REVIEW ARTICLE

Platyhelminth Venom Allergen-Like (VAL) proteins: revealing structural diversity, class-specific features and biological associations across the phylum

IAIN W. CHALMERS* *and* KARL F. HOFFMANN

Institute of Biological, Environmental and Rural Sciences (IBERS), Room 2.30, Edward Llwyd Building, Penglais Campus, Aberystwyth University, Aberystwyth SY23 3DA, UK

SUMMARY

During platyhelminth infection, a cocktail of proteins is released by the parasite to aid invasion, initiate feeding, facilitate adaptation and mediate modulation of the host immune response. Included amongst these proteins is the Venom Allergen-Like (VAL) family, part of the larger sperm coating protein/Tpx-1/Ag5/PR-1/Sc7 (SCP/TAPS) superfamily. To explore the significance of this protein family during Platyhelminthes development and host interactions, we systematically summarize all published proteomic, genomic and immunological investigations of the VAL protein family to date. By conducting new genomic and transcriptomic interrogations to identify over 200 VAL proteins (228) from species in all 4 traditional taxonomic classes (Trematoda, Cestoda, Monogenea and Turbellaria), we further expand our knowledge related to platyhelminth VAL diversity across the phylum. Subsequent phylogenetic and tertiary structural analyses reveal several class-specific VAL features, which likely indicate a range of roles mediated by this protein family. Our comprehensive analysis of platyhelminth VALs represents a unifying synopsis for understanding diversity within this protein family and a firm context in which to initiate future functional characterization of these enigmatic members.

Key words: SCP/TAPS domain, Platyhelminthes, VAL, CRISP, ancylostoma secreted proteins, venom allergen-like, trematode, cestode, turbellaria.

INTRODUCTION

The phylum Platyhelminthes possesses a bewildering array of free-living, ectoparasitic and endoparasitic species amongst its 100 000 extant members (Littlewood, 2006). Within the 4 platyhelminth classes (Trematoda, Cestoda, Turbellaria and Monogenea), a range of lifestyle adaptations has developed that maximizes an individual's evolutionary success in the face of challenging ecological niches. The urgent need to develop novel drugs and vaccines for the medically and veterinary important platyhelminth species (such as schistosomes and tapeworms) has fueled an interest in the function of conserved protein families during parasitism. One protein family that is associated with platyhelminth parasitic infection processes is the Venom Allergen-Like (VAL) family, part of the larger sperm coating protein/Tpx-1/Ag5/PR-1/Sc7 (SCP/TAPS) superfamily. Here, we briefly summarize what is known about this protein family across the eukaryotes and review our current understanding into VAL diversity throughout the Platyhelminthes.

SCP/TAPS proteins

The SCP/TAPS superfamily consists of a large group of proteins all containing a distinctive 3-layer α-β-α sandwich tertiary structure domain named the SCP/TAPS domain. The presence of SCP/TAPS family members in Archeae, Eubacteria and Eukarya species suggests that this domain was present in the common ancestor of all life forms (Gibbs *et al.* 2008). Whilst the SCP/TAPS domain has yet to be ascribed an activity, several superfamily members have been characterized, providing strong evidence for the importance of these proteins in a range of biological processes.

In plants, SCP/TAPS proteins form the pathogenesis-related 1 (PR-1) family, first identified as a class of tobacco plant proteins upregulated in response to infection with tobacco mosaic virus (Loon *et al.* 1987). The PR-1 proteins have subsequently been shown to be involved in plant immune responses to a range of pathogens (Loon

* Corresponding author: Tel: +011 44 (0) 1970 621511. Fax: +011 44 (0) 1970 621981. E-mail: iwc@aber.ac.uk

et al. 1987, 2006). In *Arabidopsis thaliana*, the PR-1 proteins form a diverse family encoded by 22 distinct genes, though the precise role of the PR-1 proteins remains enigmatic (van Loon *et al*. 2006). Functional characterization of SCP/TAPS proteins is most advanced in studies involving the mammalian members. Reviewed extensively by Gibbs *et al*. (2008), research into mammalian SCP/TAPS proteins show they are associated with a diverse array of biological processes such as sperm maturation (murine CRISP1 and 2) immune responses (human CRISP3; (Udby *et al*. 2002)) and lung development (rat lgl; (Oyewumi *et al*. 2003)). Furthermore, protein interaction studies have uncovered various mammalian CRISP binding partners such as α1B-glycoprotein (Udby *et al*. 2004), β-Microseminoprotein (Udby *et al*. 2005), ryanodine receptor-type Ca^{2+} ion channels (Gibbs *et al*. 2006), mitogen-activated protein kinase kinase kinase II (Gibbs *et al*. 2007) and gametogenetin 1 (Jamsai *et al*. 2008).

In Arthropods, SCP/TAPS protein research has focused on the Antigen 5 (Ag5) proteins – one of the 3 major allergens in hornet and yellow jacket venoms (Lu *et al*. 1993). Antibody-based, cross-reactivity studies provide evidence that there is considerable antigenic similarity between the Ag5 proteins of hymenopteran (family: Vespidae) species but that anti-SCP/TAPS IgE cross-reactivity does not extend to the related fire ant (family: Formicidae) orthologue Sol i 3 (Hoffman, 1993; Lu *et al*. 1993). Another notable group of SCP/TAPS proteins within the Arthropoda are those identified in the salivary gland of haematophagous dipterans such as *Aedes aegypti* (yellow fever vector, (Valenzuela *et al*. 2002)), *Anopheles gambiae* (malaria vector, (Francischetti *et al*. 2002)), *Culex pipiens quinquefasciatus* (Bancroftian filariasis vector, (Ribeiro *et al*. 2004)) and *Glossina morsitans* (sleeping sickness vector, (Li *et al*. 2001)). Additionally, other important haematophagous arthropods such as *Triatoma brasiliensis* (order: Hemiptera, Chagas' disease vector (Santos *et al*. 2007)), *Xenopsylla cheopis* (order: Siphonaptera, human plague vector, (Andersen *et al*. 2007)) and *Ixodes scapularis* (order: Acari, Lyme disease vector, (Ribeiro *et al*. 2006)) also have salivary gland-associated SCP/TAPS transcripts. Due to the global nature of these studies, however, no information other than their sequences has been reported.

The association of SCP/TAPS proteins within parasitic arthropods is mirrored in the phylum Nematoda. Comprehensively reviewed by Cantacessi *et al*. (2009), a number of parasitic nematode species from different taxonomic clades are known to secrete SCP/TAPS proteins into the host during infection. Crucially, several of these proteins also possess immunomodulatory effects such as platelet aggregation inhibition (*Ancylostoma caninum* HPI, (Del Valle *et al*. 2003)), neutrophil chemotaxis alteration (*Necator americanus* ASP-2, (Bower *et al*. 2008)), neutrophil binding (Ac-NIF, (Moyle *et al*. 1994; Rieu *et al*. 1996)) and angiogenesis stimulation (*Onchocerca volvulus* ASP-1, (Tawe *et al*. 2000)). The importance of SCP/TAPS proteins in hookworm infections has been highlighted by a range of vaccination studies where mice, dogs and hamsters immunized with *Ancylostoma*-secreted proteins (ASPs – SCP/TAPS proteins found in soil-transmitted nematodes) were found to be partially protected against hookworm infection (Sen *et al*. 2000; Goud *et al*. 2004; Bethony *et al*. 2005). In hookworm-infected humans, IgE antibody responses to ASP-2 are negatively correlated while IgG4 levels are positively correlated with heavy worm burdens (Bethony *et al*. 2005). These data led to the belief that *N. americanus* ASP-2 would be an effective human hookworm vaccine. However, a phase I clinical trial was immediately halted when Brazilian volunteers who previously had a hookworm infection, developed IgE-dependent generalized urticaria to Na-ASP-2 immunization, demonstrating the potent allergenicity of this protein (Diemert *et al*. 2008). Further research is necessary to determine if any SCP/TAPS proteins are suitable for immunoprophylaxis.

PUBLISHED STUDIES ON PLATYHELMINTH VAL PROTEINS

Cestode VALs – McCrisp proteins

Whilst numerous SCP/TAPS proteins have been identified and characterized in the phylum Nematoda, comparably little is known about SCP/TAPS family members in the other major phylum containing worms of medical importance, the Platyhelminthes. As this review and others have highlighted (Gibbs *et al*. 2008; Cantacessi *et al*. 2009), there is a wide range of naming conventions for SCP/TAPS proteins depending on the species discussed (i.e. PR-1 proteins for plants, ASP proteins for hookworms and CRISP proteins in humans). For this review, and according to our previous naming convention (Chalmers *et al*. 2008), we have decided to refer to these platyhelminth proteins as the Venom Allergen-Like (VAL) family. The first published report of platyhelminth VAL family members originated from investigations on the cestode *Mesocestoides corti* – a mouse model for host/cestode relationships (Britos *et al*. 2007). After serendipitously discovering a VAL family member while searching for homeobox containing genes, Britos *et al*. (2007) amplified 4 different VAL transcripts from the larval parasite life stage (tetrathyridia). Due to strong sequence similarity to human CRISP proteins, the authors named these VAL transcripts McCrisp1–4 (Table 1). Of the 4 *M. corti* VAL family members, only the full-length sequence of McCrisp2

Table 1. Published findings on platyhelminth venom allergen-like proteins

Species	VAL protein[a]	Findings	Reference
		Studies on Platyhelminth VALs	
Mesocestoides corti	McCrisp1, 2, 3 & 4	- 4 different transcripts cloned from tetrathyridia lifestage. - McCrisp2 transcript localised to the apical massif in the tetrahyrida stage and localized to the proglottids in the segmented worm.	Britos *et al*. (2007)
Schistosoma mansoni	SmVAL1–28	- Division of SmVAL family into Group 1/Group 2 proteins - qPCR lifecycle profiles for SmVAL1-13 - Alternative splicing in SmVAL6	Chalmers *et al*. (2008)
Schistosoma japonicum	SjVAL1	- High IgG1 responses after day 42 of infection to recombinant protein in Infected mice. - Protein enriched in Egg E/S products, also present in cercariae	Chen *et al*. (2010)
		Platyhelminth VAL proteins identified in global proteomic studies	
S. mansoni	SmVAL6 (TC10634 & TC10635)	Proteomic identification from adult tegument sample	van Balkom *et al*. (2005)
	SmVAL4, 10 & 18 (SmSCPb, a and c)	Proteomic identification from cercarial/somule secretions	Curwen *et al*. (2006)
	SmVAL2, 3, 5 & 9	Proteomic identification from egg secretions	Cass *et al*. (2007)
	SmVAL2, 3/23, 5/15, 9, 26/28, 27 & 29	Proteomic identification from miracidial/sporocyst secretions	Wu *et al*. (2009)
	SmVAL26/28	Proteomic identification from hatching fluid and developed egg secretions	Mathieson and Wilson (2010)
	SmVAL4	Proteomic identification from cercarial Tunnels	Hansell *et al*. (2008)
S. japonicum	SjVAL1, 11, 13 & 14	Proteomic identification from egg samples	Liu *et al*. (2006)
	SjVAL1, 11, 13 & 14	Proteomic identification from Miracidial samples	
	SjVAL6 & 11	Proteomic identification from cercarial samples	
	SjVAL9 & 11	Proteomic identification from schistosomule samples	
	SjVAL1, 6, 11, 14	Proteomic identification from adult worm samples	
Opisthorchis viverrini	OvEL619323	Proteomic identification from adult worm E/S products	Mulvenna *et al*. (2010)
Schmidtea mediterranea	SmdVAL1, 2, 3, 4, 5, 6, 11, 14, 21, 31, 32, 36, 37, 38, 41, 45, 46, 49, 50	Proteomic identification from adult samples	Adamidi *et al*. (2011)

[a] Names of VAL proteins listed in 'Studies on platyhelminth VALs' section are as listed in the original publication. For the 'Platyhelminth VAL proteins identified in global proteomic studies' section, names are derived from this review's platyhelminth VAL analysis and are listed Supplementary File 1, online version only. Names used in the original publications are present in parentheses.

was determined. Analysing the full-length sequence, the authors were able to determine that McCrisp2 encoded a protein containing a signal peptide with a complete SCP/TAPS domain. Additional *in situ* hybridization experiments revealed that McCrisp2 expression was focused to the proglottids in adult worms and to the apical region (where the frontal gland develops) in tetrathyridia. This latter observation suggested that cestode VALs could be involved in host/parasite inter-relationships. Indeed,

platyhelminth VAL expression in larval secretory glands/secretions has also been discovered in several trematode species (detailed below), further supporting a role for VALs in host interactions.

Trematode VALs – SmVAL, SjVAL and OvVAL proteins

In 2006, a study examining *S. mansoni* cercarial/schistosomule excretion/secretion (E/S) products by 2-D gel electrophoresis paired with Tandem mass spectrometry (MS/MS) analysis identified 3 VAL proteins (20–25kDA in size) released by *in vitro* cultured parasites (Curwen *et al*. 2006). Now named SmVAL4, SmVAL10 and SmVAL18 (previously named SmSCP_a, SmSCP_c and SmSCP_b respectively), these were the first SCP/TAPS family proteins described in a trematode species (Table 1). Further characterization of these VAL family members was hampered at the time of publication due to the incomplete nature of the *S. mansoni* genome. However, the same research group did discover that SmVAL10 and 18 were glycosylated in a later study (Jang-Lee *et al*. 2007). In 2008, using the version 4 assembly of the *S. mansoni* genome as a reference, a comprehensive analysis of the SmVAL family was performed, identifying 28 (SmVAL1–28) genes encoding complete SCP/TAPS domains (Table 1; (Chalmers *et al*. 2008)). Using a combination of genomic, transcriptomic, phylogenetic and tertiary structure analyses, it was discovered that the SmVAL family contain 2 distinct types of SCP/TAPS proteins. Group 1 SmVALs (SmVAL1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 14, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27 and 28) contain signal peptides, 3 conserved disulphide bonds and an extended first loop region, while group 2 SmVALs (SmVAL6, 11, 13, 16 and 17) do not possess these features but do contain other unique elements such as highly conserved histidine and tyrosine residues (i.e. His21-Tyr82 in SmVAL13). It has been postulated that these conserved amino acids help to stabilize the first and third helices of group 2 SCP/TAPS domains by intramolecular hydrogen bond formation (Chalmers, 2009). Further, multi-species phylogenetic analysis has discovered that group 1 and group 2 proteins were not limited to *S. mansoni* but are present in all examined species of the Kingdom Animalia (Chalmers, 2009). Examples of group 2 proteins include Hs-GAPR-1 in humans ((Eberle *et al*. 2002), CG4270 in *Drosophila* (Kovalick and Griffin, 2005) and Ss-NIE in nematodes (Ravi *et al*. 2002). Functionally, several of the group-defining SmVAL characteristics (such as disulphide bonds) suggest different cellular localizations, with group 1 SmVALs likely to be extracellular in nature while group 2 SmVALs are enriched in intracellular compartments (Chalmers *et al*. 2008). This assertion is now supported by findings derived from several global

proteomic studies (see Table 1, (van Balkom *et al*. 2005; Curwen *et al*. 2006; Wu *et al*. 2009)).

Group 1 schistosome VALs

As previously noted, 3 group 1 VAL proteins (SmVAL4, 10 and 18) were discovered during analysis of *in vitro* cultured cercarial/schistosomule E/S products (Curwen *et al*. 2006). Importantly, SmVAL4 (the most abundantly expressed of the 3, as determined by normalized spot volume (Curwen *et al*. 2006)) was also found during an ingenious study in which parasite and host proteins were identified by liquid chromatography coupled with tandem MS (LC-MS/MS) in infection tunnels of human skin experimentally exposed to *S. mansoni* cercariae (Hansell *et al*. 2008). These collective studies, therefore, confirm that SmVAL4, 10 and 18 are all associated with mammalian host invasion. In an intriguing symmetry, proteomic studies of *S. mansoni* miracidia/sporocyst E/S products show that a different set of group 1 SmVALs are likely to be involved in molluscan parasitism (Wu *et al*. 2009). Employing an *in vitro* protocol, which mimics the transformation of free-living miracidia to snail-residing sporocyst life-cycle stages, Wu *et al*. (2009) collected the E/S products and used 1D gel electrophoreses paired with nano LC-MS/MS to identify the released proteins. Of the 99 proteins identified in this study, 5 group 1 SmVALs were conclusively identified – SmVAL2, 9, 15, 27 and the newly identified SmVAL29 (SchistoGeneDB ID, smp_120670) (Table 1; (Wu *et al*. 2009)). At least 2 other SmVAL proteins were identified in the study but due to the high level of sequence similarity between them (e.g. SmVAL3 and 23, SmVAL26 and 28), it is unclear which SmVAL was detected. Interestingly, several of these SmVALs (SmVAL2, 3, 5 and 9) were also detected in a global proteomic study of egg E/S products, indicating that some group 1 SmVAL proteins may be secreted by both egg and miracidial lifestages (Table 1; (Cass *et al*. 2007)). Further research is required to confirm whether SmVALs are actively secreted from the egg. However, as Mathieson and Wilson (2010) demonstrated, that at least 1 SmVAL (the SmVAL26/28 isoprotein, Table 1) is present in the fluid released during miracidial hatching but could not be detected in the egg E/S products (Mathieson and Wilson, 2009). Irrespective of whether SmVAL proteins are secreted from the egg or are only released after hatching or damage, the evidence above suggests that human hosts are encountering a complex set of group 1 SmVAL proteins during chronic infection (i.e. SmVAL4, 10 and 18 during cercariae invasion and SmVAL2, 3, 5, 9 26/28 during egg embolization or tissue translocation). It, therefore, remains a high priority to characterize if/how these SmVALs modulate/stimulate the mammalian immune system.

In the Asian schistosome (*S. japonicum*), initial steps have been made to address these immunological questions by studying how mice respond to the group 1 *S. japonicum* VAL-1 protein (Table 1; (Chen *et al.* 2010)). Amplified from *S. japonicum* egg cDNA, the *Sj*-VAL-1 transcript encodes a protein most closely related to SmVAL15 (58% amino acid (AA) identity). Transcript and immunolocalization studies detected *Sj*-VAL-1 in both cercariae and eggs, although expression was considerably more pronounced in the egg samples (Table 1; (Chen *et al.* 2010)). Analysis of anti-*Sj*-VAL-1 antibody responses during a chronic murine infection revealed a Th2 bias with anti-*Sj*-VAL-1 IgG$_1$ predominating (IgG$_1$ > IgG$_{2a}$) (Chen *et al.* 2010). Due to maximal *Sj*-VAL-1 production being found in the egg stage, increases in murine anti-*Sj*-VAL-1 IgG$_1$ were, unsurprisingly, correlated with the onset of schistosome egg production (5–6 weeks post-infection). Unfortunately, examination of anti-*Sj*-VAL-1 IgE was not performed in this study, so it is currently unknown whether this allergen-like protein is the target of host IgE responses similar to those found for hookworm Na-ASP-2 (Bethony *et al.* 2005). While no other members of the SjVAL family have yet been examined in detail, evidence of 5 additional SjVAL proteins (in addition to *Sj*-VAL-1) can be found by searching the Liu *et al.* (2006) proteomic dataset derived from 5 different *S. japonicum* life-cycle stages (cercariae, 2-week schistosomula, 6-week mixed sex adult worms, eggs and miracidia; see Table 1) (Liu *et al.* 2006). Outside of the *Schistosoma* genus, VAL proteins have been experimentally detected in only 1 other trematode species – the human liver fluke *Opisthorchis viverrini*. Notably, a group 1 VAL protein (GeneBank Accession EL619323) was identified in the proteomic study of E/S products released from adult *O. viverrini*. This datum suggests that *O. viverrini* group 1 VALs, similar to *Schistosoma* VALs, are also present at the mammalian host/adult parasite interface (Table 1; (Mulvenna *et al.* 2010)).

Group 2 schistosome VALs

Whilst there is growing evidence that many group 1 VAL proteins are associated with parasite secretions in trematode species (e.g. *S. mansoni*, *S. japonicum* and *O. viverrini*), information related to group 2 VAL proteins is sparse. The one exception is the highly unusual SmVAL6 – a group 2 SmVAL expressed throughout the mammalian *S. mansoni* lifestages (cercariae through adult, (Chalmers *et al.* 2008)). While other group 1 and group 2 SmVAL family members possess very few amino acids outside of the SCP/TAPS domain, SmVAL6 contains a C-terminal region of variable length and sequence (40–295AA) with no similarity to any characterized protein. Examination of the SmVAL6 gene revealed a complex structure of 34 exons (ranging from 6 to 294 bp in size) encoding the C-terminal region, which provided a template for extensive alternative splicing detected in the SmVAL6 transcripts (Chalmers *et al.* 2008). Intriguingly, the presence of 17 exons less than 20 bp in length, allied with the high level of alternative splicing over this region, suggests that the C-terminal region of SmVAL6 is related to the recently discovered Micro-Exon Gene (MEG) families (Berriman *et al.* 2009; DeMarco *et al.* 2010; Verjovski-Almeida and DeMarco, 2011).

Defined by their gene structure, which is comprised of several micro-exons (< 36 bp) flanked by conventional exons (> 36 bp) at the 5′ and 3′ ends, MEGs are exclusive to the *Schistosoma* genus with 18 separate families identified to date (DeMarco *et al.* 2010). While the function of these proteins is unknown, recent proteomic analysis has detected members of the MEG-3 family in E/S products derived from *in vitro*-transformed schistosomula and mature eggs, while members of the MEG-2 family were identified in mature egg E/S products only (DeMarco *et al.* 2010). The secretion of MEG proteins during mammalian host lifestages has led to the hypothesis that the high levels of alternative splicing in MEG transcripts is an attempt to evade the host immune response. While SmVAL6 cannot truly be classified as a MEG (due to the presence of conventional exons and a non-schistosome-specific SCP/TAPS domain), the proposal by Verjovski-Almeida and DeMarco (2011) that an SmVAL6 ancestor was formed by the combination of a MEG gene and a group 2 VAL gene is highly plausible.

As the only known MEG-like protein with a characterized domain, study of the SmVAL6 protein may well provide insight into both the function of group 2 VALs and MEG proteins. Interestingly, proteomic evidence by van Balkom *et al.* (2005) shows that SmVAL6 (referred to as TC10634 and TC10635 by the authors in the study) is found in adult worm tegumental preparations (van Balkom *et al.* 2005). However, the absence of SmVAL6 in proteomic studies examining surface tegumental membrane preparations suggests it is, like the human group 2 member Hs-GAPR-1, an intracellular protein (Braschi and Wilson, 2006; Castro-Borges *et al.* 2011). Recently, microarray analysis of different parasite tissues/regions has provided further localization data for SmVAL6, identifying the transcript to be 31-fold enriched in the female head region when compared to the whole female worm (Nawaratna *et al.* 2011). In contrast to SmVAL6, the SmVAL13 transcript, which is also a group 2 member, was found to be 14-fold enriched in the male head. Additional studies are required to shed light on the role of these different group 2 members at these locations, and to investigate whether these roles are conserved across platyhelminth species.

Monogenean and Tubellarian VALs

Currently there are no experimental studies of VAL proteins from either monogenean or turbellarian species, which limits our understanding of this family in either of these platyhelminth classes. However, a recent large-scale proteomic study of the turbellarian *Schmidtea mediterranea* provides evidence that at least 19 *S. mediterranea* VALs (SmdVALs) are present in the adult worm (Table 1; (Adamidi *et al.* 2011)). These data suggest that VAL proteins are also participating in aspects of non-parasitic platyhelminth biology. With 19 SmdVALs identified in the adult worm, potential issues of functional redundancy (especially when using RNA interference) may hamper ascertaining functions for these proteins.

As this overview suggests, research into platyhelminth VAL family members has not progressed as quickly as that performed on the nematode VAL homologues (reviewed by Cantacessi *et al.* (2009)). One of the main reasons for this has been the paucity of characterized platyhelminth genomic and transcriptomic datasets in comparison to those elucidated for the nematodes. In the last 5–10 years, however, a number of small-, medium- and large-scale platyhelminth transcriptomes (Verjovski-Almeida *et al.* 2003; Zayas *et al.* 2005; Liu *et al.* 2006; Morris *et al.* 2006; Young *et al.* 2010*a*,*b*, 2011) have been made publicly available in addition to the genomes of *S. mansoni* (Berriman *et al.* 2009), *S. japonicum*, (2009) and *S. mediterranea*, (Robb *et al.* 2008). Interrogating these datasets in a systematic fashion has facilitated the first large-scale comparative genomic/ transcriptomics/phylogenetic analysis of VAL diversity across the Platyhelminthes.

LARGE-SCALE PLATYHELMINTH VAL GENOMIC, TRANSCRIPTOMIC AND PHYLOGENETIC ANALYSES

VAL proteins are present in all classes of platyhelminth species

To identify VAL homologues from these newly-available nucleotide datasets, BLAST searches and protein domain interrogation were combined (see Table 2 legend for full description of methods), resulting in the identification of 228 complete VAL family members from 18 different platyhelminth species (Table 2; sequences excluded due to incomplete SCP/TAPS domains are listed in Supplementary File 1, online version only). Of the 59 published VAL proteins (Table 1), 56 were reassuringly found in this dataset with only McCrisp3, McCrisp4 and OvEL619323 excluded due to the incomplete nature of their respective SCP/TAPS domains. At the time this analysis was performed (11/11/2011), the *S. haematobium* genome predictions were not publicly available. Since that date, the *S. haematobium* genome was published

Table 2. *Venom allergen-like family distribution across the phylum Platyhelminthes*

Species	No. of VAL members[a]	Group 1/2[b]
Class Trematoda		
Schistosoma mansoni	29	24/5
Schistosoma japonicum	18	12/6
Schistosoma haematobium	5	1/4
Opisthorchis viverrini	16	9/7
Fasciola hepatica	12	6/6
Fasciola gigantica	7	1/6
Clonorchis sinensis	14	8/6
Class Cestoda		
Mesocestoides corti	2	2/0
Taenia asiatica	1	1/0
Taenia solium	2	2/0
Taenia saginata	1	1/0
Moniezia expansa	1	1/0
Echinococcus multilocularis	1	1/0
Class Monogenea		
Neobenedenia melleni	41	38/3
Class Turbellaria		
Dugesia japonica	5	3/2
Dugesia ryukyuensis	12	9/3
Schmidtea mediterranea	51	46/5
Macrostomum lignano	10	5/5
TOTAL	**228**	**170/58**

VAL members from platyhelminth species were identified by tBLASTn searches of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi), Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/) and Gasser Laboratory (http://gasser-research.vet.unimelb.edu.au/) EST databases and genome gene predictions for *S. mansoni*, (http://www.genedb.org/Homepage/Smansoni), *S. japonicum* (http://www.genedb.org/Homepage/Sjaponicum) and *S. mediterranea* (http://smedgd.neuro.utah.edu) using SmVAL1-29 protein sequences. All sequences with a tBLASTn e-value of <1 e-04 were then clustered to create a non-redundant dataset using a CAP3 clustering and additional pair-wise alignment interrogation (98% match over 150 bp minimum). Database searches were preformed on the 11 November 2011 (a) Number of VAL members refers to the number of unique sequences encoding a protein sequence containing at least 90% of a SCP/TAPS domain as defined by Pfam (PF00188). (b) Number of Group 1 and Group 2 members were defined by phylogenetic clustering (Fig. 2) with known SmVAL group 1 and group 2 members.

(Young *et al.* 2012), allowing a preliminary examination of VAL diversity within this species. Here, a total of 21 ShVAL genes were found using a Pfam domain search (see Supplementary file 1, online version only). However, a comprehensive analysis is required to identify the full repertoire of ShVAL diversity. Of the 21 ShVALs present in the genome Pfam list, only SHA_103186 is represented in this analysis (ShVAL6).

Examination of the platyhelminth VALs by species distribution reveals this protein family to be present across all 4 classes within the phylum (Table 2). Notably, this is the first published description of VAL family members in several of these species – *S. haematobium*, *Fasciola hepatica*, *Fasciola gigantica*, *Clonorchis sinensis* (Class: Trematoda), *Moniezia expansa*, *Echinococcus multilocularis*, *Taenia asiatica*, *Taenia solium*, *Taenia saginata* (Class: Cestoda), *Neobenedenia melleni* (Class: Monogenea), *Macrostomum lignano*, *Dugesia japonica*, *Dugesia ryukyuensis* and *S. mediterranea* (Class: Turbellaria). Interestingly, while the experimental data on platyhelminth VAL proteins (described above) have found a strong association with early events in parasite infection, the largest VAL family is present in the free-living planarian *S. mediterranea* (51 members – including the 19 identified proteomically in Adamidi *et al.* (2011)). Whereas the final number of *S. mediterranea* VALs (SmdVALs) may be amended as newer versions of the *S. mediterranea* genome are assembled and annotated, our analysis finds transcriptomic support (EST coverage over gene prediction; 98% match over 150 bp minimum) for 32 of the 51 SmdVALs, confirming that a larger protein family exists in this species than *S. mansoni* (Supplementary File 1, online version only). It is interesting to note that a recent bioinformatic study of G protein-coupled receptors (GPCRs) found that the *S. mediterranea* genome contained 4 times the number of GPCRs in comparison to the *S. mansoni* genome (Zamanian *et al.* 2011). Whether this reflects a general trend for larger gene families in free-living compared to parasitic platyhelminths needs to be further investigated by comparative genomics.

Cestodes provided the fewest numbers of VAL proteins with only 8 members identified across the 6 analysed species. This general under-representation of cestode VALs implies that fewer family members are required in these species. Caution must be made when drawing this conclusion, however, as the cestode EST databases currently available are represented by small-scale studies using few lifestages, while many VALs are known to have expression profiles tightly restricted to particular developmental forms (Chalmers *et al.* 2008). A clearer view of the cestode VAL family will undoubtedly arrive when cestode genome projects (such as *T. solium*, *E. multilocularis*, *E. granulosus* and *Hymenolepis microstoma*; reviewed by (Olson *et al.* 2011)) are published. Although the publicly available *E. multilocularis* genomic assembly (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/Echinococcus) is not annotated with gene predictions or fully assembled, a preliminary, non-exhaustive search for VAL genes identifies 5 scaffolds containing at least 5 different group 1 VAL genes (pathogen_EMU_scaffold_006139, _62143, _007768, _47586 and _007761; data not shown) and 2 different scaffolds containing

at least 4 different group 2 VAL genes (pathogen_EMU_scaffold_007285, _007768 and _008000; data not shown). One example of a probable *E. multilocularis* VAL gene is present on EMU_scaffold_008000 (1226851–1235374 bp). This gene (named EmVAL11 in this study) possesses the same structure as SmVAL11 over the SCP/TAPS regions with a 50% identity at the amino acid level (Fig. 1). The detection of group 2 VAL genes in the draft *E. multilocularis* genome is especially important as our analysis (using cestode ESTs) failed to identify a group 2 cestode VAL (Table 2). Further research is required to confirm whether EmVAL11, or any other *E. multilocularis* group 2 gene is transcribed.

Overall, the presence of large VAL families in both parasitic (e.g. *N. melleni*) and non-parasitic (e.g. *S. mediterranea*) species most likely is explained by these proteins participating in functions critical to platyhelminth life cycles, regardless of trophic strategy. Whether these functions are the same in all platyhelminth organisms is currently unknown. However, detailed interrogation of phylogenetic relationships (described below and illustrated in Fig. 2 and Supplementary File 2, online version only) indicates that conservation of function across species may differ between group 1 and group 2 VAL proteins.

Group 1/Group 2 VAL division is maintained across platyhelminth species

As first identified in the *D. melanogaster* and *S. mansoni* VAL family studies (Kovalick and Griffin, 2005; Chalmers *et al.* 2008), our phylogenetic reconstruction confirms that the major division within the platyhelminth VAL members is between group 1 and group 2 proteins (Fig. 2, Bayesian inference 100% support; Supplementary File 2, online version only; Maximum Likelihood 90% support). This division of platyhelminth group 1 or group 2 VALs is also supported by evidence from multiple sequence alignment and signal peptide analysis (summarized in Supplementary File 2, online version only).

Examination of the platyhelminth VALs showed that the vast majority of group 1 members (87%; 150/171 SCP/TAPS domains) contain all 6 disulphide bond-forming cysteines characteristic of group 1 SmVALs (C1-C6) (indicated in Supplementary File 2, online version only). These 6 cysteines were absent in all group 2 proteins analysed (Supplementary File 2, online version only), as previously found for group 2 SmVALs. Signal peptide analysis confirmed that the presence of signal peptides was, as found by Chalmers *et al.* (2008) in the SmVAL family, to be characteristic of group 1 VALs with a majority (74%) of the platyhelminth group 1 proteins encoding a signal peptide (as defined by SignalP 3.0 Neural
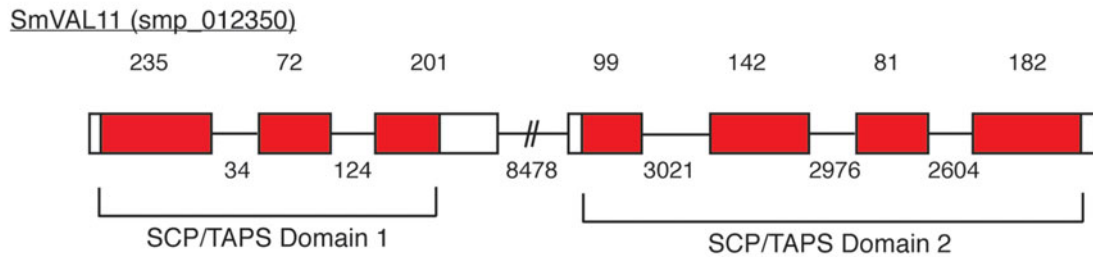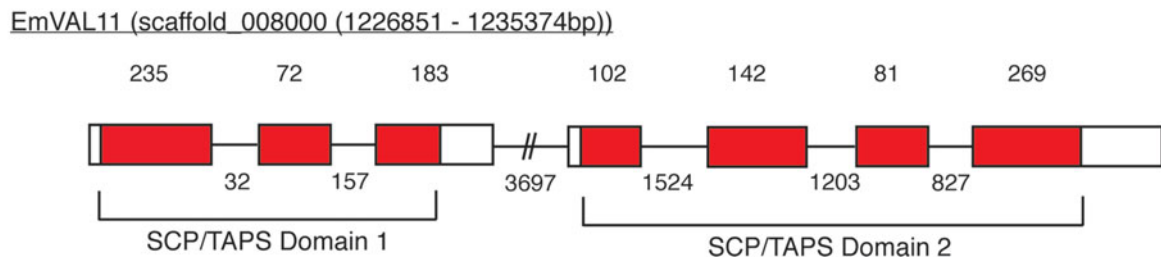
Fig. 1. Comparison of *Schistosoma mansoni* and *Echinococcus multilocularis* VAL11 gene structure. (A) SmVAL11 gene structure over the SCP/TAPS domain-encoding exons. Structure obtained from *S. mansoni* genome v. 4 (http://www.genedb.org/Homepage/Smansoni). (B) EmVAL11 gene structure over the SCP/TAPS domain-encoding exons. The genomic region (scaffold_008000) was identified by a tBLASTn search of the *E. multilocularis* genome (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/Echinococcus) using SmVAL11. EmVAL11 gene structure was manually predicted with all exon/intron junctions conforming to the consensus (GT/AG) splice donor/acceptor sequences for eukaryotes. Exons are represented by boxes with the length shown in base pairs above. Introns are represented by lines with the length shown in base pairs below. Exon regions coloured red represent regions encoding the SCP/TAPS domain.

Network analysis, using default *Dscore* threshold). The prevalence of this feature was similar across all 4 taxonomic classes (Trematoda – 78%, Cestoda – 87%, Monogenea – 68% and Turbellaria – 71%; Supplementary File 1, online version only). In contrast to this result, not one group 2 VAL encoded a signal peptide, indicating that these members are likely to be found as intracellular proteins.

Group 1 VALs are restricted to class-specific clades

One of the most notable findings from phylogenetic inspection is the strong evidence for multiple group 1 class-specific VAL clades (Fig. 2). For example, 7 of the 8 group 1 cestode VALs (McCrisp3, McCrisp2, TsVAL1, TsVAL2, TsgVAL1, TaVAL1 and MeVAL1) form a single, cestode-specific clade (Fig. 2; cestode VALs highlighted yellow). Further interrogation of all group 1 VALs demonstrates that this observation is ubiquitous across the phylum with 92% of family members (157/171 SCP/TAPS domains) contained within class-specific clades (Fig. 2), thus having no clear orthologue outside of that taxonomic class. Within the turbellarian group 1 VALs, taxonomic subdivisions are also reflected, with 43 of the 55 VALs from *Dugesidea* species (*D. japonica*, *D. ryukyuensis* and *S. mediterranea*) present in a single clade (highlighted blue in Fig. 2),

while the distantly related *Macrostomum lignano* VALs are present in additional species-specific clades. Within the trematodes (Fig. 2; coloured red), all group 1 schistosome VALs are present in class-specific clades with the exception of SmVAL20, which does not cluster within any clade. Interestingly, the 3 cercarial/schistosomal E/S SmVALs (4, 10 and 18) form a distinct clade (along with SmVAL19) lacking orthologues from other species (Fig. 2, posterior probability support 0·82; Maximum Likelihood 53% support). This finding provides molecular evidence for potential species specificity in these mammalian-associated, invasion proteins. Monogenean group 1 VALs also showed clear class specificity with 79% of the *N. melleni* VALs clustering into class-specific clades. Of the 171 group 1 SCP/TAPS domains examined, only 3 clustered in a non-class-specific clade – NmVAL4 (*N. melleni*; Monogenea), SmdVAL4 (*S. mediterranea*; Turbellaria) and DrVAL12 (*D. ryukyuensis*; Turbellaria) (posterior probability score 0·79; Fig. 2). This clade, however, was not observed by Maximum Likelihood analysis (Supplementary File 2, online version only), casting doubt on the relationship of these 3 proteins.

In contrast to the divergent relationship amongst group 1 proteins (i.e. class-specific members), the platyhelminth group 2 proteins are more highly conserved across the phylum. Phylogenetic analysis
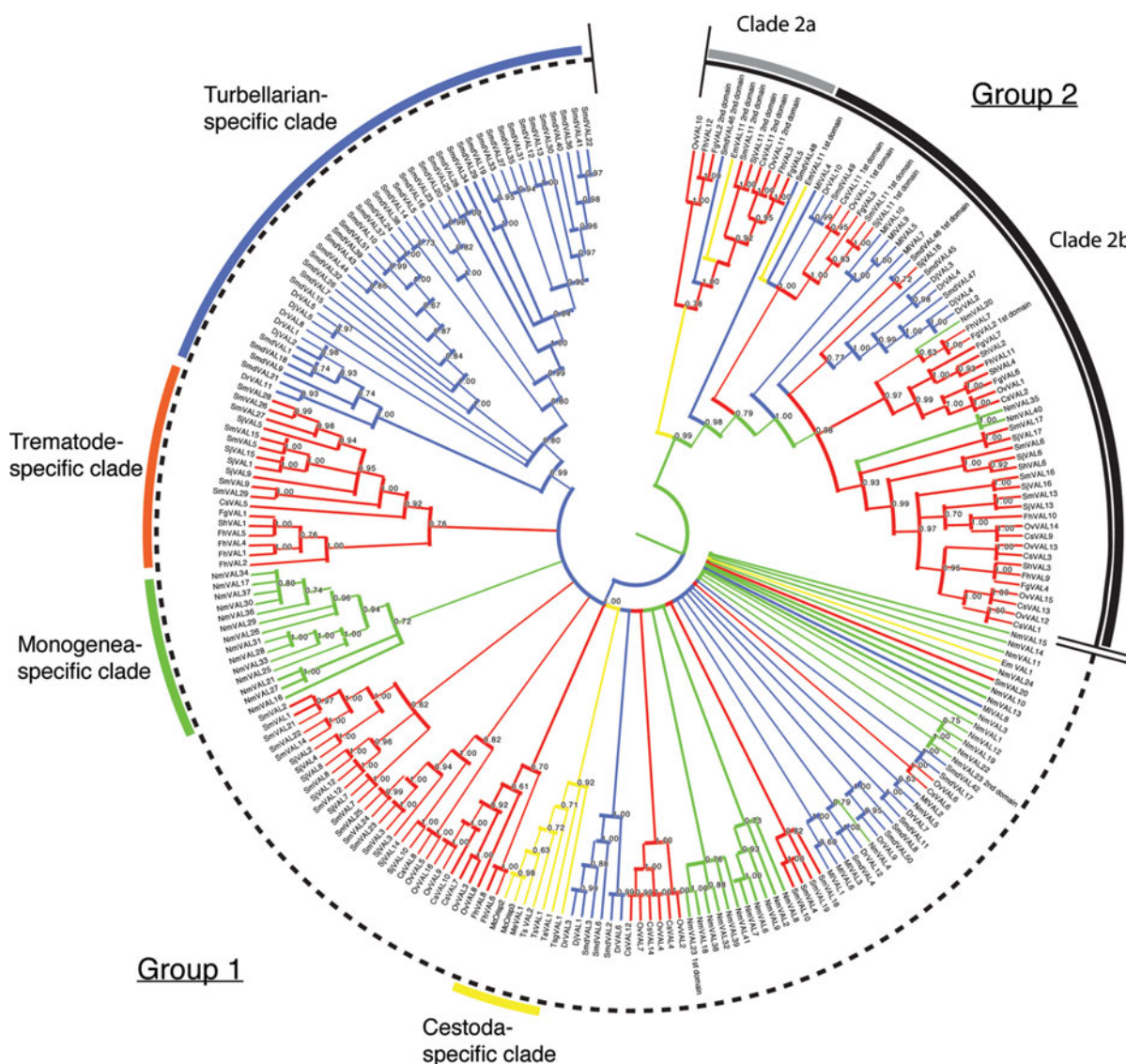
Fig. 2. Phylogenetic analysis of platyhelminth VAL proteins. In total, 237 platyhelminth SCP/TAPS domain amino acid sequences were aligned using ClustalW (Larkin *et al.* 2007) with Bayesian inference phylogenetic analysis performed using MrBayes software (version 3.1.2, WAG protein substitution model used, 3×10^6 generations run). The resulting unrooted consensus phylogenetic tree was visualized using Mesquite software. Branches are coloured to indicate the taxonomic class each sequence derives from: Trematoda (red), Cestoda (yellow), Turbellaria (blue) or Monogenea (green). Group 1 (dashed black line) and Group 2 proteins (solid black line) are indicated, as are the 2 major group 2 clades – Clade 2a (light grey line) and 2b (dark grey line). Examples of class-specific group 1 clades are highlighted red (trematode-specific), yellow (cestode-specific), green (monogenean-specific) or blue (turbellarian-specific) depending on the taxonomic class. Bayesian posterior probability support values greater than 0·6 are indicated. Species identifiers are as follows; *Schistosoma mansoni* (Sm), *Schistosoma japonicum* (Sj), *Schistosoma haematobium* (Sh), *Opisthorchis viverrini* (Ov), *Fasciola hepatica* (Fh), *Fasciola gigantica* (Fg), *Clonorchis sinensis* (Cs), *Mesocestoides corti* (Mc), *Taenia asiatica* (Ta), *Taenia solium* (Ts), *Taenia saginata* (Tsg), *Moniezia expansa* (Me), *Echinococcus multilocularis* (Em), *Neobenedenia melleni* (Nm), *Dugesia japonica* (Dj), *Dugesia ryukyuensis* (Dr), *Schmidtea mediterranea* (Smd) and *Macrostomum lignano* (Ml).

of the 65 group 2 SCP/TAPS domains provides strong support (Fig. 2, posterior probability score 0·99; Maximum Likelihood, 81% support) for at least 2 major clades within the Platyhelminthes – Clade 2a and 2b (Fig. 2, highlighted in grey (2a) and black (2b)). The presence of turbellarian, cestode and trematode members in both clades provides evidence that these two group 2 clades diverged early in platyhelminth evolution and have both been

maintained across taxa. Published genomic structure analysis of the group 2 SmVALs (Chalmers *et al.* 2008) supports this early divergence, finding different intron boundary positions over exons encoding the N-terminal and C-terminal SCP/TAPS domain. Of the two clades, Clade 2b contains the vast majority of the group 2 SCP/TAPS domains, while Clade 2a contains only 11 members – 9 from trematode species, 1 from the *E. multilocularis* genome and 1

from the turbellarian *S. mediterranea*. Interestingly, all of the double SCP/TAPS domain group 2 proteins identified (SmVAL11, SjVAL11 CsVAL11, OvVAL11, FgVAL2, EmVAL11 and SmdVAL46) possess a Clade 2a N-terminal SCP/TAPS domain and a Clade 2b C-terminal SCP/TAPS domain. These 7 double-domain VALs from 7 different species are highly likely to represent orthologous proteins. Given the early divergence of these two group 2 domain types, it is likely that each domain type possesses a different function. Double SCP/TAPS domain VALs such as SmVAL11 (Fig. 1), therefore, would possess 2 different functions mediated through the different SCP/TAPS domains.

In addition to an SmVAL11 orthologue (SjVAL11; 89% amino acid identity over N-terminal SCP/TAPS domain, 82% ID for C-terminal SCP/TAPS domain), the *S. japonicum* genome also contains orthologues for all group 2 SmVALs – SmVAL6 (SjVAL6; 90% ID), SmVAL13 (SjVAL13; 76% ID), SmVAL16 (SjVAL16; 93% ID) and SmVAL17 (SjVAL17; 85% ID). Surprisingly, one group 2 SjVAL (SjVAL18) does not appear to have an *S. mansoni* orthologue. Derived from 2 *S. japonicum* ESTs (AY811609 and BU780182), the SjVAL18 transcript has no gene prediction in the current *S. japonicum* genome and must therefore be viewed with caution (see Supplementary File 1, online version only). In contrast to the group 2 VALs, no clear orthologues can be ascertained for a number of group 1 SmVALs (i.e. SmVAL4, 10, 18, 19 and 20). Further, where group 1 orthologues are identified, the percentage amino acid identities between the Sj and SmVAL members is also consistently lower than those observed in the group 2 analysis, with only SjVAL5's similarity to SmVAL28 above 80% (summarized in Supplementary File 1, online version only).

The class- and species- specificity of group 1 platyhelminth VALs (in comparison to the group 2 proteins) indicates that these particular SCP/TAPS members undergo rapid evolutionary changes. High levels of divergence between SCP/TAPS families from related species is not unprecedented. For example, only 1 potential orthologue was detected in a phylogenetic comparison of the *Arabidopsis thaliana* (22 family members) and rice (*Oryza sativa*; 32 members) PR-1 family (van Loon *et al*. 2006). This conservation level is very low in comparison to other gene families such as the serine protease proteins, where nearly 40% of *A. thaliana* members have identifiable orthologues in rice (Tripathi and Sowdhamini, 2006). The authors explained the near-complete, non-overlap of SCP/TAPS members as being due to gene duplication/gene loss and sequence evolution after the divergence of these 2 species. As with *Arabidopsis* PR-1 family, the *S. mansoni* genome contains evidence of local gene duplication events expanding the gene repertoire, with clusters of group

1 SmVAL genes present in particular chromosomal regions (Chalmers *et al*. 2008). Detailed evolutionary studies are required to address whether gene duplication/loss or sequence divergence are driving the differences observed in this study. If the group 1 platyhelminth VALs are indeed rapidly changing in amino acid sequence, this may support the view that a key role of the SCP/TAPS domain is providing a structural scaffold for functions performed by residues on the loop regions, glycans and/or additional domains N-terminal or C-terminal to the SCP/TAPS domain (Gibbs *et al*. 2008). If VAL functional residues are not present in the core SCP/TAPS fold, considerable sequence variation found here would not affect function. Alternatively, as many group 1 VALs are likely to function after excretion/secretion into the environment, the protein differences between species could reflect the co-evolution of these proteins with specific environmental interacting partners (e.g. host proteins for parasitic platyhelminths).

Distinct protein domains are found within Group 1 VAL C-terminal regions

While the phylogenetic analysis focused only on the SCP/TAPS domain regions, comparison of the platyhelminth VALs outside of the SCP/TAPS domain identified further differences in protein structure amongst taxonomic classes. Similar to SCP/TAPS proteins from other phyla (e.g. PR-1 proteins and Hs-GAPR-1), the majority (98%; 223/228) of the platyhelminth VALs encode no protein domains other than an SCP/TAPS domain (as determined by Pfam searches). Only 5 transcripts were found to encode other protein domains; 4 group 1 turbellarian VALs (DrVAL12, DrVAL9, SmdVAL8 and SmdVAL4) encoded a fibronectin type 2 domain (FN2; PF00040) and one group 1 monogenean VAL (NmVAL27) encoded 3 low-density lipoprotein receptor domains (LDL; PF00057) C-terminal to the SCP/TAPS domain (Fig. 3). The identification of FN2 domains in 4 tubellarian VALs is unusual as proteins containing FN2 domains are thought to only be present in vertebrate species (Ozhogina *et al*. 2001). From the published literature, invertebrates should only contain the ancestor of the FN2 domain, the Kringle domain (PF00015) (Ozhogina and Bominaar, 2009). However, the 4 FN2 regions found within the turbellarian VALs conform in both size and composition (i.e. conserved residues) to the FN2 domain (data not shown). If these turbellarian VALs do contain functional FN2 domains, then this would indicate a role for these proteins in collagen and/or gelatin binding (Banyai *et al*. 1994). Protein interaction studies are essential to address whether this represents a novel function for an SCP/TAPS protein.
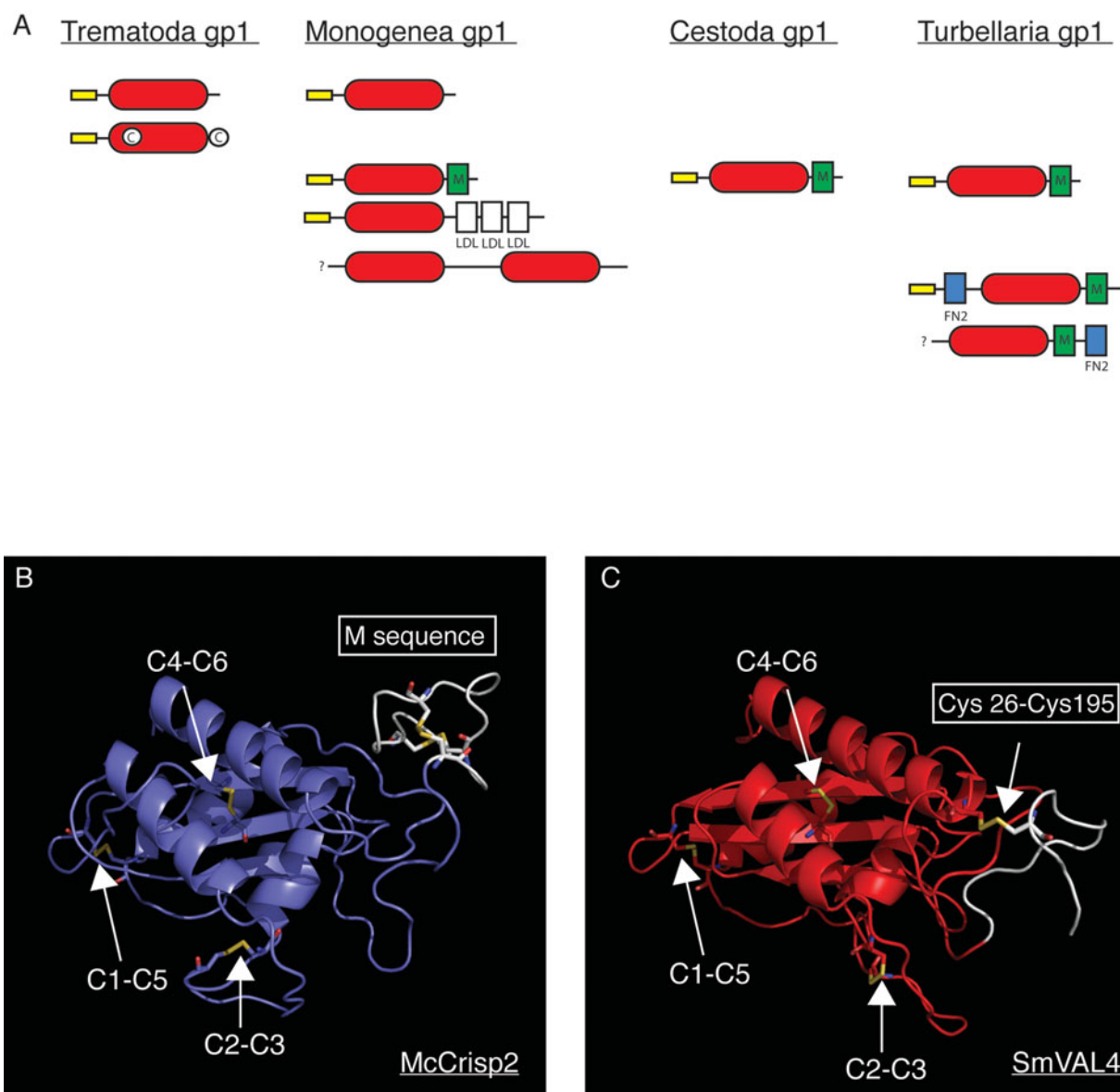
Fig. 3. Diversity of domain architectures across platyhelminth VALs. (A) Cartoon representation of different domain architectures within platyhelminth group 1 VALs across different taxonomic classes. Signal peptides (represented by a yellow box) were identified by SignalP searches. A question mark indicates when the incomplete nature of the sequences did not allow for presence/absence of a signal peptide to be determined. Protein domains were identified by Pfam searches (red boxes represent SCP/TAPS domains (PF00188), white boxes represent low-density lipoprotein receptor domains (PF00057) and blue boxes represent fibronectin 2 domains (PF00040)). The M sequence subdomain (represented by green boxes) was identified by manual inspection of the alignment using the following amino acid convention derived from Gibbs *et al*. (2008) – C-X(2)-C-X(5-10)-C-X(5-15)-C (where C indicates a cysteine residue and X indicates any amino acid). SCP/TAPS domains containing the additional disulphide bond are represented by 2 circled 'C' letters. (B) Homology model of *M. corti* Crisp2 protein. The McCrisp2 M sequence subdomain is coloured white. Potential disulphide bonds are coloured yellow, with the cysteines involved in the formation of each disulphide bond labelled C1-C6. (C) Homology model of *S. mansoni* VAL4 (SmVAL4) protein. The SmVAL4 C-terminal region is coloured white, potential disulfide bonds are coloured yellow and the additional disulphide bond between Cysteine 26 and Cysteine 195 (where the starting Methionine is the first amino acid) indicated by an arrow. Homology models were produced, optimized and verified as described by Chalmers *et al*. (2008) using MODELLER version 9.1 (Eswar, 2006). Specific constraints employed to model the SmVAL4 Cys26-Cys195 disulphide bond did not adversely affect model quality by PROSA-web analysis (Wiederstein and Sippl, 2007). Models were visualized using MacPyMOL (DeLano Scientific LLC).

One subdomain not included in the Pfam database is the M (metazoan) sequence (also known as the Hinge region; (Gibbs *et al.* 2008)). First identified in the snake venom SteCRISP crystal structure

C-terminal to the SCP/TAPS domain (Guo *et al*. 2005), the M sequence is a small (∼ 25AA) subdomain present in multiple group 1 metazoan VAL structures such as Na-ASP-2 and mCRISP2

(Asojo *et al.* 2005; Gibbs *et al.* 2006). The M sequence comprises 2 anti-parallel beta-strands containing 4 disulphide bond-forming cysteines (Fig. 3B) with the following pattern: C-X(2)-C-X(5-10)-C-X(5-15)-C (where C indicates a cysteine residue and X indicates any amino acid). Crucially, the M sequence is known to be essential in mCRISP2 binding to MAP3KII and gametogenetin 1 (Gibbs *et al.* 2007; Jamsai *et al.* 2008), suggesting that this is a critical region for certain protein-protein interactions. In mammalian and reptile CRISP proteins, the M sequence is paired with the vertebrate-specific ion channel regulator subdomain (ICR). However, in other SCP/TAPS proteins, such as those found in *Drosophila* and the Nematoda, it is the only identifiable C-terminal subdomain. Importantly, the presence/absence of the M sequence appears to be a major area of divergence between the trematode VALs and other platyhelminth VALs (Fig. 3). Visual inspection of alignments finds that all trematode group 1 proteins have lost the M sequence, whereas at least one group 1 VAL from the turbellarians, cestodes and monogeneans contains it. For example, greater than 90% of turbellarian group 1 proteins (57/63) contain the M sequence (summarized in Supplementary File 1, online version only). This number is likely to be 100% as the 6 turbellarian VALs not possessing an M sequence are *S. mediterranea* gene predictions without any EST support. Thus, these sequences may represent incorrect gene models. Support for this assertion is found in the phylogenetic analysis where these 6 SmdVALs cluster with M sequence-containing VALs (Fig. 2). In cestodes, 63% (5/8) group 1 VALs contain the M sequence (including the published McCrisp2 and 3; McCrisp2 homology model in Fig. 3B). The 3 cestode VALs missing the M sequence originate from EST sequences encoding no 3′ stop codon, thus likely only missing the M sequence due to incomplete sequence. Finally, approximately 50% (20/38) of the monogenean group 1 VALs encode a C-terminal M sequence. The lack of M sequences in some monogenean VALs does not appear to be due to incomplete EST coverage as the majority of these sequences (15/18) encode a 3′ stop codon. Overall, our sequence analyses suggest that this subdomain is differentially found amongst the Platyhelminthes.

Given the near ubiquity of the M sequence in metazoan group 1 VALs (Chordata; (Gibbs *et al.* 2008), Arthropoda; (Kovalick and Griffin, 2005), Nematoda; (Asojo *et al.* 2005), Gastropoda; (Milne *et al.* 2003)), the complete loss of this subdomain in the trematode VAL family is highly unusual but not unique. For example, the Ag5 wasp venoms do not possess an M sequence (Henriksen *et al.* 2001). However, these SCP/TAPS domain containing proteins differ from the trematode VALs in that they possess an insect-specific N-terminal subdomain named the I (insect) domain. Oddly, due to the lack of an M sequence, it could be argued that the trematode group 1 VALs most closely resemble the plant PR-1 proteins (Fernandez *et al.* 1997). However, a subset of the trematode group 1 VALs (e.g. SmVAL4) appear to contain a trematode-specific structural feature (Fig. 3). Identified by multiple sequence alignment, 2 cysteine residues are co-conserved in 36 trematode VALs originating from all 7 species used in this analysis (Fig. 3). With 1 cysteine present after the first helix of the SCP/TAPS domain and the other C-terminal to the SCP/TAPS domain, this conserved pair of cysteines is unique to these trematode VALs (Fig. 3C). Crucially, homology modelling of SmVAL4 confirms that these 2 cysteines (Cys26-Cys195) could create a disulphide bond within a monomer (Fig. 3C), forming a distinct C-terminal region (Fig. 3C; coloured white). Phylogenetic analysis shows that this fourth disulphide bond is not always maintained, as SmVAL7, SjVAL7 and SmVAL10 do not contain either of the cysteines despite being located in clades containing VALs with the additional disulphide bond (Fig. 2). Further research must be performed to address whether this trematode-specific disulphide bond leads to immunological and/or functional differences in these proteins.

CONCLUSIONS

This review has shown that VAL proteins are present in numerous platyhelminth species in all 4 traditional taxonomic classes. There is strong proteomic evidence that group 1 VALs are secreted by several trematode species during parasite infections, specifically the invasive stages, suggesting that these proteins could perform immunomodulatory functions similar to parasitic nematode homologues such as Na-ASP-2 (Bower *et al.* 2008). Studies into the mammalian CRISP proteins, however, have highlighted the importance of the related subdomains (such as the M sequence) in mediating different protein functions (Gibbs *et al.* 2007). Therefore, close examination of the Platyhelminthes VAL repertoire at the genomic, phylogenetic and structural levels are essential for helping to elucidate functional, immunological and evolutionary roles across the phylum.

The study included in this review has begun this process, finding evidence that phylogenetic and structural differences are more likely to occur between the extracellular group 1 VALs compared to the intracellular group 2 proteins within the phylum. These findings (in combination with studies from across the SCP/TAPS superfamily field) lead to the conclusions that platyhelminth VALs are highly unlikely to all possess the same biological function, although they may all broadly perform the same role (i.e. protein-protein interactions). Even within the group 1 proteins, the class-specific clustering and

clear structural differences observed between VALs suggest that a number of distinct functions have evolved. In parasitic species, this divergence may be driven by parasite/host interactions either directly (VAL proteins interacting with host proteins which differ between hosts) or indirectly (interactions with other parasite proteins involved in parasitism). For the intracellular group 2 proteins, our findings suggest that functions will be largely conserved across platyhelminth species, particularly in the case of the double domain SmVAL11 orthologues present in trematodes, cestodes and turbellarians. Evidence from Hs-GAPR-1, a human group 2 protein, suggests that these group 2 functions will be related to the Golgi complex, specifically at lipid rafts (Eberle, 2002). The wide array of different protein complexes that form at lipid rafts (Lingwood and Simons, 2010), may hint at a role for group 2 proteins in coordinating protein-protein interactions at this site.

Undoubtedly, elucidation of new platyhelminth genomes (Holroyd and Sanchez-Flores, 2011) as well as implementation of multi-species comparative genomic analyses (Swain *et al.* 2011) will provide greater scope for understanding the evolution of VAL families across the phylum. The most urgent studies required, however, are investigations that attempt to ascribe functions or identify interacting partners for the different platyhelminth VAL types (such as group 1 trematode-specific VALs, group 1 with/without M domain VALs, Group 2a VALs and Group 2b VALs). Understanding the particular role of each VAL family member during platyhelminth developmental biology would likely lead to cross-phyla insight important for the full appreciation of this enigmatic, but widely distributed, protein superfamily.

REFERENCES

Adamidi, C., Wang, Y., Gruen, D., Mastrobuoni, G., You, X., Tolle, D., Dodt, M., Mackowiak, S. D., Gogol-Doering, A., Oenal, P., Rybak, A., Ross, E., Sanchez Alvarado, A., Kempa, S., Dieterich, C., Rajewsky, N. and Chen, W. (2011). De novo assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics. *Genome Research* **21**, 1193–1200.

Andersen, J. F., Hinnebusch, B. J., Lucas, D. A., Conrads, T. P., Veenstra, T. D., Pham, V. M. and Ribeiro, J. M. (2007). An insight into the sialome of the oriental rat flea, *Xenopsylla cheopis* (Rots). *BMC Genomics* **8**, 102.

Asojo, O. A., Goud, G., Dhar, K., Loukas, A., Zhan, B., Deumic, V., Liu, S., Borgstahl, G. E. and Hotez, P. J. (2005). 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. *Journal of Molecular Biology* **346**, 801–814.

Banyai, L., Tordai, H. and Patthy, L. (1994). The gelatin-binding site of human 72 kDa type IV collagenase (gelatinase A). *The Biochemical Journal* **298**, 403–407.

Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., Cerqueira, G. C., Mashiyama, S. T., Al-Lazikani, B., Andrade, L. F., Ashton, P. D., Aslett, M. A., Bartholomeu, D. C., Blandin, G., Caffrey, C. R., Coghlan, A., Coulson, R., Day, T. A., Delcher, A., DeMarco, R., Djikeng, A., Eyre, T., Gamble, J. A., Ghedin, E., Gu, Y., Hertz-Fowler, C., Hirai, H., Hirai, Y., Houston, R., Ivens, A., Johnston, D. A., Lacerda, D., Macedo, C. D., McVeigh, P., Ning, Z., Oliveira, G., Overington, J. P., Parkhill, J., Pertea, M., Pierce, R. J., Protasio, A. V., Quail, M. A., Rajandream, M. A., Rogers, J., Sajid, M., Salzberg, S. L., Stanke, M., Tivey, A. R., White, O., Williams, D. L., Wortman, J., Wu, W., Zamanian, M., Zerlotini, A., Fraser-Liggett, C. M., Barrell, B. G. and El-Sayed, N. M. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature, London* **460**, 352–358.

Bethony, J., Loukas, A., Smout, M., Brooker, S., Mendez, S., Plieskatt, J., Goud, G., Bottazzi, M. E., Zhan, B., Wang, Y., Williamson, A., Lustigman, S., Correa-Oliveira, R., Xiao, S. and Hotez, P. J. (2005). Antibodies against a secreted protein from hookworm larvae reduce the intensity of hookworm infection in humans and vaccinated laboratory animals. *The FASEB Journal* **19**, 1743–1745.

Bower, M. A., Constant, S. L. and Mendez, S. (2008). *Necator americanus*: the Na-ASP-2 protein secreted by the infective larvae induces neutrophil recruitment in vivo and in vitro. *Experimental Parasitology* **118**, 569–575.

Braschi, S. and Wilson, R. A. (2006). Proteins exposed at the adult schistosome surface revealed by biotinylation. *Molecular and Cellular Proteomics* **5**, 347–356.

Britos, L., Lalanne, A. I., Castillo, E., Cota, G., Senorale, M. and Marin, M. (2007). *Mesocestoides corti* (syn. vogae, cestoda): characterization of genes encoding cysteine-rich secreted proteins (CRISP). *Experimental Parasitology* **116**, 95–102.

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. and Gasser, R. B. (2009). A portrait of the "SCP/TAPS" proteins of eukaryotes–developing a framework for fundamental research and biotechnological outcomes. *Biotechnology Advances* **27**, 376–388.

Cass, C. L., Johnson, J. R., Califf, L. L., Xu, T., Hernandez, H. J., Stadecker, M. J., Yates, J. R., 3rd and Williams, D. L. (2007). Proteomic analysis of *Schistosoma mansoni* egg secretions. *Molecular and Biochemical Parasitology* **155**, 84–93.

Castro-Borges, W., Dowle, A., Curwen, R. S., Thomas-Oates, J. and Wilson, R. A. (2011). Enzymatic shaving of the tegument surface of live schistosomes for proteomic analysis: a rational approach to select vaccine candidates. *PLoS Neglected Tropical Diseases* **5**, e993.

Chalmers, I. W. (2009). Characterisation of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. PhD thesis, University of Cambridge, Cambridge, UK.

Chalmers, I. W., McArdle, A. J., Coulson, R. M., Wagner, M. A., Schmid, R., Hirai, H. and Hoffmann, K. F. (2008). Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. *BMC Genomics* **9**, 89.

Chen, J., Hu, X., He, S., Wang, L., Hu, D., Wang, X., Zheng, M., Yang, Y., Liang, C., Xu, J. and Yu, X. (2010). Expression and immune response analysis of *Schistosoma japonicum* VAL-1, a homologue of vespid venom allergens. *Parasitology Research* **106**, 1413–1418.

Curwen, R. S., Ashton, P. D., Sundaralingam, S. and Wilson, R. A. (2006). Identification of novel proteases and immunomodulators in the secretions of schistosome cercariae that facilitate host entry. *Molecular and Cellular Proteomics* **5**, 835–844.

Del Valle, A., Jones, B. F., Harrison, L. M., Chadderdon, R. C. and Cappello, M. (2003). Isolation and molecular cloning of a secreted hookworm platelet inhibitor from adult *Ancylostoma caninum*. *Molecular and Biochemical Parasitology* **129**, 167–177.

DeMarco, R., Mathieson, W., Manuel, S. J., Dillon, G. P., Curwen, R. S., Ashton, P. D., Ivens, A. C., Berriman, M., Verjovski-Almeida, S. and Wilson, R. A. (2010). Protein variation in blood-dwelling schistosome worms generated by differential splicing of micro-exon gene transcripts. *Genome Research* **20**, 1112–1121.

Diemert, D. J., Bethony, J. M., Pinto, A. G., Freire, J., Santiago, H., Correa-Oliveira, R. and Hotez, P. J. (2008). Clinical development of the Na-ASP-2 hookworm vaccine in previously-infected Brazilian adults. *The American Journal of Tropical Medicine and Hygiene* **79**. Suppl. 345.

Eberle, H. B., Serrano, R. L., Fullekrug, J., Schlosser, A., Lehmann, W. D., Lottspeich, F., Kaloyanova, D., Wieland, F. T. and Helms, J. B. (2002). Identification and characterization of a novel human plant pathogenesis-related protein that localizes to lipid-enriched microdomains in the Golgi complex. *Journal of Cell Science* **115**, 827–838.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. and Sali, A. (2006). Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics* (ed. Baxevanis, Andreas D., Pearson, William R., Stein, Lincoln D., Stormo, Gary D. and Yates III, John R.), pp. 1–30. Wiley Online Library, doi:10.1002/0471250953.

Fernandez, C., Szyperski, T., Bruyere, T., Ramage, P., Mosinger, E. and Wuthrich, K. (1997). NMR solution structure of the pathogenesis-related protein P14a. *Journal of Molecular Biology* **266**, 576–593.

Francischetti, I. M., Valenzuela, J. G., Pham, V. M., Garfield, M. K. and Ribeiro, J. M. (2002). Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. *The Journal of Experimental Biology* **205**, 2429–2451.

Gibbs, G. M., Bianco, D. M., Jamsai, D., Herlihy, A., Ristevski, S., Aitken, R. J., Kretser, D. M. and O'Bryan, M. K. (2007). Cysteine-rich secretory protein 2 binds to mitogen-activated protein kinase kinase kinase 11 in mouse sperm. *Biology of Reproduction* **77**, 108–114.

Gibbs, G. M., Roelants, K. and O'Bryan, M. K. (2008). The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins–roles in reproduction, cancer, and immune defense. *Endocrine Reviews* **29**, 865–897.

Gibbs, G. M., Scanlon, M. J., Swarbrick, J., Curtis, S., Gallant, E., Dulhunty, A. F. and O'Bryan, M. K. (2006). The cysteine-rich secretory protein domain of Tpx-1 is related to ion channel toxins and regulates ryanodine receptor Ca2 + signaling. *The Journal of Biological Chemistry* **281**, 4156–4163.

Goud, G. N., Zhan, B., Ghosh, K., Loukas, A., Hawdon, J., Dobardzic, A., Deumic, V., Liu, S., Dobardzic, R., Zook, B. C., Jin, Q., Liu, Y., Hoffman, L., Chung-Debose, S., Patel, R., Mendez, S. and Hotez, P. J. (2004). Cloning, yeast expression, isolation, and vaccine testing of recombinant *Ancylostoma*-secreted protein (ASP)-1 and ASP-2 from Ancylostoma ceylanicum. *Journal of Infectious Diseases* **189**, 919–929.

Guo, M., Teng, M., Niu, L., Liu, Q., Huang, Q. and Hao, Q. (2005). Crystal structure of the cysteine-rich secretory protein stecrisp reveals that the cysteine-rich domain has a K + channel inhibitor-like fold. *The Journal of Biological Chemistry* **280**, 12405–12412.

Hansell, E., Braschi, S., Medzihradszky, K. F., Sajid, M., Debnath, M., Ingram, J., Lim, K. C. and McKerrow, J. H. (2008). Proteomic analysis of skin invasion by blood fluke larvae. *PLoS Neglected Tropical Diseases* **2**, e262.

Henriksen, A., King, T. P., Mirza, O., Monsalve, R. I., Meno, K., Ipsen, H., Larsen, J. N., Gajhede, M. and Spangfort, M. D. (2001). Major venom allergen of yellow jackets, Ves v 5: structural characterization of a pathogenesis-related protein superfamily. *Proteins* **45**, 438–448.

Hoffman, D. R. (1993). Allergens in Hymenoptera venom. XXV: The amino acid sequences of antigen 5 molecules and the structural basis of antigenic cross-reactivity. *Journal of Allergy and Clinical Immunology* **92**, 707–716.

Holroyd, N. and Sanchez-Flores, A. (2011). Producing parasitic helminth reference and draft genomes at the Wellcome Trust Sanger Institute. *Parasite Immunology* **34**, 100–107. doi: 10.1111/j.1365-3024.2011.01311.x.

Jamsai, D., Bianco, D. M., Smith, S. J., Merriner, D. J., Ly-Huynh, J. D., Herlihy, A., Niranjan, B., Gibbs, G. M. and O'Bryan, M. K. (2008). Characterization of gametogenetin 1 (GGN1) and its potential role in male fertility through the interaction with the ion channel regulator, cysteine-rich secretory protein 2 (CRISP2) in the sperm tail. *Reproduction* **135**, 751–759.

Jang-Lee, J., Curwen, R. S., Ashton, P. D., Tissot, B., Mathieson, W., Panico, M., Dell, A., Wilson, R. A. and Haslam, S. M. (2007). Glycomics analysis of *Schistosoma mansoni* egg and cercarial secretions. *Molecular and Cellular Proteomics* **6**, 1485–1499.

Kovalick, G. E. and Griffin, D. L. (2005). Characterization of the SCP/TAPS gene family in *Drosophila melanogaster*. *Insect Biochemistry and Molecular Biology* **35**, 825–835.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.

Li, S., Kwon, J. and Aksoy, S. (2001). Characterization of genes expressed in the salivary glands of the tsetse fly, *Glossina morsitans morsitans*. *Insect Molecular Biology* **10**, 69–76.

Lingwood, D. and Simons, K. (2010). Lipid rafts as a membrane-organizing principle. *Science* **327**, 46–50.

Littlewood, D. (2006). The evolution of parasitism in flatworms. In *Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology and Physiology* (ed. Maule, A. G.), pp. 1–36. CABI, Wallingford, UK.

Liu, F., Lu, J., Hu, W., Wang, S. Y., Cui, S. J., Chi, M., Yan, Q., Wang, X. R., Song, H. D., Xu, X. N., Wang, J. J., Zhang, X. L., Zhang, X., Wang, Z. Q., Xue, C. L., Brindley, P. J., McManus, D. P., Yang, P. Y., Feng, Z., Chen, Z. and Han, Z. G. (2006). New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathogens* **2**, e29.

Loon, L. C., Gerritsen, Y. A. M. and Ritter, C. E. (1987). Identification, purification, and characterization of pathogenesis-related proteins from virus-infected Samsun NN tobacco leaves. *Plant Molecular Biology* **9**, 593–609.

Lu, G., Villalba, M., Coscia, M. R., Hoffman, D. R. and King, T. P. (1993). Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps, and yellow jackets. *The Journal of Immunology* **150**, 2823–2830.

Mathieson, W. and Wilson, R. A. (2010). A comparative proteomic study of the undeveloped and developed *Schistosoma mansoni* egg and its contents: the miracidium, hatch fluid and secretions. *International Journal for Parasitology* **40**, 617–628.

Milne, T. J., Abbenante, G., Tyndall, J. D., Halliday, J. and Lewis, R. J. (2003). Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *The Journal of Biological Chemistry* **278**, 31105–31110.

Morris, J., Ladurner, P., Rieger, R., Pfister, D., Del Mar De Miguel-Bonet, M., Jacobs, D. and Hartenstein, V. (2006). The *Macrostomum lignano* EST database as a molecular resource for studying platyhelminth development and phylogeny. *Development Genes and Evolution* **216**, 695–707.

Moyle, M., Foster, D. L., McGrath, D. E., Brown, S. M., Laroche, Y., De Meutter, J., Stanssens, P., Bogowitz, C. A., Fried, V. A., Ely, J. A. and et al. (1994). A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. *The Journal of Biological Chemistry* **269**, 10008–10015.

Mulvenna, J., Sripa, B., Brindley, P. J., Gorman, J., Jones, M. K., Colgrave, M. L., Jones, A., Nawaratna, S., Laha, T., Suttiprapa, S., Smout, M. J. and Loukas, A. (2010). The secreted and surface proteomes of the adult stage of the carcinogenic human liver fluke *Opisthorchis viverrini*. *Proteomics* **10**, 1063–1078.

Nawaratna, S. S., McManus, D. P., Moertel, L., Gobert, G. N. and Jones, M. K. (2011). Gene Atlasing of digestive and reproductive tissues in *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases* **5**, e1043.

Olson, P. D., Zarowiecki, M., Kiss, F. and Brehm, K. (2011). Cestode genomics – progress and prospects for advancing basic and applied aspects of flatworm biology. *Parasite Immunology* **34**, 130–150.

Oyewumi, L., Kaplan, F. and Sweezey, N. B. (2003). Lgl1, a mesenchymal modulator of early lung branching morphogenesis, is a secreted glycoprotein imported by late gestation lung epithelial cells. *The Biochemical Journal* **376**, 61–69.

Ozhogina, O. A. and Bominaar, E. L. (2009). Characterization of the kringle fold and identification of a ubiquitous new class of disulfide rotamers. *Journal of Structural Biology* **168**, 223–233.

Ozhogina, O. A., Trexler, M., Banyai, L., Llinas, M. and Patthy, L. (2001). Origin of fibronectin type II (FN2) modules: structural analyses of distantly-related members of the kringle family idey the kringle domain of neurotrypsin as a potential link between FN2 domains and kringles. *Protein Science* **10**, 2114–2122.

Ravi, V., Ramachandran, S., Thompson, R. W., Andersen, J. F. and Neva, F. A. (2002). Characterization of a recombinant immunodiagnostic antigen (NIE) from *Strongyloides stercoralis* L3-stage larvae. *Molecular and Biochemical Parasitology* **125**, 73–81.

Ribeiro, J. M., Alarcon-Chaidez, F., Francischetti, I. M., Mans, B. J., Mather, T. N., Valenzuela, J. G. and Wikel, S. K. (2006). An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochemistry and Molecular Biology* **36**, 111–129.

Ribeiro, J. M., Charlab, R., Pham, V. M., Garfield, M. and Valenzuela, J. G. (2004). An insight into the salivary transcriptome and proteome of the adult female mosquito *Culex pipiens quinquefasciatus*. *Insect Biochemistry and Molecular Biology* **34**, 543–563.

Rieu, P., Sugimori, T., Griffith, D. L. and Arnaout, M. A. (1996). Solvent-accessible residues on the metal ion-dependent adhesion site face of integrin CR3 mediate its binding to the neutrophil inhibitory factor. *The Journal of Biological Chemistry* **271**, 15858–15861.

Robb, S. M., Ross, E. and Sanchez Alvarado, A. (2008). SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Research* **36** (Database issue): D599–606.

Santos, A., Ribeiro, J. M., Lehane, M. J., Gontijo, N. F., Veloso, A. B., Sant'Anna, M. R., Nascimento Araujo, R., Grisard, E. C. and Pereira, M. H. (2007). The sialotranscriptome of the blood-sucking bug *Triatoma brasiliensis* (Hemiptera, Triatominae). *Insect Biochemistry and Molecular Biology* **37**, 702–712.

Sen, L., Ghosh, K., Bin, Z., Qiang, S., Thompson, M. G., Hawdon, J. M., Koski, R. A., Shuhua, X. and Hotez, P. J. (2000). Hookworm burden reductions in BALB/c mice vaccinated with recombinant *Ancylostoma* secreted proteins (ASPs) from *Ancylostoma duodenale*, *Ancylostoma caninum* and *Necator americanus*. *Vaccine* **18**, 1096–1102.

Swain, M. T., Larkin, D. M., Caffrey, C. R., Davies, S. J., Loukas, A., Skelly, P. J. and Hoffmann, K. F. (2011). *Schistosoma* comparative genomics: integrating genome structure, parasite biology and anthelmintic discovery. *Trends in Parasitology* **27**, 555–564.

Tawe, W., Pearlman, E., Unnasch, T. R. and Lustigman, S. (2000). Angiogenic activity of *Onchocerca volvulus* recombinant proteins similar to vespid venom antigen 5. *Molecular and Biochemical Parasitology* **109**, 91–99.

The *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium. (2009). The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature, London* **460**, 345–351.

Tripathi, L. P. and Sowdhamini, R. (2006). Cross genome comparisons of serine proteases in Arabidopsis and rice. *BMC Genomics* **7**, 200.

Udby, L., Calafat, J., Sorensen, O. E., Borregaard, N. and Kjeldsen, L. (2002). Identification of human cysteine-rich secretory protein 3 (CRISP-3) as a matrix protein in a subset of peroxidase-negative granules of neutrophils and in the granules of eosinophils. *Journal of Leukocyte Biology* **72**, 462–469.

Udby, L., Lundwall, A., Johnsen, A. H., Fernlund, P., Valtonen-Andre, C., Blom, A. M., Lilja, H., Borregaard, N., Kjeldsen, L. and Bjartell, A. (2005). Beta-Microseminoprotein binds CRISP-3 in human seminal plasma. *Biochemical and Biophysical Research Communications* **333**, 555–561.

Udby, L., Sorensen, O. E., Pass, J., Johnsen, A. H., Behrendt, N., Borregaard, N. and Kjeldsen, L. (2004). Cysteine-rich secretory protein 3 is a ligand of alpha1B-glycoprotein in human plasma. *Biochemistry* **43**, 12877–12886.

Valenzuela, J. G., Pham, V. M., Garfield, M. K., Francischetti, I. M. and Ribeiro, J. M. (2002). Toward a description of the sialome of the adult female mosquito *Aedes aegypti*. *Insect Biochemistry and Molecular Biology* **32**, 1101–1122.

van Balkom, B. W., van Gestel, R. A., Brouwers, J. F., Krijgsveld, J., Tielens, A. G., Heck, A. J. and van Hellemond, J. J. (2005). Mass spectrometric analysis of the *Schistosoma mansoni* tegumental sub-proteome. *Journal of Proteome Research* **4**, 958–966.

van Loon, L. C., Rep, M. and Pieterse, C. M. (2006). Significance of inducible defense-related proteins in infected plants. *Annual Review of Phytopathology* **44**, 135–162.

Verjovski-Almeida, S. and DeMarco, R. (2011). Gene structure and splicing in schistosomes. *Journal of Proteomics* **74**, 1515–1518.

Verjovski-Almeida, S., DeMarco, R., Martins, E. A., Guimaraes, P. E., Ojopi, E. P., Paquola, A. C., Piazza, J. P., Nishiyama, M. Y., Jr., Kitajima, J. P., Adamson, R. E., Ashton, P. D., Bonaldo, M. F., Coulson, P. S., Dillon, G. P., Farias, L. P., Gregorio, S. P., Ho, P. L., Leite, R. A., Malaquias, L. C., Marques, R. C., Miyasato, P. A., Nascimento, A. L., Ohlweiler, F. P., Reis, E. M., Ribeiro, M. A., Sa, R. G., Stukart, G. C., Soares, M. B., Gargioni, C., Kawano, T., Rodrigues, V., Madeira, A. M., Wilson, R. A., Menck, C. F., Setubal, J. C., Leite, L. C. and Dias-Neto, E. (2003). Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics* **35**, 148–157.

Wiederstein, M. and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* **35**(Web Server issue): W407–410.

Wu, X. J., Sabat, G., Brown, J. F., Zhang, M., Taft, A., Peterson, N., Harms, A. and Yoshino, T. P. (2009). Proteomic analysis of *Schistosoma mansoni* proteins released during in vitro miracidium-to-sporocyst transformation. *Molecular and Biochemical Parasitology* **164**, 32–44.

Young, N. D., Campbell, B. E., Hall, R. S., Jex, A. R., Cantacessi, C., Laha, T., Sohn, W. M., Sripa, B., Loukas, A., Brindley, P. J. and Gasser, R. B. (2010*a*). Unlocking the transcriptomes of two carcinogenic parasites, *Clonorchis sinensis* and *Opisthorchis viverrini*. *PLoS Neglected Tropical Diseases* **4**, e719.

Young, N. D., Hall, R. S., Jex, A. R., Cantacessi, C. and Gasser, R. B. (2010*b*). Elucidating the transcriptome of *Fasciola hepatica* – a key to fundamental and biotechnological discoveries for a neglected parasite. *Biotechnology Advances* **28**, 222–231.

Young, N. D., Jex, A. R., Cantacessi, C., Hall, R. S., Campbell, B. E., Spithill, T. W., Tangkawattana, S., Tangkawattana, P., Laha, T. and Gasser, R. B. (2011). A portrait of the transcriptome of the neglected trematode, *Fasciola gigantica* – biological and biotechnological implications. *PLoS Neglected Tropical Diseases* **5**, e1004.

Young, N. D., Jex, A. R., Li, B., Liu, S., Yang, L., Xiong, Z., Li, Y., Cantacessi, C., Hall, R. S., Xu, X., Chen, F., Wu, X., Zerlotini, A., Oliveira, G., Hofmann, A., Zhang, G., Fang, X., Kang, Y., Campbell, B. E., Loukas, A., Ranganathan, S., Rollinson, D., Rinaldi, G., Brindley, P. J., Yang, H., Wang, J. and Gasser, R. B. (2012). Whole-genome sequence of *Schistosoma haematobium*. *Nature Genetics* **44**, 221–225. doi: 10.1038/ng.1065

Zamanian, M., Kimber, M. J., McVeigh, P., Carlson, S. A., Maule, A. G. and Day, T. A. (2011). The repertoire of G protein-coupled receptors in the human parasite Schistosoma mansoni and the model organism Schmidtea mediterranea. *BMC Genomics* **12**, 596–608. doi: 10.1186/1471-2164-12-596

Zayas, R. M., Hernandez, A., Habermann, B., Wang, Y., Stary, J. M. and Newmark, P. A. (2005). The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: analysis of ESTs from the hermaphroditic strain. *Proceedings of the National Academy of Sciences, USA* **102**, 18491–18496.