W, Ray KP, Solari R, Johnson CD, Ewbank JJ. A reverse genetic analysis of components of the Toll signaling pathway in *Caenorhabditis elegans*. Curr Biol 2001;11:809-21.

Qian W, He X, Chan E, Xu H, Zhang J. Measuring the evolutionary rate of protein-protein interaction. Proc Natl Acad Sci USA 2011;108:8725-30.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. Nucleic Acids Res 2005;33:116-20.

Rehman A, Jasmer DP. A tissue specific approach for analysis of membrane and secreted protein antigens from *Haemonchus contortus* gut and its application to diverse nematode species. Mol Biochem Parasitol 1998;97:55-68.

Reinke V, Gil IS, Ward S, Kazmer K. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. Development 2004;131:311-23.

Ren C, Ren CH, Li L, Goltsov AA, Thompson TC. Identification and characterization of RTVP1/GLIPR1-like genes, a novel p53 target gene cluster. Genomics 2006;88:163-72.

Rieu P, Sugimori T, Griffith DL, Arnaout MA. Solvent-accessible residues on the metal ion-dependent adhesion site face of integrin CR3 mediate its binding to the neutrophil inhibitory factor. J Biol Chem 1996;271:15858-61.

ACCEPTED MANUSCRIPT

Robertson NG, Skvorak AB, Yin Y, Weremowicz S, Johnson KR, Kovatch KA, Battey JF, Bieber FR, Morton CC. Mapping and characterization of a novel cochlear gene in human and in mouse: a positional candidate gene for a deafness disorder, DFNA9. Genomics 1997;46:345-54.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-40.

Ronquist F, Huelsenbeck J. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003;19:1572-4.

Rual JF, Ceron J, Koreth J, Hao T, Nicot AS, Hirozane-Kishikawa T, Vandenhaute J, Orkin SH, Hill DE, van den Heuvel S, Vidal M. Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. Genome Res 2004;14:2162-8.

Saini V, Marchese A, Tang W-J, Majetschak M. Structural determinants of ubiquitin-CXC chemokine receptor 4 interaction. J Biol Chem 2011;286:44145-52.

Sali A, Blundell T. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779-815.

von Samson-Himmelstjerna G, Harder A, Sangster NC, Coles GC. Efficacy of two cyclooctadepsipeptides, PF1022A and emodepside, against anthelmintic-resistant nematodes in sheep and cattle. Parasitology 2005;130:343-7.

Schallig HD, van Leeuwen MA, Verstrepen BE, Cornelissen AW. Molecular characterization and expression of two putative protective excretory secretory proteins of *Haemonchus contortus*. Mol Biochem Parasitol 1997;88:203-13.

Schwarz EM, Korhonen PK, Campbell BE, Young ND, Jex AR, Jabbar A, Hall RS, Mondal A, Howe AC, Pell J, Hofmann A, Boag PR, Zhu XQ, Gregory TR, Loukas A, Williams BA, Antoshechkin I, Brown CT, Sternberg PW, Gasser RB. The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. Genome Biol 2013;14:R89.

Schwarz EM, Hu Y, Antoshechkin I, Miller MM, Sternberg PW, Aroian RV. The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. Nat Genet 2015;47:416-22.

Shaye DD, Greenwald I. OrthoList: a compendium of *C. elegans* genes with human orthologs. PLoS One 2011;6:e20085.

Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PV, Kamath RS, Fraser AG, Ahringer J, Plasterk RH. Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol 2003;1:E12.

Sönnichsen B, Koski LB, Walsh A, Marschall P, Neumann B, Brehm M, Alleaume AM, Artelt J, Bettencourt P, Cassin E, Hewitson M, Holz C, Khan M, Lazik S, Martin C, Nitzsche B, Ruer M, Stamford J, Winzi M, Heinkel R, Röder M, Finell J, Häntsch H, Jones SJ, Jones M, Piano F, Gunsalus KC, Oegema K, Gönczy P, Coulson A, Hyman AA, Echeverri CJ. Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. Nature 2005;434:462-9.

Sutherland I, Scott I. Gastrointestinal Nematodes of Sheep and Cattle: Biology and Control. Wiley-Blackwell; West Sussex, UK, 2010.

Swoboda P, Adler HT, Thomas JH. The RFX-type transcription factor DAF-19 regulates sensory cilium formation in *C. elegans.* Mol Cell 2000;5:411-21.

Tang YT, Gao X, Rosa BA, Abubucker S, Hallsworth-Pepin K, Martin J, Tyagi R, Heizer E, Zhang X, Bhonagiri-Palsikar V, Minx P, Warren WC, Wang Q, Zhan B, Hotez PJ, Sternberg PW, Dougall A, Gaze ST, Mulvenna J, Sotillo J, Ranganathan S, Rabelo EM, Wilson RK, Felgner PL, Bethony J, Hawdon JM, Gasser RB, Loukas A, Mitreva M. Genome of the human hookworm *Necator americanus*. Nat Genet 2014;46:261-9.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 2008; 18:1979-90.

Tonkin LA, Saccomanno L, Morese DP, Brodigan T, Krause M, Bass BL. RNA editing by ADARs is important for normal behavior in *Caenorhabditis elegans*. EMBO J 2002;21:6025-35.

Trent C, Tsung N, Horvitz HR. Egg-laying defective mutants of the nematode *Caenorhabditis elegans*. Genetics 1983;104:619-47.

ACCEPTED MANUSCRIPT

Tsai IJ, Otto TD, Berriman M. Improving draft assemblies by iterative mapping and assemblies of short reads to eliminate gaps. Genome Biol 2010;11:R41.

Tuck S. The control of cell growth and body size in *Caenorhabditis elegans*. Exp Cell Res 2014;321:71-6.

Veglia F. The anatomy and life-history of the *Haemonchus contortus.* Vet Res 1915;4:347-500.

Vlaminck J, Borloo J, Vercruysse J, Geldhof P, Claerebout E. Vaccination of calves against *Cooperia oncophora* with a double-domain activation-associated secreted protein reduces parasite egg output and pasture contamination. Int J Parasitol 2015;45:209-13.

Vowels JJ, Thomas JH. Genetic analysis of chemosensory control of dauer formation in *Caenorhabditis elegans*. Genetics 1992;130:105-23.

Wang CK, Broder U, Weeratunga S, Gasser RB, Loukas A, Hofmann A. SBAL: a practical tool to generate and edit structure-based amino acid sequence alignments. Bioinformatics 2012;28:1026-7.

Wang CK, Hofmann A. An integrated Java tool for generating amino acid sequence alignments with mapped secondary structure elements. Biotechnology 2015;5:87-92.

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 2010;38:W214-20.

Wolstenholme AJ, Fairweather I, Prichard R, von Samson-Himmelstjerna G, Sangster NC. Drug resistance in veterinary helminths. Trends Parasitol 2004;20:469-76.

Wolstenholme AJ, Martin RJ. Anthelmintics – from discovery to resistance. Int. J. Parasitol. Drugs Drug Resist 2014;4:218-19.

Yamazaki Y, Hyodo F, Morita T. Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. Arch Biochem Biophys . 2003;412:133-41.

Yang J, Yan R, Roy R, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: Protein structure and function prediction. Nat Methods 2015;12:7-8.

Yatsuda AP, Krijgsveld J, Cornelissen AW, Heck AJ, de Vries E. Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition. J Biol Chem 2003;278:16941-51.

Zdobnov EM, Apweiler R. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics 2001;17:847-8.

Zhong W, Sternberg PW. Genome-wide prediction of *C. elegans* genetic interactions. Science 2006;311:1481-4.

Zhu XQ, Korhonen PK, Cai H, Young ND, Nejsum P, von Samson-Himmelstjerna G, Boag PR, Tan P, Li Q, Min J, Yang Y, Wang X, Fang X, Hall RS, Hofmann A, Sternberg PW, Jex AR, Gasser RB. Genetic blueprint of the zoonotic pathogen *Toxocara canis*. Nat Commun 2015;6:6145.

ACCEPTED MANUSCRIPT

Table 1. Full-length transcripts representing the different CAP protein genes (sd: single domain; dd: double domain; td: triple domain) and the genomic scaffolds to which they map, including the length of inferred amino acid sequences and InterProScan matches (including domains and motifs) for each predicted protein.

Fig. 1. (A) The phylogenetic relationships of CAP proteins predicted from the draft genome and transcriptomes of *H. contortus* based on Bayesian inference. *Groups* 1 to 5 are indicated; single-domain [sd], double-domain [dd] and triple-domain [td] proteins. The posterior probability (pp) supporting each clade is indicated. (B) Heatmap displaying transcription profiles for genes *Hc-cap-1* to *Hc-cap-45* (cf. Table 1). Transcription levels (see colour scale): low (red), medium (orange), high (yellow), very high (white).

Fig. 2. Summary of criteria for the classification of *Hc*-CAP proteins.

Fig. 3. (A) The homology model of *Hc*-CAP-15 illustrates the characteristic α-β-α sandwich fold of CAP (= SCP/TAPS) proteins. In case of *Hc*-CAP-15, this central domain is preceded by a rather extended N-terminal moiety (blue). The location of the four CAP motifs is indicated by pale colours and labels. Disulphide bonds are shown as explicit stick models, and the LON-1 specific disulphide bridges are indicated. (B) The *Hc*-CAP-15 model is shown as surface representation and colour mapped by the electrostatic surface potential (red: negative, blue: positive). The left image is the same orientation as the cartoon model in (A), the image in (B) shows the view after a 180° turn around the vertical axis. Images were rendered with PyMOL (De Lano, 2002), and the electrostatic potential was calculated using PDB2PQR and APBS (Dolinsky et al., 2007).

ACCEPTED MANUSCRIPT

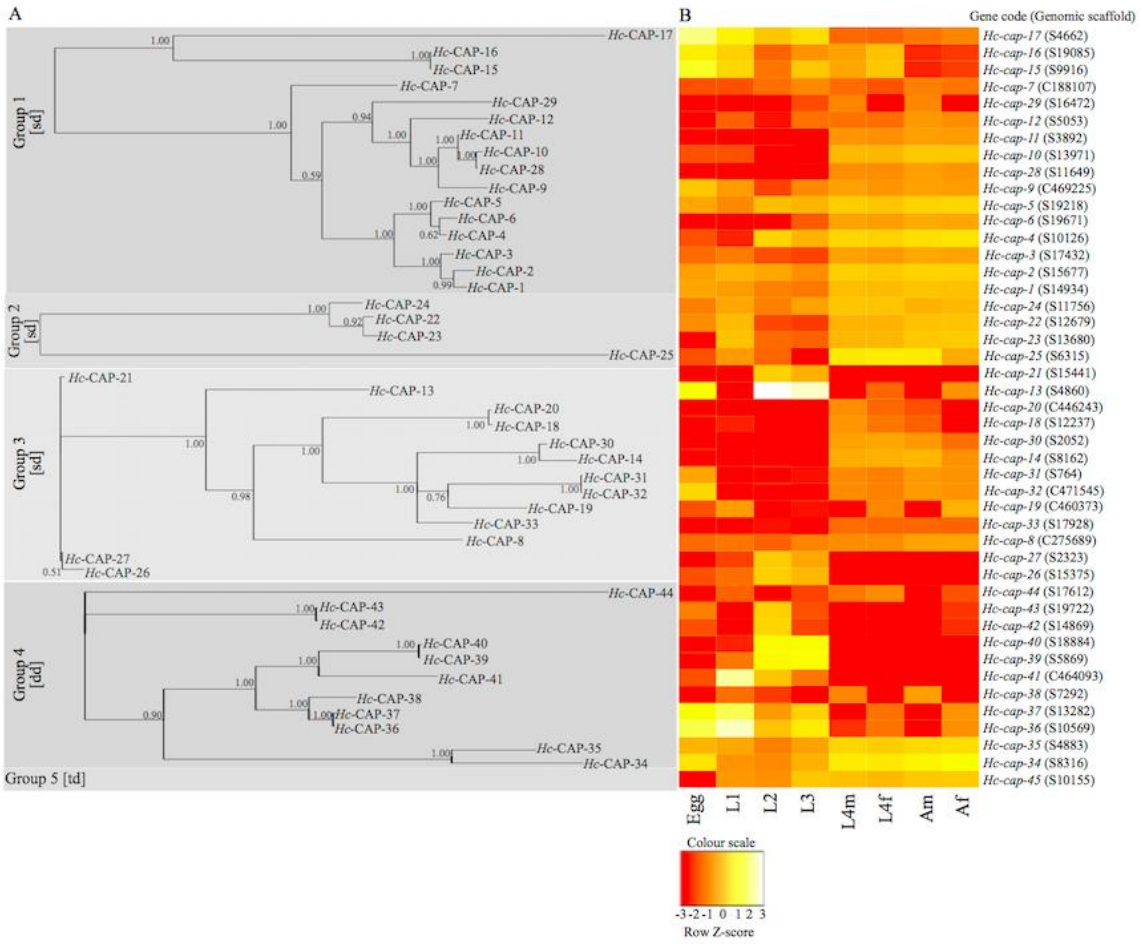
Table 1

Gene code	Genomic scaffold*	Corresponding transcripts*	Accession number	Length of inferred amino acid sequence	InterProScan matches: PTHR10334 (Cysteine-rich secretory protein family) Subfamily (SF)	PR00837 (Allergen V5/Tpx-1 family signature)	PR00838 (Venom allergen 5 signature)	SSF55797 (PR-1-like domain)	SM00198 (SCP domain)	PF00188 (CAP domain)	PS01009 (Allergen V5/Tpx-1 related, conserved site CRS P-1)	PS01010 (Allergen V5/Tpx-1 related, conserved site CRS P-2)	Signal peptide	C-region	H-region	N-region	Signal-TM	Signal-noTM	Non-cytoplasmic domain
									Length (in aa)				Signal peptide						
Single-domain																			
Hc-cap-1	S14934	Locus_4956_Transcript_2/3_Confidence_0.588_Length_534+ Locus_1281_Transcript_1/3_Confidence_0.714_Length_263	222	21-222 (201 (S F173))	71-89,150-166,122-135,176-189 (18,13,16,13)	120-135,149-168,27-43 (16,15,19)	7-196 (189)	31-189 (158)	23-213 (190)	37-181 (144)	151-161 (10)	-	1-19 (18)	15-19 (4)	4-14 (10)	1-3 (2)	1-21 (20)	1-19 (18)	20-222 (202)
Hc-cap-2	S15677	Locus_1260_Transcript_2/3_Confidence_0.500_Length_418+ Locus_11807_Transcript_1/1_Confidence_1.000_Length_365 +Locus_1281_Transcript_3/3_Confidence_0.714_Length_294	221	4-221 (217)	71-89,149-165,175-188,122-135 (18,13,16,13)	-	6-195 (189)	31-188 (157)	23-212 (189)	37-180 (143)	-	-	1-19 (18)	15-19 (4)	4-14 (10)	1-3 (2)	1-21 (20)	1-19 (18)	20-221 (201)
Hc-cap-3	S17432	Locus_1260_Transcript_3/3_Confidence_0.857_Length_412+ Locus_3258_Transcript_1/2_Confidence_0.833_Length_375	220	21-220 (199 (S F173))	69-87,120-133,174-187,148-164 (18,13,16,13)	27-43,147-166,118-133,43-63 (16,20,15,19)	6-194 (188)	31-187 (156)	23-211 (188)	37-178 (141)	149-159 (10)	-	1-19 (18)	15-19 (4)	4-14 (10)	1-3 (2)	1-21 (20)	1-19 (18)	20-220 (200)
Hc-cap-4	S10126	Locus_8857_Transcript_3/4_Confidence_0.400_Length_700	221	7-221 (214)	174-187,69-87,120-133,148-164 (18,13,16,13)	27-43,43-63,147-166 (16,20,19)	12-194 (182)	31-187 (156)	24-210 (186)	38-179 (141)	149-159 (10)	-	1-19 (18)	15-19 (4)	7-14 (7)	1-6 (5)	1-27 (26)	1-19 (18)	20-221 (201)
Hc-cap-5	S19218	Locus_2924_Transcript_2/2_Confidence_0.889_Length_203+ Locus_3595_Transcript_1/2_Confidence_0.857_Length_787+ Locus_2140_Transcript_3/3_Confidence_0.824_Length_280+ Locus_421_Transcript_5/6_Confidence_0.579_Length_367	221	4-221 (217)	174-187,69-87,148-164,120-133 (18,13,16,13)	147-166,27-43,43-63 (16,20,19)	12-194 (182)	31-187 (156)	24-210 (186)	38-179 (141)	149-159 (10)	-	1-19 (18)	15-19 (4)	3-14 (11)	1-2 (1)	-	1-19 (18)	20-221 (201)
Hc-cap-6	S19671	Locus_3220_Transcript_1/5_Confidence_0.650_Length_425+ Locus_8857_Transcript_4/4_Confidence_0.600_Length_598+ Locus_1617_Transcript_2/3_Confidence_0.500_Length_434	221	7-197 (190)	174-187,148-164,120-133,69-87 (18,13,16,13)	147-166,27-43,43-63 (16,20,19)	12-194 (182)	31-187 (156)	24-217 (193)	38-172 (134)	149-159 (10)	-	1-19 (18)	15-19 (4)	7-14 (7)	1-6 (5)	1-27 (26)	1-19 (18)	20-221 (201)
Hc-cap-7	C188107	Locus_6767_Transcript_1/1_Confidence_1.000_Length_1079	184	1-183 (183)	147-160,121-137,42-60 (18,16,13)	16-36,120-139,42-60 (20,18,19)	1-167 (166)	4-160 (154)	2-183 (181)	10-151 (141)	-	-	-	-	-	-	-	-	-
Hc-cap-8	C275689	Locus_1502_Transcript_1/2_Confidence_0.667_Length_679	226	1-226 (225)	175-188,149-165,120-133 (13,16,13)	26-42,148-167,42-62,1 18-133 (16,20,15,19)	14-195 (181)	30-188 (156)	22-208 (186)	36-180 (144)	-	-	1-18 (17)	14-18 (4)	2-13 (11)	1-2 (1)	-	1-18 (17)	19-226 (207)
Hc-cap-9	C469225	Locus_3277_Transcript_4/7_Confidence_0.476_Length_769	213	1-210 (209 (S F157))	116-129,144-160,170-183,69-87 (18,13,16,13)	143-162,27-43,43-63,69-87 (16,20,18,19)	16-190 (184)	31-183 (152)	25-210 (185)	37-173 (136)	145-155 (10)	-	1-19 (18)	15-19 (4)	3-14 (11)	1-2 (1)	1-21 (20)	1-19 (18)	20-213 (293)
Hc-cap-10	S13971	Locus_3225_Transcript_6/8_Confidence_0.216_Length_698	216	4-214 (210)	69-87,144-160,170-183 (18,16,13)	43-63,69-87,143-162,27-43 (16,20,18,19)	7-190 (183)	31-183 (152)	25-210 (187)	37-175 (138)	145-155 (10)	-	1-19 (18)	15-19 (4)	3-14 (11)	1-2 (1)	1-19 (18)	1-19 (18)	20-216 (196)
Hc-cap-11	S3892	Locus_3225_Transcript_2/8_Confidence_0.324_Length_701	225	36-222 (186)	178-191,152-168,77-95 (18,16,13)	151-170,77-95,35-51,51-71 (16,20,18,19)	29-198 (168)	39-191 (152)	33-218 (185)	45-183 (138)	153-163 (10)	-	1-19 (19)	14-19 (5)	2-13 (11)	1-2 (1)	1-19 (19)	1-19 (19)	20-224 (204)
Hc-cap-12	S5053	Locus_2107_Transcript_1/1_Confidence_1.000_Length_671	213	1-210 (209)	69-87,144-160,170-183 (18,16,13)	43-63,143-162,27-43,69-87 (16,20,18,19)	12-190 (178)	31-183 (152)	25-211 (186)	37-175 (138)	-	-	1-19 (18)	14-19 (5)	2-13 (11)	1-2 (1)	1-19 (19)	1-19 (18)	20-213 (193)
Hc-cap-13	S4860	Locus_7786_Transcript_1/1_Confidence_1.000_Length_1072	298	103-297 (194)	235-251,261-274,153-171,203-216 (18,13,16,13)	-	85-281 (196)	114-274 (160)	104-297 (193)	120-266 (146)	-	-	-	-	-	-	-	-	-
Hc-cap-14	S8162	Locus_397_Transcript_1/5_Confidence_0.714_Length_882	241	4-217 (213 (S F174))	-	[illegible]	6-217 (211)	43-210 (167)	37-217 (180)	53-202 (149)	-	-	1-23 (22)	19-23 (4)	7-18 (11)	1-6 (5)	1-32 (31)	22	24-241 (217)
Hc-cap-15	S9916	Locus_7119_Transcript_1/1_Confidence_1.000_Length_1548	297	62-255 (193)	159-175,93-111,194-207,136-149 (18,13,16,13)	-	64-205 (141)	68-207 (139)	69-246 (177)	77-199 (122)	160-170 (10)	-	1-22 (22)	16-22 (6)	4-15 (11)	1-3 (2)	1-26 (25)	1-19 (18)	23-297 (274)
Hc-cap-16	S19085	Locus_120_Transcript_1/1_Confidence_1.000_Length_1168	297	62-255 (193)	159-175,93-111,194-207,136-149 (18,13,16,13)	-	64-205 (141)	68-207 (139)	69-246 (177)	77-199 (122)	160-170 (10)	-	1-22 (21)	16-22 (6)	4-15 (11)	1-3 (2)	1-26 (25)	1-19 (18)	23-297 (274)
Hc-cap-17	S4662	Locus_8892_Transcript_2/2_Confidence_0.600_Length_1404	275	61-249 (275 (S F82))	119-137,193-209,221-234,171-184 (18,13,16,13)	169-184,192-21 1,119-137 (18,15,19)	98-232 (134)	98-234 (136)	96-247 (151)	104-224 (120)	-	-	-	-	-	-	-	-	-
Hc-cap-18	S12237	Locus_5034_Transcript_1/1_Confidence_1.000_Length_786	214	28-211 (183 (S F174))	145-161,172-185,70-88 (18,16,13)	-	16-198 (174)	32-185 (153)	23-212 (189)	40-177 (137)	-	172-183 (11)	1-18 (17)	14-18 (4)	2-13 (11)	1-2 (1)	1-19 (18)	1-18 (17)	19-214 (195)
Hc-cap-19	C460373	Locus_6436_Transcript_1/2_Confidence_0.875_Length_881	280	86-275 (189)	-	-	80-255 (175)	90-248 (158)	86-273 (187)	97-240 (143)	-	-	1-20 (19)	16-20 (4)	5-15 (10)	1-4 (3)	1-22 (21)	1-20 (19)	21-280 (259)
Hc-cap-20	C446243	Locus_2214_Transcript_1/4_Confidence_0.556_Length_785	214	28-211 (183 (S F174))	145-161,172-185,70-88 (18,16,13)	-	8-190 (182)	32-185 (153)	23-212 (189)	40-177 (137)	-	172-183 (11)	1-18 (17)	14-18 (4)	2-13 (11)	1-2 (1)	1-20 (19)	1-18 (17)	19-214 (195)
Hc-cap-21	S15441	Locus_2119_Transcript_1/2_Confidence_0.667_Length_842	240	48-239 (165)	92-110,174-190,200-213 (18,16,13)	-	43-220 (177)	53-213 (160)	47-239 (192)	59-205 (146)	-	-	1-21 (20)	17-21 (4)	5-16 (11)	1-4 (3)	-	1-21 (20)	22-240 (218)
Hc-cap-22	S12679	Locus_1028_Transcript_3/6_Confidence_0.524_Length_468+ Locus_3375_Transcript_3/11_Confidence_0.545_Length_798 +Locus_3375_Transcript_5/11_Confidence_0.582_Length_760 +Locus_2590_Transcript_3/3_Confidence_0.417_Length_464 +Locus_3697_Transcript_1/1_Confidence_1.000_Length_416	221	-	-	-	12-176 (158) - 8-170 (162)	-	39-209 (170) - 39-209 (170)	-	-	-	1-24 (23) - 1-22 (21)	18-24 (6) - 15-22 (7)	6-17 (11) - 3-14 (11)	1-5 (4) - 1-2 (1)	1-24 (23) - 1-20 (19)	1-24 (23) - 1-20 (19)	25-221 (196) - 23-221 (198)
Hc-cap-23	S13680																		
Hc-cap-24	S11756	Locus_10133_Transcript_1/1_Confidence_1.000_Length_414	226	-	-	-	14-171 (157)	-	41-210 (169)	-	-	-	1-20 (19)	16-20 (4)	6-15 (9)	1-5 (4)	1-20 (19)	1-19 (18)	21-226 (205)
Hc-cap-25	S6315	Locus_1568_Transcript_1/2_Confidence_0.750_Length_949	206	-	-	-	13-175 (162)	17-171 (154)	19-204 (185)	25-156 (131)	-	-	1-16 (15)	12-16 (4)	3-11 (8)	1-2 (1)	-	1-16 (15)	17-206 (189)
Hc-cap-26	S15375	Locus_6723_Transcript_2/4_Confidence_0.556_Length_952	236	48-213 (165)	92-110,174-190,200-213 (18,16,13)	-	47-214 (167)	53-213 (160)	47-215 (168)	59-205 (146)	-	-	1-21 (20)	17-21 (4)	5-16 (11)	1-4 (3)	-	1-17 (16)	22-236 (214)
Hc-cap-27	S2323	Locus_2119_Transcript_2/2_Confidence_0.667_Length_877	241	49-240 (191)	93-111,201-214,175-191 (18,16,13)	-	47-221 (174)	54-214 (160)	48-240 (192)	60-206 (146)	-	-	1-18 (17)	14-18 (4)	3-13 (10)	1-2 (1)	-	1-18 (17)	19-241 (222)
Hc-cap-28	S11649	Locus_3225_Transcript_3/8_Confidence_0.324_Length_698	216	4-214 (210)	69-87,144-160,170-183 (18,16,13)	43-63,69-87,143-162,27-43 (16,20,18,19)	7-190 (183)	31-183 (170)	23-210 (187)	37-175 (138)	145-155 (10)	-	1-19 (18)	15-19 (4)	3-14 (11)	1-2 (1)	1-19 (18)	1-19 (18)	20-216 (196)
Hc-cap-29	S16472	Locus_7322_Transcript_5/6_Confidence_0.500_Length_974	252	26-234 (208)	153-169,179-192 (16,13)	152-171,26-42,42-62 (16,20,19)	16-199 (183)	30-192 (162)	24-220 (196)	37-184 (147)	-	-	1-18 (17)	14-18 (4)	2-13 (11)	1-2 (1)	1-18 (17)	1-18 (17)	19-252 (233)
Hc-cap-30	S2052	Locus_397_Transcript_4/5_Confidence_0.500_Length_1051	300	95-276 (181 (S F153))	230-246,141-159,256-269 (18,16,13)	[illegible]	89-276 (187)	102-269 (167)	95-276 (181)	110-261 (151)	-	-	1-26 (25)	19-26 (7)	7-18 (11)	1-6 (5)	1-28 (27)	1-26 (25)	27-300 (273)
Hc-cap-31	S764	Locus_6381_Transcript_2/4_Confidence_0.812_Length_968	298	1-294 (293)	-	-	95-272 (177)	107-265 (158)	102-291 (189)	114-257 (143)	-	-	-	-	-	-	-	-	-
Hc-cap-32	C471545	Locus_6381_Transcript_4/4_Confidence_0.812_Length_968	273	1-294 (293)	-	-	95-272 (177)	107-265 (158)	102-291 (189)	114-257 (143)	-	-	-	-	-	-	-	-	-
Hc-cap-33	S17928	Locus_10262_Transcript_2/3_Confidence_0.750_Length_976	245	57-236 (179)	195-208,169-185,89-107 (18,16,13)	-	25-215 (190)	51-208 (151)	43-234 (191)	57-208 (143)	-	-	-	-	-	-	1-34 (33)	-	-
Double-domain																			
Hc-cap-34	S8316	Locus_3010_Transcript_2/2_Confidence_0.833_Length_256+ Locus_2113_Transcript_7/10_Confidence_0.473_Length_1050 +Locus_2676_Transcript_2/5_Confidence_0.619_Length_624	459	25-231 (206 (S F173))	72-90,125-138,152-168,413-426 (18,13,16,13)	72-90,46-66,151-170,30-46 (16,20,18,19)	9-203,252-433 (194,181)	34-196,270-426 (162, 156)	264-450,25-220 (195,186)	41-176,276-418 (135,142)	-	-	17	14-18 (4)	3-13 (10)	1-2 (1)	1-18 (17)	1-18 (17)	19-459 (440)
Hc-cap-35	S4883	Locus_3010_Transcript_1/1_Confidence_1.000_Length_194+ Locus_4810_Transcript_2/4_Confidence_0.810_Length_655+ Locus_4312_Transcript_1/3_Confidence_0.875_Length_230 + Locus_3207_Transcript_1/4_Confidence_0.857_Length_548 + Locus_4977_Transcript_1/4_Confidence_0.714_Length_377	449	31-402 (371 (S F173))	403-416,121-134,70-88,149-165 (18,13,16,13)	269-389,253-269,374-393,344-359 (16,20,15,19)	248-423, 8-195 (187,175)	257-416,32-188 (156,159)	24-211,249-447 (187,198)	38-180,263-408 (142,145)	150-160 (10)	-	1-19 (18)	13-19 (6)	4-12 (8)	1-3 (2)	-	1-19 (18)	20-449 (429)
Hc-cap-36	S10569	Locus_4740_Transcript_1/2_Confidence_0.750_Length_1438	431	1-428 (428 (S F173))	278-296,329-342,386-399,360-376 (18,13,16,13)	359-378,252-272,236-252 (16,20,19)	210-406,9-196 (187,196)	240-399,37-189 (152,159)	231-428,28-219 (191,197)	246-391,43-181 (138,145)	361-371 (10)	-	1-21 (20)	18-21 (3)	6-17 (11)	1-5 (4)	1-26 (25)	1-21 (20)	22-431 (409)
Hc-cap-37	S13282	Locus_2149_Transcript_1/1_Confidence_1.000_Length_1337 +Locus_727_Transcript_1/2_Confidence_0.833_Length_1393	431	1-428 (427 (S F173))	278-296,329-342,386-399,360-376 (18,13,16,13)	359-378,252-272,236-252 (16,20,19)	210-406,9-196 (186,184)	240-399,37-189 (152,159)	231-428,28-219 (191,197)	43-181,246-391 (138,145)	361-371 (10)	-	1-21 (20)	18-21 (3)	6-17 (11)	1-5 (4)	1-26 (25)	1-21 (20)	22-431 (409)
Hc-cap-38	S7292	Locus_4043_Transcript_1/2_Confidence_0.750_Length_1550	429	1-426 (425 (S F173))	330-343,387-400,279-297,361-377 (18,13,16,13)	360-379,237-253,253-273 (16,20,19)	19-197,213-407 (178,194)	241-400,38-190 (152, 159)	29-220,232-426 (191,194)	44-182,247-392 (138,145)	362-372 (10)	-	1-22 (21)	16-22 (6)	4-15 (11)	1-3 (2)	1-24 (23)	1-22 (23)	23-429 (406)
Hc-cap-39	S5869	Locus_5503_Transcript_2/4_Confidence_0.667_Length_1875	453	19-450 (431 (S F173))	351-364,95-113,382-398,408-421 (18,13,16,13)	258-274,381-400,274-294 (16,20,19)	251-428,27-216 (189,177)	58-209,262-421 (151,159)	253-450,50-239 (189,197)	64-201,268-413 (137, 145)	-	-	1-38 (37)	34-38 (4)	22-33 (11)	1-21 (20)	-	-	39-453 (414)
Hc-cap-40	S18884	Locus_1113_Transcript_1/1_Confidence_1.000_Length_1424	459	25-456 (431 (S F173))	101-119,357-370,388-404,414-427 (18,13,16,13)	280-300,387-406,264-280 (16,20,19)	33-222,257-434 (189,177)	268-427,64-215 (151,159)	56-245,259-456 (189,197)	70-207,274-419 (137,145)	-	-	-	-	-	-	-	-	-
Hc-cap-41	C464093	Locus_1113_Transcript_1/1_Confidence_1.000_Length_1424	462	1-390 (389 (S F173))	333-346,76-94,364-380 (18,13,16)	363-382,256-276,240-256 (16,20,19)	233-388 (178,155)	39-191,244-406 (152,162)	30-221,235-397 (191,162)	45-183,250-400 (138,150)	365-375 (10)	-	1-21 (20)	17-21 (4)	5-16 (11)	1-4 (3)	-	1-19 (18)	22-426 (404)
Hc-cap-42	S14869	Locus_2406_Transcript_1/1_Confidence_1.000_Length_1648	380	1-377 (376 (S F183))	227-245,309-325,135-348,278-291 (18,13,16,13)	308-327,184-200,227-245 (16,18,19)	3-137,181-355 (134,174)	188-348 (160)	4-161,182-377 (157,195)	194-340,9-122 (1 13,146)	310-320 (10)	-	-	-	-	-	-	-	-
Hc-cap-43	S19722	Locus_4162_Transcript_1/2_Confidence_0.750_Length_1886	394	1-391 (390 (S F183))	241-259,292-305,323-339,349-362 (18,13,16,13)	322-341,198-214,241-259 (16,18,19)	195-369,3-137 (134,174)	202-362 (160)	196-391,4-147 (143,195)	9-122,208-354 (1 13,146)	324-334 (10)	-	-	-	-	-	-	-	-
Hc-cap-44	S17612	Locus_4006_Transcript_1/2_Confidence_0.667_Length_1359	429	37-206 (167 (S F153))	-	-	229-393,22-203 (181,164)	42-192 (150)	233-421,35-220 (185,188)	48-179,236-388 (137,152)	-	-	1-20 (19)	16-20 (4)	8-15 (7)	1-7 (6)	1-20 (20)	1-20 (19)	21-429 (408)
Triple-domain																			
Hc-cap-45	S10155	Locus_6393_Transcript_2/4_Confidence_0.867_Length_614 +Locus_3225_Transcript_1/8_Confidence_0.351_Length_911 +Locus_10231_Transcript_2/3_Confidence_0.538_Length_872 +Locus_5949_Transcript_1/4_Confidence_0.500_Length_503 +Locus_3830_Transcript_1/2_Confidence_0.769_Length_516	700	23-402 (379 (S F173))	313-331,366-379,393-409,654-667 (18,13,16,13)	392-41 1,313-331,269-285 (16,18,19)	259-444,496-674,1-195 (194,185,178)	273-437,32-188,51 1-667 (156,164,156)	505-691,265-461,23-21 1 (188,196,186)	279-422,517-659,39-180 (141,143,142)	[illegible]	-	1-20 (19)	14-20 (6)	2-13 (11)	1-1 (1)	1-20 (19)	1-20 (19)	21-700 (679)

*c.f. Schwarz et al., 2013

ACCEPTED MANUSCRIPT

Fig. 1



ACCEPTED MANUSCRIPT

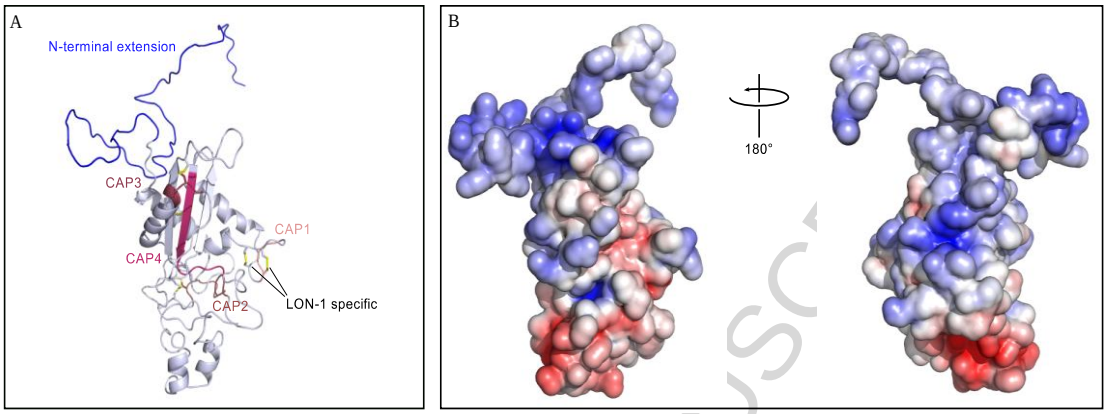
Fig. 2

Group	Domain	CAP4	CAP3	CAP1	CAP2	No. of *Hc*-CAP proteins
1		Conserved	Conserved	Variable	Variable	16
2		Variable	Absent	Variable	Variable	4
3		Conserved	Variable	Variable	Variable	13
4	N-terminus	Conserved	Variable	Variable	Variable	11
	C-terminus	Conserved	Variable	Variable	Variable	
5	N-terminus	Conserved	Variable	Variable	Variable	1
	Middle	Conserved	Conserved	Variable	Variable	
	C-terminus	Conserved	Variable	Variable	Variable	

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Fig. 3



ACCEPTED MANUSCRIPT