Venom allergen-like protein diversification in flatworms

by

Breanna N. Sipley

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 3, 2019

Keywords: CRISP domain, signal peptide, parasitism, evolution

Approved by

Jamie Oaks, Chair, Assistant Professor of Biological Sciences
Tonia Schwartz, Assistant Professor of Biological Sciences
Sarah Zohdy, Assistant Professor of Disease Ecology, Forestry & Wildlife Sciences

Abstract

Flatworm Venom Allergen-like Proteins (VAPs) modulate mammalian and avian host immune responses, but their evolutionary origins and functions within symbioses (especially parasitism) remain poorly studied. Previous studies suggest Group 1 VAPs in particular play important roles in host-parasite interactions. These studies have historically focused on taxa of medical and economic value, which represent a small number of the more than 22,000 nominal parasitic flatworm species. The extent to which this observation holds against increased taxonomic sampling remains unknown. To address this gap in taxonomic sampling and better understand the diversification of VAPs across flatworm diversity, we mined new transcriptome assemblies from 21 previously unsampled blood fluke species infecting 19 non-mammalian vertebrate host species, in addition to 26 flatworm transcriptomes and genomes available from public databases. We predicted (1) VAPs would be expressed across flatworm diversity; (2) Group 1, unlike Group 2, VAPs would contain signal peptides consistent with roles in host-parasite interactions; and (3) Group 1 VAPs would exhibit faster rates of amino acid substitutions than Group 2 VAPs. Our bioinformatics approach identified 474 novel VAPs expressed in 45 of 47 flatworm species, including 273 Group 1 VAPs (185 with predicted signal peptides) and 201 Group 2 VAPs (only 2 with predicted signal peptides). We also found evidence of accelerated molecular evolution in Group 1 VAPs. This study is of medical relevance as it is a necessary first step to elucidating VAP members and specific VAP residues that mediate host-specificity and pathogenicity in blood flukes. Moreover, understanding the evolution of gene families involved in flatworm-vertebrate host interactions is important for understanding the genomic bases of parasitism.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No.1414475 awarded to BNS; the NSF Division of Environmental Biology under Grant No.1051106 awarded to Stephen A. Bullard and Kenneth M. Halanych; and the Auburn University Cellular and Molecular Biosciences Graduate Research Fellowship under two Peaks of Excellence awards to BNS. The funding bodies played no role in the design, collection, analysis, interpretation of data, or writing. This work was completed in part with resources provided by the Auburn University Hopper Cluster. I'd like to thank Dr. Iain Chalmers for valuable feedback pertinent to this study; Kenneth M. Halanych for sponsoring several educational and professional development opportunities relevant to this work; Rapheal Orélis-Ribero and Kevin M. Kocot for generating transcriptomes; Stephen "Ash" Bullard, Kerry Cobb, Ryan Cook, Kyle David, Brian Folt, Kenneth M. Halanych, Randy Klabacka, Jamie R. Oaks, Tonia Schwartz, Perry L. Wood Jr., and Sarah Zohdy for manuscript suggestions; Jackson R. Roberts and Ash for sharing their knowledge on the organismal biology, ecology, and systematics of several organisms included in this study; and members of the Molette Biological Laboratory for Environmental and Climate Change Studies for their conversations in support of the methods of this study.

To my committee chair Jamie R. Oaks, thank you for adopting me as a Phyletician even though I lacked a soft spot for herps. I know I've benefited immensely from your levelheadedness, patience, kindness, openness, knowledge, guidance, and time. You've introduced me to more things statistical phylogenetics, molecular evolution, and computational biology than I could've possibly imagined, and it's been fun to push myself outside my comfort zones. I owe similar thanks to many other faculty, especially Sarah Zohdy, Tonia Schwartz, Geoff Hill, and Wendy Hood. Thank you, too, to Raymond Henry, Mary Mendonça, and many other advisors, mentors, colleagues, confidantes, and friends. I'd like to give a special shout-out to Kayleigh Chalkowski, James Goodwin, Jesse Grismer, Shaelyn Smith, Samantha "Sammie"

Smoot, Aundrea Westfall, and Bonnie Wilson for your moral support—as well as a handful of healthcare practitioners and psychologists for more than I can put into words. I might've come here to work on marine invertebrates, but I have no regrets about all the other cool organisms I got to work on, people I got to meet, places I got to see, and conversations I got to have instead! Never in my wildest dreams did I think I'd find myself learning Malagasy while collecting blood from wild lemurs or sharing my love of live-long learning with male prisoners in Alabama or growing so confident in my skills as a budding computational biologist.

I dedicate this thesis to my Mema, Patricia "Pat" Joann Greenwood Sipley Kaiser Voss, August 16, 1939 - April 16, 2019, and my best friend Glen Argirion, February 27, 1968 - September 22, 2017. Thank you for supporting my dreams despite many a setback, walking alongside me through my darkest valleys, and loving me when I wasn't strong enough to love myself. To my Papa, G. Hugh Carlton Jr, October 17, 1923 - September 23, 2017, I regret I discovered our shared love of Biology only after your passing but I will always remember you as the wholesome, earnest, and silly yet sensible lover of life who taught horses to sit like dogs and got rattlesnakes to train beagles; I can't help but feel a special kinship with you and I hope I may have even half your knack for "coaching" and inspiring others. To my Grandpa Jack, October 7, 1927 - December 14, 2000, even though you passed when I was just a young buck, I think about you all the time. Thank you for teaching me the importance of knowing there are other perfectly valid ways to live life than just the ones I've seen with my own eyes and to be open to allowing my opinions and understanding of things to evolve when confronted with new information. I doubt you knew you were training me to be a scientist during that game of Tic-Tac-Toe; I wish you could see how far I've come. To Sey Hee Park, September 19, 1989 - January 9, 2018, I'm so sad for the world that we lost you; you inspired so many of us through your determination, sense of humor, and compassion; thank you for being exactly the friend I needed, and Go Gators! Finally, thank you to all those who advocate for greater inclusivity, diversity, and equity; weathering, bearing witness, and courageously standing up against the storms that threaten to tear the fabric of our shared humanity. Your commitment to helping raise the voices of those underprivileged and systematically oppressed gives me hope that, together, we *can* make the world a better place.

Table of Contents

List of Figures

List of Tables

Chapter 1

Venom allergen-like protein diversity in flatworms: implications for understanding the evolution of parasitism

1.1 Introduction

Parasitism is the most common lifestyle among metazoans and has evolved independently more than 200 times (Windsor, 1998; Weinstein & Kuris, 2016). The molecular evolution of genetic pathways involved in evading and coping with host immune responses must underlie transitions to parasitism, but the degree to which independent transitions to parasitism have converged on similar molecular mechanisms remains unknown. One mechanism by which parasites modulate host immune responses is by releasing secretory proteins and extracellular vesicles that interact with and manipulate host immune signaling pathways (Pearce & Sher, 1987; Salzet, Capron, & Stefano, 2000; Hewitson, Grainger, & Maizels, 2009; Harnett, 2014; Coakley, Buck, & Maizels, 2016). Adaptive evolutionary changes such as mutations changing the binding affinity in genes involved in host-parasite interactions allow parasites to evade host immune responses (Buhot et al., 2004). Likewise, such genes are expected to be under reciprocal selective pressures with immune response genes in hosts (Carrillo-Bustamante, Keşmir, & de Boer, 2015; Talbot et al., 2017). In fact, there are many examples of coevolutionary arms races that result in rapid adaptive evolution and genetic divergence in genes encoding host and pathogen binding proteins (Obbard, Jiggins, Halligan, & Little, 2006; Eizaguirre, Lenz, Kalbe, & Milinski, 2012; Paterson et al., 2010).

Among parasite genes families of interest are Venom Allergen-Like Proteins (VAL Proteins; sensu Chalmers et al., 2008 and Chalmers & Hoffmann, 2012; henceforth, VAPs), which are 1) homologous to proteins in plants, yeast, and vertebrates (humans) and 2) associated with defense and immune systems across classes of organisms. VAPs, which have also been referred to as Sperm-Coating Protein (SCP)-like proteins and many other names (see Cantacessi et al., 2009), contain a Cysteine-Rich Secretory Protein (CRISP) domain and belong to the CAP superfamily, so named because of recognized sequence similarity between Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 proteins (Gibbs & O'Bryan, 2007; Gibbs, Roelants, & O'bryan, 2008; Darwiche, Kelleher, Hudspeth, Schneiter, & Asojo, 2016). VAPs share sequence similarity to venom allergen (Antigen) 5 from vespid wasps—which illicit allergic reactions in some humans (Monsalve, Lu, et al., 1999; reviewed in Cantacessi et al., 2009)—and pathogenesis-related 1 proteins—which are ubiquitously expressed across plants and differentially upregulated during pathogen infection (Breen, Williams, Outram, Kobe, & Solomon, 2017). VAPs, venom allergen 5, pathogenesis-related 1 proteins, and other VAP homologues, including GAPR and GLIPR in humans and Pry1 in yeast, have been shown to function in transporting sterols (Cantacessi et al., 2012; Schneiter & Di Pietro, 2013; Kelleher et al., 2014; Darwiche et al., 2016; Breen et al., 2017; Fernandes et al., 2017; Asojo et al., 2018; Darwiche, El Atab, Cottier, & Schneiter, 2018; Darwiche, Lugo, et al., 2018). Interestingly, sterols have been shown to regulate the function of immune cells (reviewed in Spann & Glass, 2013).

In particular, helminth VAPs have been shown to modulate host immune responses. VAPs are strongly immunogenic, inducing vertebrate antibodies (specifically, Immunoglobulin E, IgE) (Farias et al., 2012). Moreover, VAPs are differentially expressed across parasite life history stages and host tissues (Rofatto et al., 2012; Leontovyč et al., 2016; Fernandes et al., 2017). Taken together, these findings suggest VAPs from parasitic flatworms (Asojo et al., 2018; Caraballo, Acevedo, & Zakzuk, 2018) and parasitic nematodes (Lozano-Torres et al., 2014; Cooper & Eleftherianos, 2016) have immunoregulatory functions.

VAPs have been most well-studied in schistosomes due to their medical relevance (Schistosomiasis is the second most devastating parasitic disease in the world according to the CDC). Chalmers et al., 2008 described 28 VAPs from transcripts of the human blood parasite *Schistosoma mansoni* that clustered into two main groups. Chalmers & Hoffmann, 2012 showed that VAPs are expressed across a diversity of free-living and parasitic flatworm lineages. They hypothesized that Group 1 VAPs are fast-evolving secretory proteins, whereas Group 2 VAPs are intracellular proteins that are structurally and functionally conserved. Furthermore, expression of Group 1 VAPs 1, 4, and 10 is upregulated in life stages associated with definitive (i.e., vertebrate) host invasion (Chalmers et al., 2008). Similar results were observed for Group 1 VAP 8 in *Trichobilharzia regenti*, which infects water fowl (Leontovyč et al., 2016). Moreover, Rofatto et al., 2012 found tissue-specific expression of VAPs in *S. mansoni*: VAP 7 was localized to the esophageal gland and VAP 6 was localized to the tegument, suggesting VAP 7, a Group 1 VAP, may play a more active role in host immunoregulation and nutrient acquisition (e.g., blood-feeding) than VAP 6, a Group 2 VAP. If Group 1 VAPs are indeed secretory proteins associated with evading host immune responses, antagonistic coevolution with host-binding proteins may result in faster rates of molecular evolution on average in Group 1 VAPs than Group 2 VAPs, as the latter may function less intimately in host-parasite interactions.

With limited taxonomic sampling, it appears VAP evolution may be driven by host-parasite interactions, but it remains to be seen whether this is true across all flatworms. In particular, although Chalmers & Hoffmann, 2012 included representatives from every major parasitic flatworm clade including monogenoids, tapeworms, and digeneans (including schistosomes), sampling of blood flukes (Schistosomatoidea) has been limited to mammalian and avian schistosomes, even though related schistosomatoideans infect turtles, crocodiles, ray-finned fishes, and sharks and rays (Orélis-Ribeiro, Arias, Halanych, Cribb, & Bullard, 2014). Members of Platyhelminthes provide an interesting system to investigate VAP gene family evolution within a large, ancient radiation of parasites that require a vertebrate to complete their life cycle. The collective ecological diversity of parasitic flatworms is remarkable, comprising ectoparasites (Monongenoidea) that infect the body surfaces, gill, and olfactory lamellae of aquatic vertebrates (chondrichthyans, ray-finned fishes, turtles); flukes (Digenea) that infect the mucosal epithelium of the gut, and rarely gill and skin epithelium, and the blood of vertebrates; tapeworms (Cestoidea) that principally infect the intestinal lumen; and the marine and freshwater free-living flatworms, colloquially referred to by many as “Turbellarians.” Here, we use a bioinformatics approach to investigate the diversity of VAPs expressed across flatworm diversity with an emphasis on VAPs from previously unsampled blood parasitic lineages. Following Chalmers & Hoffmann, 2012, we hypothesize that Group 1 VAPs are secretory proteins that were important in the evolution of parasitism in flatworms and test the following predictions: 1) VAPs are expressed across flatworm diversity, 2) Group 1 VAPs contain signal peptides consistent with signaling roles absent in Group 2 VAPs, and 3) Group 1 VAPs exhibit, on average, faster rates of molecular evolution than Group 2 VAPs.

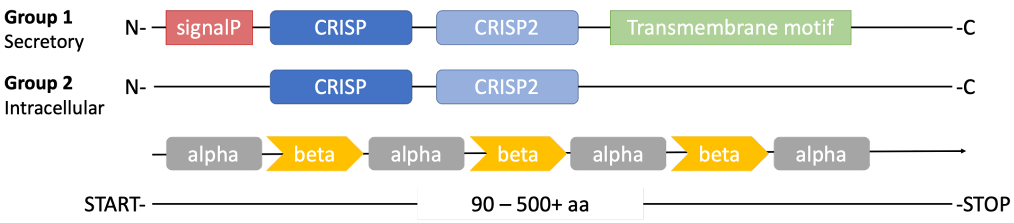
1.2 Materials and methods

1.2.1 Taxonomic sampling, sample collection, and transcriptome data collection

We used 47 flatworm transcriptomes and/or bioinformatically predicted gene models from genomes (Table 1.1) in this study, including 21 *de novo* transcriptomes from non-schistosome blood fluke species (Supplementary Table A.1). We included representatives from all blood fluke families (but see Roberts, Platt, Orélis-Ribeiro, & Bullard, 2016 regarding paraphyly of "Spirorchiidae"). Upon collection, all individuals were stored either at -80 °C or in RNAlater (Life Technologies Inc). RNA extraction, library preparation, sequencing, and transcriptome assembly were performed according to Whelan, Kocot, Moroz, & Halanych, 2015. Briefly, total RNA extracted was purified using TRIzol (Invitrogen) with the RNeasy kit (Qiagen). Single-stranded cDNA libraries were then reverse transcribed with the SMART cDNA Library Construction kit (Clontech) and double-stranded cDNA libraries were synthesized using the Advantage 2 PCR system (Clontech). Library preparation and paired-end sequencing (2 x 100bp) were performed by the Genomic Services Lab at the Hudson Alpha Institute in Huntsville, Alabama on an Illumina HiSeq platform (to be deposited in SRA). Raw reads were digitally normalized by k-mer coverage of 30 using `normalize-by-median.py` (Brown, Howe, Zhang, Pyrkosz, & Brom, 2012) and remaining reads were assembled with `Trinity` r2013-02-25 (Grabherr et al., 2011) using default settings. Additionally, we retrieved 26 publicly available transcriptomes from WormBase ParaSite (`http://parasite.wormbase.org/`) (Howe, Bolt, Shafie, Kersey, & Berriman, 2017), PlanMine (`http://planmine`

`.mpi-cbg.de/`) (Rozanski et al., 2018), trematode.net (`http://trematode.net/`), and the National Center for Biotechnology Information (NCBI). We attempted to represent the greatest diversity in non-blood fluke flatworms currently available and included seven free-living flatworms, two monogenoids, nine cestodes, and eight digeneans (including five non-blood flukes) (Figure 1.1b).

(a) VAP domain architecture



(b) Taxonomic sampling

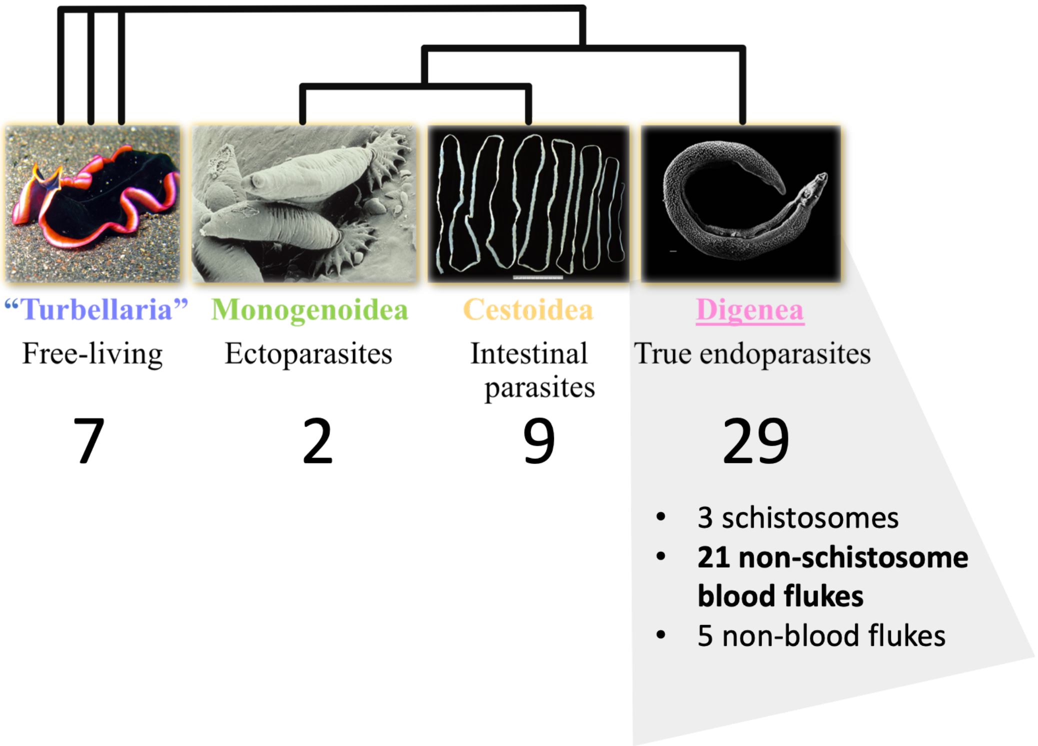


Figure 1.1: VAP inclusion criteria and flatworm diversity represented in this study. (a) A novel VAP must (1) be homologous to known VAPs; (2) contain at least one CRISP domain; (3) contain a start and stop codon. Group 1 VAPs may also contain a secretory +/- transmembrane domain. (b) Number of species sampled mapped onto a simple flatworm phylogeny. Every major parasitic clade as well as the paraphyletic free-living flatworms, colloquially referred to as "Turbellaria" are represented in this study, which includes transcriptomes from 21 previously unsampled non-schistosome blood fluke lineages

Table 1.1: Transcriptomes included in this study

	Taxa	Lifestyle	Genus	species	Host[1]	#contigs	Accession	Reference
1	“Turbellaria”[2]	Free-living	*Dendrocoelum*	*lacteum*	Free-living	82142	N/A	http://planmine.mpi-cbg.de
2	“Turbellaria”	Free-living	*Macrostomum*	*ligN/Ao*	Free-living	60534	PRJNA284736	Wasik et al., 2015
3	“Turbellaria”	Free-living	*Planaria*	*torva*	Free-living	54591	N/A	http://planmine.mpi-cbg.de
4	“Turbellaria”	Free-living	*Polycelis*	*nigra*	Free-living	46985	N/A	http://planmine.mpi-cbg.de
5	“Turbellaria”	Free-living	*Polycelis*	*tenuis*	Free-living	53363	N/A	http://planmine.mpi-cbg.de
6	“Turbellaria”	Free-living	*Schmidtea*	*mediterranea*	Free-living	29850	PRJNA12585	Robb, Ross, & Alvarado, 2007
7	“Turbellaria”	Free-living	*Schmidtea*	*polychroa*	Free-living	54433	N/A	http://planmine.mpi-cbg.de
8	Monogenoidea	Ectoparasite	*Gyrodactylus*	*salaris*	Salmon	15436	PRJNA244375	Hahn, Fromm, & Bachmann, 2014
9	Monogenoidea	Ectoparasite	*Protopolystoma*	*xenopodis*	African clawed frog	37906	PRJEB1201	Unpublished
10	Cestoda	Gut parasite	*Echinococcus*	*multilocularis*	Canid; Hominid	10669	PRJEB122	Tsai et al., 2013
11	Cestoda	Gut parasite	*Hymenolepis*	*diminuta*	Arthropoda; Rodentia	11271	PRJEB507	Unpublished
12	Cestoda	Gut parasite	*Hymenolepis*	*microstoma*	Arthropoda; Rodentia	12373	PRJEB124	Tsai et al., 2013
13	Cestoda	Gut parasite	*Mesocestoides*	*corti*	Arthropoda; Carnivora; Rodentia	10614	PRJEB510	Unpublished
14	Cestoda	Gut parasite	*Schistocephalus*	*solidus*	Aves; Rodentia	20228	PRJEB527	Unpublished
15	Cestoda	Gut parasite	*Spirometra*	*erinaceieuropaei*	Carnivora; Hominid	39557	PRJEB1202	Bennett et al., 2014
16	Cestoda	Gut parasite	*Taenia*	*asiatica*	Suid; Hominid	10331	PRJEB532	Unpublished
17	Cestoda	Gut parasite	*Taenia*	*saginata*	Bovid; Hominid	13161	PRJNA71493	Wang et al., 2016
18	Cestoda	Gut parasite	*Taenia*	*solium*	Bovid; Hominid	12481	PRJNA170813	Tsai et al., 2013
19	Digenea	Endoparasite	*Acipensericola*	*petersoni*	American paddlefish	152140	TBD	This study
20	Digenea	Endoparasite	*Cardicola*	*currani*	Red drum	86962	TBD	This study
21	Digenea	Endoparasite	*Cardicola*	*palmeri*	Black drum	52837	TBD	This study
22	Digenea	Endoparasite	*Clonorchis*	*sinensis*	Hominid	13634	PRJDA72781	Huang et al., 2013
23	Digenea	Endoparasite	*Coeuritrema*	*platti*	Chinese softshell turtle	173090	TBD	This study
24	Digenea	Endoparasite	*Elaphrobates*	*euzeti*	Northern red snapper	118013	TBD	This study
25	Digenea	Endoparasite	*Elopicola*	*bristowi*	Hawaiian ladyfish	58171	TBD	This study
26	Digenea	Endoparasite	*Elopicola*	*franksi*	Atlantic tarpon	64384	TBD	This study
27	Digenea	Endoparasite	*Elopicola*	*nolancribbi*	Northern ladyfish/ tenpounder	222375	TBD	This study
28	Digenea	Endoparasite	*Fasciola*	*hepatica*	Ovis	15739	PRJNA179522	McNulty et al., 2017
29	Digenea	Endoparasite	*Hapalorhynchus*	*foliorchis*	Common snapping turtle	42863	TBD	This study
30	Digenea	Endoparasite	*Littorellicola*	*billhawkinsi*	Florida pompano	55698	TBD	This study
31	Digenea	Endoparasite	*Microphallus*	*livelyi*	Gastropoda; Aves/Rodentia?	15423	PRJNA360970	Bankers & Neiman, 2017
32	Digenea	Endoparasite	*Myliobaticola*	sp. nov.	Caribbean electric ray	73883	TBD	This study
33	Digenea	Endoparasite	*Myliobaticola*	*richardheardi*	Atlantic stingray	15147	TBD	This study
34	Digenea	Endoparasite	gen. nov.	sp. nov.	Thai yellow catfish?	145041	TBD	This study
35	Digenea	Endoparasite	*Nomasanguinicola*	*canthoensis*	Broadhead catfish	30615	TBD	This study
36	Digenea	Endoparasite	*Paragonimus*	*westermani*	Gastropoda; Hominid	27842	PRJNA219632	Unpublished
37	Digenea	Endoparasite	*Phthinomita*	sp.	Orange-spotted spinefoot	17908	TBD	This study
38	Digenea	Endoparasite	*Psettarium*	*anthicum*	Cobia	109559	TBD	This study
39	Digenea	Endoparasite	*Schistosoma*	*haematobium*	Gastropoda; Hominid	13073	PRJNA78265	Young et al., 2012
40	Digenea	Endoparasite	*Schistosoma*	*japonicum*	Gastropoda; Hominid	12738	PRJEA34885	Liu et al., 2009
41	Digenea	Endoparasite	*Schistosoma*	*mansoni*	Gastropoda; Hominid	11774	PRJEA36577	Berriman et al., 2009
42	Digenea	Endoparasite	*Selachohemecus*	*olsoni*	Atlantic sharpnose shark	135169	TBD	This study
43	Digenea	Endoparasite	*Spirorchis*	*haematobius*	Common snapping turtle	132292	TBD	This study
44	Digenea	Endoparasite	*Spirorchis*	*picta*	Pond slider turtle	121315	TBD	This study
45	Digenea	Endoparasite	*Spirorchis*	*scripta*	Pond slider turtle	96187	TBD	This study
46	Digenea	Endoparasite	*Trichobilharzia*	*regenti*	Gastropoda; Aves	22185	PRJEB4662	Unpublished
47	Digenea	Endoparasite	*Vasotrema*	sp.	Spiny softshell turtle	20819	TBD	This study

---

[1] All new parasites from this study were sampled from vertebrate hosts

[2] Colloquial term used to describe free-living flatworms; not monophyletic

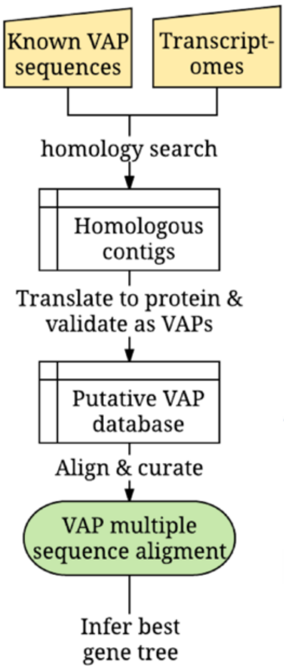


Figure 1.2: Schematic of bioinformatics pipeline. Known VAP sequences from *Schistosoma mansoni* were blasted against a database containing all transcriptomes included in this study. Contigs with significant hits (e-cutoff = 0.001) to known VAP sequences were translated and subjected to a series of filtering and validation steps. This putative VAP database was then aligned with a previously aligned flatworm VAP database containing only the CRISP domain. The CRISP domain region was then manually extracted from the alignment to produce the multiple sequence alignment used to infer the best VAP CRISP domain tree from 474 new and 273 previously identified VAPs across flatworm diversity.

1.2.2 Bioinformatics pipeline

Our bioinformatics pipeline is summarized in Figure 1.2.

We define a venom allergen-like protein as a protein sequence that has (1) sequence similarity to known *Schistosoma mansoni* VAPs (`SmVAL proteins`; Chalmers et al., 2008), (2) more than 75% coverage for at least one cysteine-rich secretory protein (CRISP) domain (Gibbs et al., 2006) `(PMID:16339766)`; and (3) start and stop codons (Figure 1.1a).

To identify novel VAPs, we first queried known `SmVAL proteins` against a database consisting of the flatworm transcriptomes and/or bioinformatically predicted gene models from genomes listed in Table 1.1. For bait sequences, we retrieved the longest protein sequence available for each SmVAL protein member from Uniprot on 2017-08-14 (27 total protein sequences as no representative for SmVAL protein 28 was available) and confirmed VAP identity by the presence of a CRISP domain using `hmmscan` within the HMMER web server (`https://www.ebi.ac.uk/Tools/hmmer/`). A CRISP domain was present in all SmVAL protein sequences but one (SmVAL protein 23), which was removed. Consequently, 26 of these *S. mansoni* venom allergen-like protein sequences were used as bait in this analysis. We built a cDNA BLAST+ database of all using `makeblastdb` as implemented in BLAST+

version 2.6.0 (Altschul, Gish, Miller, Myers, & Lipman, 1990). Next, we queried our SmVAL protein bait across our flatworm transcriptome database using `tblastn` (Altschul et al., 1997) with an e-value cut-off of 0.001 (Pearson, 2013).

We used the perl script `select_contigs.pl` (White, 2009) to select homologous contigs with significant hits based on e-values (cut-off = 0.001) and extracted long open reading frames (ORFs) with `TransDecoder` version 3.0.1 with a minimum ORF length of 90 amino acids, the length of the shortest SmVAL protein included in our bait. To maximize sensitivity in capturing ORFs with functional significance, we performed a `blastp` search against the `Uniprot protein database` (updated April 25, 2018), a `hmmer` search (version 3.1b2; `http://hmmer.org/`) against the `Pfam database` (updated Feb 23, 2017), and included the `SmVAL protein bait` to train a Markov model within `TransDecoder` for VAP coding sequences. Analyses with and without the `"--train"` option produced similar results. The only differences were 13 fewer and one additional translated region when the `"--train"` option was used. To be conservative, we removed the coding region predicted without the `"--train"` option as well as all sequences that lacked a start and/or stop codon.

A final domain search was performed on these complete coding region predictions using `hmmscan` and excluded any protein sequences without a CRISP domain. We further reduced our dataset by removing (1) redundant sequences with `cd-hit` (threshold: % identical = 100) (Li & Godzik, 2006; Fu, Niu, Zhu, Wu, & Li, 2012), (2) all sequences with 100% identity to known VAP sequences from Chalmers & Hoffmann, 2012, and (3) sequences without greater than 75% CRISP domain coverage, which gave us our putative VAP database.

1.2.3 Phylogenetic inference and statistical analyses

Dr. Iain Chalmers kindly provided the VAP multiple sequence alignment used in Chalmers & Hoffmann, 2012, which contains the CRISP domains from 237 predcited VAPs from each parasitic flatworm clade as well as free-living flatworms (henceforth, CH12). We validated CRISP domain presence in each of the sequences in this alignment using Hmmer. We then aligned our new VAPs to CH12 using the `"--add"` option available in the online version of `MAFFT` version 7 (Katoh & Standley, 2013), in combination with the all-pair global alignment iterative refinement method (`G-INS-i`) and default parameters except we increased the offset value to 0.1. We then manually extracted the CRISP domain regions of this alignment, i.e., the region spanning CH12 (sites 1461-2191), in `Geneious` version 11.1.4 to produce a final VAP CRISP domain multiple sequence alignment containing the CRISP domain regions from all VAPs identified in this study aligned with those previously identified in Chalmers & Hoffmann, 2012.

The perl script `ProteinModelSelection.pl` available in RaxML version 8.2.9 (Stamatakis, 2014) determined that a `WAG+GAMMA` model of protein evolution best fit our data. Therefore, we estimated the VAP CRISP domain tree using a `WAG+GAMMA` model and 50 maximum likelihood search replicates, each starting from a random tree. Using the same model, we also inferred trees from 1000 bootstrapped alignments. Nodes with less than 50% bootstrap support were collapsed. VAP group assignment for each novel VAP was determined based on clustering of CRISP domains with known VAP CRISP domains. We annotated this best VAP CRISP domain tree based on signal peptide and transmembrane motif predictions from the default settings of `SignalP` version 4.1 (Nielsen, 2017) using `ggtree` (Yu, Smith, Zhu, Guan, & Lam, 2017) in `R` version 3.4.3 (R Core Team, 2019).

To explore how parameters within SignalP (Nielsen, Engelbrecht, Brunak, & von Heijne, 1997) affect signal peptide predictions, we ran several tests including (1) lowering the sensitivity of `SignalP` version 4.1 to that of version 3.0 (D-cutoff = 3.0) (Bendtsen, Nielsen, von Heijne, & Brunak, 2004); (2) disabling N-terminal truncation of input sequences (default = 70 aa) with default sensitivity in version 4.1; and (3) using default settings in the recently released `SignalP` version 5.0 (Armenteros et al., 2019), which claims to improve signal peptide predictions using deep neural networks.

As a proxy to test the null hypothesis that rates of molecular evolution, on average, between Group 1 VAPs and Group 2 VAPs are not significantly different, we compared branch lengths of Group 1 VAP terminal nodes from the Group 1 VAP basal node to branch lengths of Group 2 VAP terminal nodes from the Group 2 VAP basal node using a Mann-Whitney's U-test.

To improve our estimate of the alignment and VAP relationships within each of the two main VAP groups that were recovered (Figure 1.3), we repeated the alignment and phylogenetic methods described above on each group separately using the complete protein sequences (including signal peptides) of our novel VAP protein sequences. The only additional difference was that `VT` was selected as the best-fit model of amino acid substitution.

1.3 Results

We identified 474 new Venom Allergen-like Proteins (VAPs) from 45 of 47 flatworm transcriptomes (Table 1.2, Supplementary Tables A.2-A.3). These VAPs, which included at least 75% of the canonical CRISP domain, ranged from 113 to 960 amino acids long (mean: 257 aa; standard deviation: 121 aa). On average, free-living flatworms expressed the greatest number of VAPs (21.7 VAPs in 7 species), followed by the monogenoid *Gyrodactylus salaris* (18.0, 1 species), tapeworms (12.6, 9 species), and digeneans (6.8, 28 species) (Table 1.3).

Of the 474 VAPs, we characterized 273 as Group 1 VAPs and 201 as Group 2 VAPs based on phylogenetic clustering of the CRISP domain region with the CRISP domain region of known VAPs (Figure 1.3), summarized in Table 1.2, (Chalmers & Hoffmann, 2012). Based on the full protein sequences for our VAPs, 187/474 VAPs were predicted to have a signal peptide and/or transmembrane protein motif, of which only two fell within Group 2 VAPs (Figure 1.3). A preliminary investigation of the two Group 2 VAP sequences with predicted signal peptides revealed their detection was not an artifact of the signal peptide prediction software we used. Manual inspection confirmed they contain N-terminal leucine rich regions, and `DeepLoc` predicted them to be localized to the cell membrane (Almagro Armenteros, Sønderby, Sønderby, Nielsen, & Winther, 2017), both consistent with signaling export. Both of these Group 2 VAPs are from endoparasitic flatworms.

Although there is considerable overlap between the distribution curves of Group 1 and Group 2 VAP branch lengths from their respective common ancestor, branch lengths are typically longer in Group 1 VAPs than Group 2 VAPs (Figure 1.4). The median branch lengths of Group 1 and Group 2 were 0.551558 and 0.4269806, respectively. Despite the ancestor-to-tip lengths not being independent, we used a Mann-Whitney's U-test to quantify the difference in

Table 1.2: Summary of novel VAPs identified in this study by taxon. Group classification was based on phylogenetic clustering of CRISP domains with CRISP domains from previously identified VAPs; signal peptide predictions for novel VAPs were based on the complete protein sequences using SignalP.

Lifestyle	Genus	species	# VAPs	# Group 1 (# SignalP)	# Group 2 (# SignalP)
Free-living	*Dendrocoelum*	*lacteum*	28	17 (11)	11 (0)
	Macrostomum	*ligo*	16	4 (4)	12 (0)
	Planaria	*torva*	25	11 (10)	14 (0)
	Polycelis	*nigra*	20	11 (11)	9 (0)
	Polycelis	*tenuis*	24	13 (11)	11 (0)
	Schmidtea	*mediterranea*	10	9 (7)	1 (0)
	Schmidtea	*polychroa*	29	15 (10)	14 (0)
Ectoparasite	*Gyrodactylus*	*salaris*	18	12 (6)	6 (0)
Intestinal parasite	*Echinococcus*	*multilocularis*	14	9 (7)	5 (0)
	Hymenolepis	*diminuta*	10	7 (6)	3 (0)
	Hymenolepis	*microstoma*	21	17 (11)	4 (0)
	Mesocestoides	*corti*	21	19 (11)	2 (0)
	Schistocephalus	*solidus*	4	2 (1)	2 (0)
	Spirometra	*erinaceieuropaei*	4	2 (2)	2 (0)
	Taenia	*asiatica*	8	6 (5)	2 (0)
	Taenia	*saginata*	18	13 (8)	5 (0)
	Taenia	*solium*	13	8 (5)	5 (0)
True endoparasite	*Acipensericola*	*petersoni*	8	4 (0)	4 (0)
	Cardicola	*currani*	5	1 (1)	4 (0)
	Cardicola	*palmeri*	3	1 (1)	2 (0)
	Clonorchis	*sinensis*	14	9 (5)	5 (0)
	Coeuritrema	*platti*	12	5 (5)	7 (0)
	Elaphrobates	*euzeti*	4	0 (0)	4 (1)
	Elopicola	*bristowi*	1	1 (0)	0 (0)
	Elopicola	*franksi*	4	0 (0)	4 (0)
	Elopicola	*nolancribbi*	1	0 (0)	1 (0)
	Fasciola	*hepatica*	6	5 (2)	1 (1)
	Hapalorhynchus	*foliorchis*	17	12 (10)	5 (0)
	Littorellicola	*billhawkinsi*	5	2 (1)	3 (0)
	Microphallus	*livelyi*	8	3 (3)	5 (0)
	Myliobaticola	sp. nov.	8	4 (4)	4 (0)
	gen. nov.	sp. nov.	7	1 (1)	6 (0)
	Nomasanguinicola	*canthoensis*	1	1 (0)	0 (0)
	Paragonimus	*westermani*	1	1 (1)	0 (0)
	Phthinomita	sp.	3	3 (0)	0 (0)
	Psettarium	*anthicum*	8	1 (1)	7 (0)
	Schistosoma	*haematobium*	9	6 (2)	3 (0)
	Schistosoma	*japonicum*	6	5 (3)	1 (0)
	Schistosoma	*mansoni*	8	7 (2)	1 (0)
	Selachohemecus	*olsoni*	3	1 (1)	2 (0)
	Spirorchis	*haematobius*	5	0 (0)	5 (0)
	Spirorchis	*picta*	9	4 (1)	5 (0)
	Spirorchis	*scripta*	11	3 (1)	8 (0)
	Trichobilharzia	*regenti*	15	13 (11)	2 (0)
	Vasotrema	sp.	9	5 (3)	4 (0)

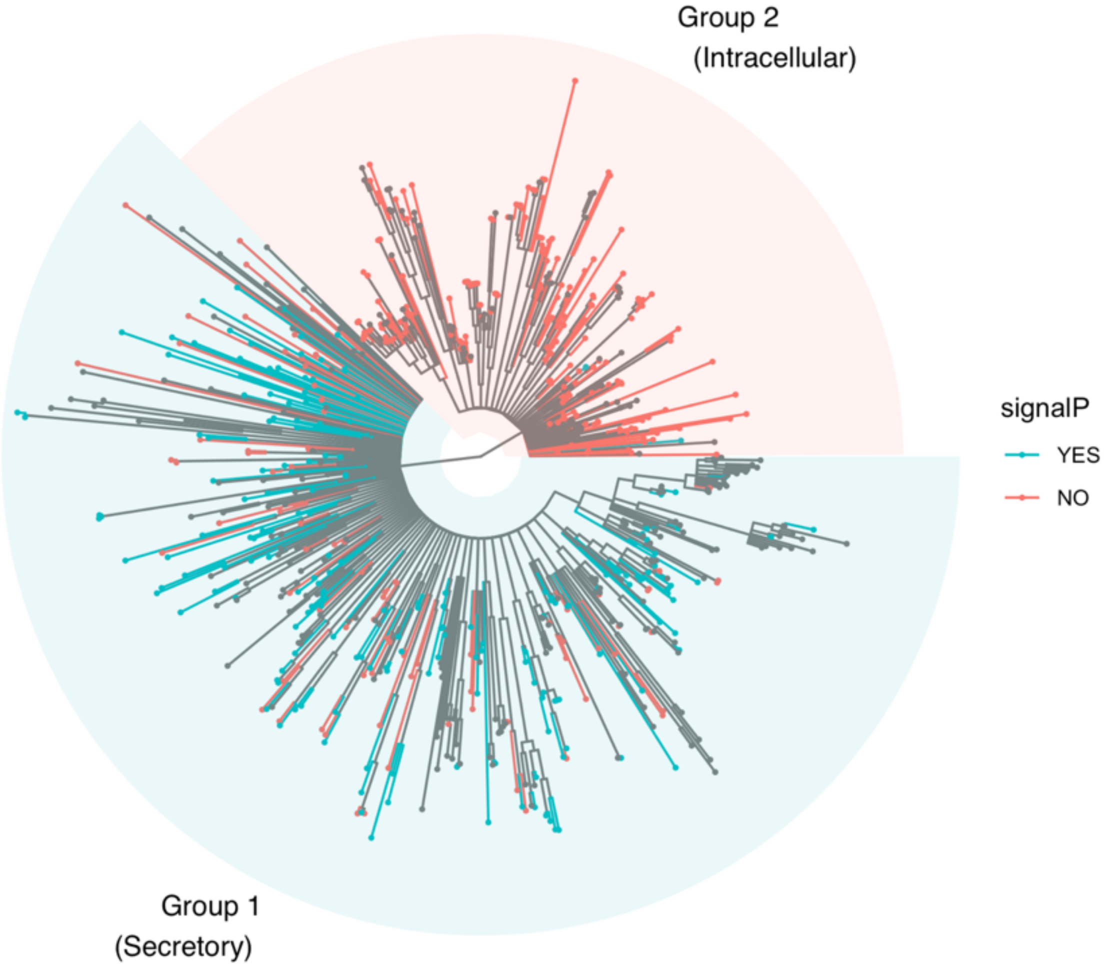


Figure 1.3: Signal peptide +/- transmembrane motif predictions based on the full protein sequences for all novel VAPs identified in this study mapped onto the maximum likelihood VAP CRISP domain tree which also includes the CRISP domains from previously identified VAPs. The best tree was estimated with RAxML using the `WAG+GAMMA` model of protein evolution and 50 maximum likelihood search replicates, each starting from a random tree. Bootstrap support was inferred from trees from 1000 bootstrapped alignments. Nodes with less than 50% bootstrap support are collapsed. VAP group assignment for each novel VAP was determined based on CRISP domain clustering with the CRISP domains of known VAPs. Bootstrap support for the branch separating Group 1 and Group 2 VAPs was 99%. Grey tips represent VAP CRISP domains from Chalmers & Hoffman (2012) for which we do not have signal peptide predictions. Despite the fact that this tree was estimated using only CRISP domains and no signal peptide sequences, there is a striking difference in where the signal peptide predictions fall in the tree. All but two signal peptide predictions are from Group 1 VAPs.

Table 1.3: Summary of the average number of novel VAPs expressed across species within each major taxonomic group. Taxonomic group here also corresponds to shared lifestyle and relative degree of dependency on a vertabrate host to complete its lifecycle. Free-living flatworms appear to express more VAPs on average than parasitic flatworms; true endoparasitic flatworms, many of which live in and feed on vertebrate blood, appear to express the fewest number of VAPs on average; intestinal parasitic and ectoparasitic flatworms appear to express more Group 1 VAPs than Group 2 VAPs on average. Note: parasitic flatworms, especially cestodes and digeneans, tend to have complex lifecyles involving several hosts, and we cannot untangle the effect of developmental stage here

	# VALs	# Group 1	# Group 2	# SignalP
”Turbellaria”/Free-living	21.7	11.4	10.3	9.1
Monogenoidea/Ectoparasite	18.0	12.0	6.0	6.0
Cestoidea/Intestinal parasite	12.6	9.2	3.3	6.2
Digenea/True endoparasite	6.8	3.5	3.3	2.2

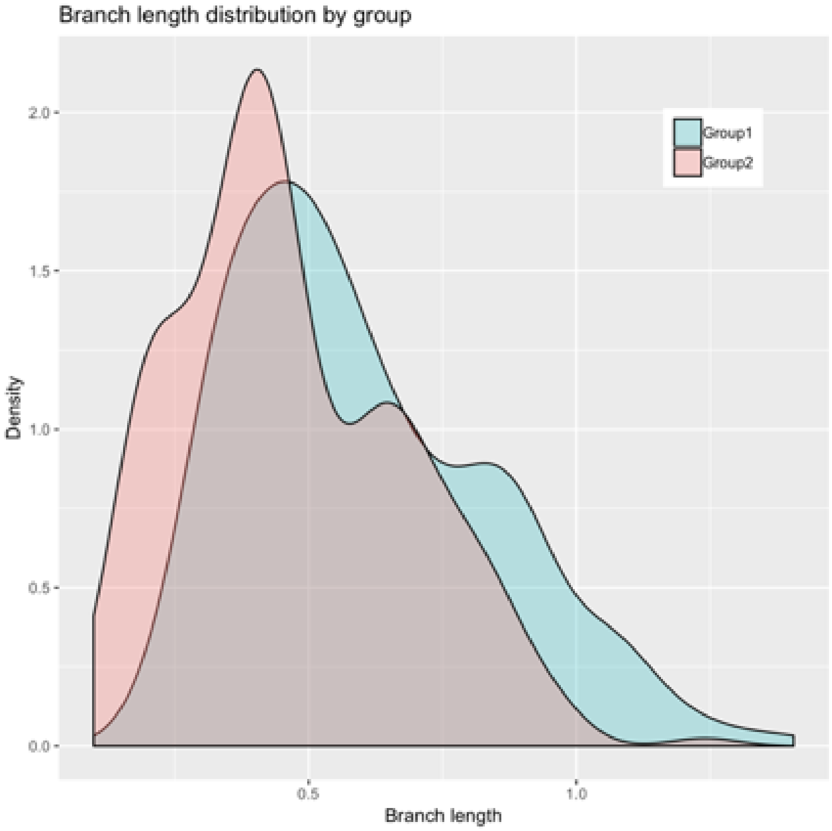


Figure 1.4: Distribution curves of VAP CRISP domain branch lengths by group measured by distance from the corresponding ancestral group node.

the branch lengths between groups. We found a significant albeit small effect of group classification (the mean ranks of Group 1 and Group 2 were 398.85 and 286.40, respectively. W = 77857, Z = 7.006797, $p<1e-12$, r = 0.2627754).

Clades with at least 50% bootstrap support contained VAPs from specific subclades of flatworms that included only free-living flatworms, digeneans, *Gyrodactylus salaris*, or tapeworms, but no combination of these, except for one Group 2 VAP clade for which there was strong support (830/1000 bootstrap replicates), a clade containing VAPs from free-living flatworms, *G. salaris*, digeneans, and tapeworms (Figure 1.5). These patterns were also found in separate trees inferred from full protein sequence alignments of Group 1 and Group 2 VAPs (Figure 1.6).

1.4 Discussion

Our study nearly triples the number of venom allergen-like proteins (VAPs) characterized in flatworms. This dataset is a necessary first step to investigate how the diversification of this gene family, which has ancestral functions unrelated to parasitism, has given rise to proteins that interact with a large diversity of invertebrate and vertebrate immune signaling pathways. Our focus on previously unsampled lineages of non-schistosome blood flukes will be particularly valuable to those who wish to understand the evolution of endoparasitism, blood parasitism, hematophagy, pathogenicity, and host-specificity in Digeneans. In addition to the 474 novel flatworm VAPs we contribute, we also share our bioinformatics pipeline which can be used to explore the potentially similarly rich diversity of VAPs in other major parasitic lineages including nematodes and arthropods (e.g., lice, ticks, mites, and fleas), or easily modified to survey the diversity of other gene families of interest. Moreover, we make available the transcriptome assemblies from these previously unsampled blood fluke taxa found inside a diversity of non-mammalian vertebrate hosts, which will be an incredible resource for using comparative transcriptomics to identify other candidate gene families of interest and understand the genomic bases of endoparasitism more broadly.

Our study strongly supports the prediction that two main groups of VAPs can be distinguished by the presence of signal peptides in Group 1 VAPs (Chalmers et al., 2008; Chalmers

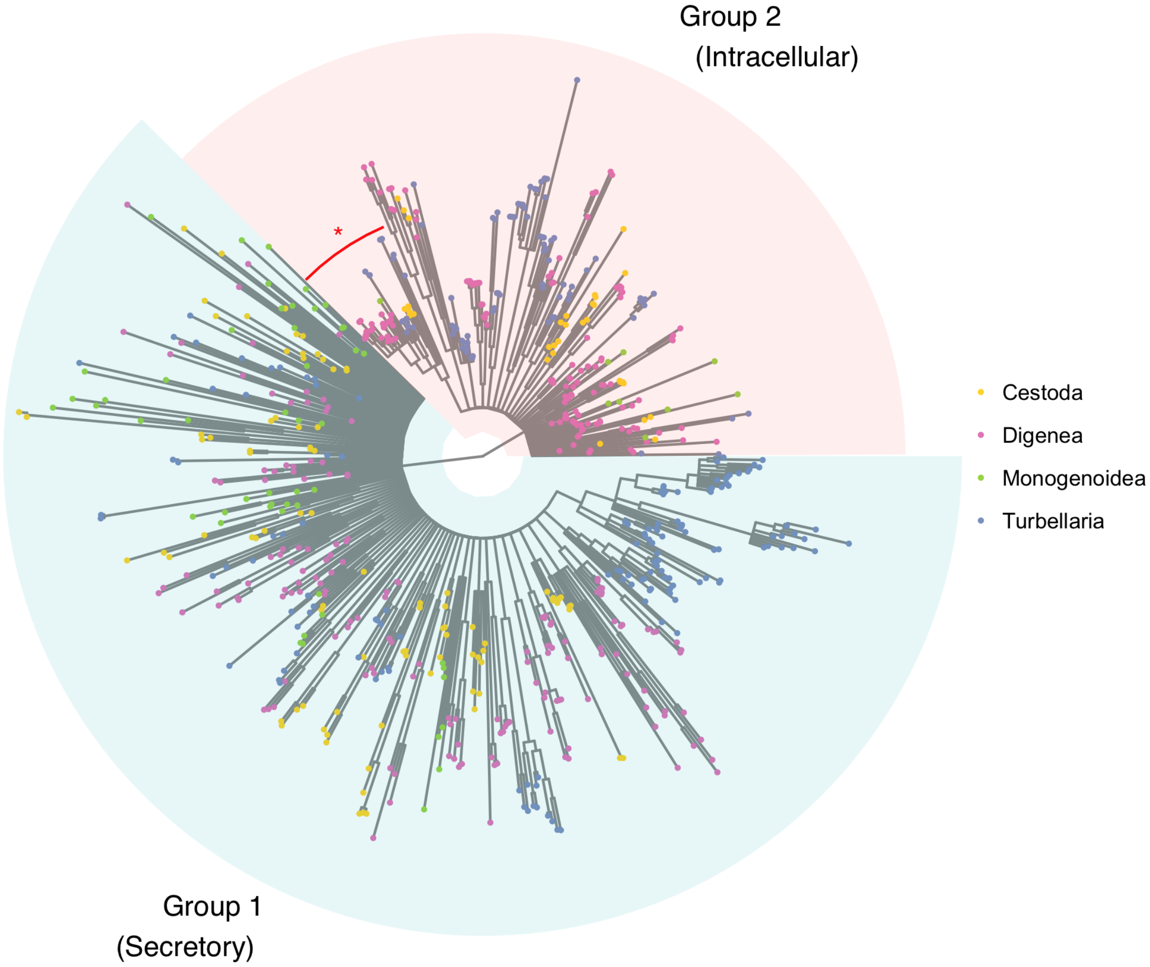


Figure 1.5: Major flatworm taxonomic groups which also correspond to shared lifestyle mapped onto the maximum likelihood VAP CRISP domain tree which also includes the CRISP domains from previously identified VAPs. The best tree was estimated with RAxML using the `WAG+GAMMA` model of protein evolution and 50 maximum likelihood search replicates, each starting from a random tree. Bootstrap support was inferred from trees from 1000 bootstrapped alignments. Nodes with less than 50% bootstrap support are collapsed. VAP group assignment for each novel VAP was determined based on CRISP domain clustering with the CRISP domains of known VAPs. Bootstrap support for the branch separating Group 1 and Group 2 VAPs was 99%. No Group 1 VAP CRISP domain clade contains VAP CRISP domains from more than one major flatworm taxonomic group; one Group 2 clade contains VAP CRISP domains from all four major flatworm groups (highlighted in red and denoted with asterisk; bootstrap support = 83%).

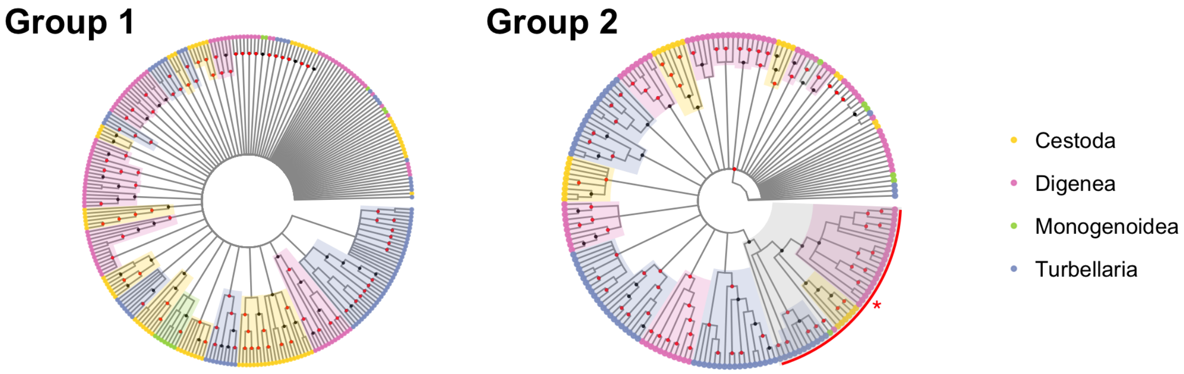


Figure 1.6: Maximum likelihood trees for novel Group 1 and Group 2 VAPs based on complete predicted protein sequences including signal peptides. Clades with less than 50% bootstrap support are collapsed. Major flatworm clades are color-coded as yellow=Cestoda; green=Monogenoidea; pink=Digenea; blue=Turbellaria. No Group 1 VAP clade contains VAPs from more than one major flatworm clade; one Group 2 clade contains VAPs from all four major flatworm clades (highlighted in red). Nodes with 75-94% bootstrap support are gray; clades with at least 95% bootstrap support are red.

& Hoffmann, 2012) in addition to phylogenetic clustering of VAP CRISP domains alone. Of the 201 new Group 2 VAPs we found, only two contained predicted signal peptides; in contrast, 185 of the 273 new Group 1 VAPs we found contained predicted signal peptides. It is important to emphasize that `SignalP` only predicts N-terminal signal peptides even though several C-terminal and internal signal peptides are known (Nielsen, 2012). Moreover, the sensitivity of `SignalP` is reportedly higher in version 4 than in version 3 (Nielsen, 2017). Exploring the effect of varying parameters on the behavior of SignalP indeed validated this claim in Group 1 VAPs, but did not affect the number of signal peptides predicted in Group 2 VAPs. Moreover, N-terminal signal peptides and other sorting signals are known to be highly divergent: evolutionary sequence divergence of sorting signals, instead of sequence conservation, has even been proposed as a more effective approach to identify sorting signals (Fukasawa, Leung, Tsui, & Horton, 2011). Altogether, this suggests that the number of Group 1 VAPs predicted to have signal peptides in this study is a conservative estimate. Nevertheless, as signal peptides are known to play roles in infectivity, functioning as trans-membrane domains and as antigens (see Owji, Nezafat, Negahdaripour, Hajiebrahimi, & Ghasemi, 2018 for a review), our study lends support to the claim that Group 1 VAPs are secretory proteins and, thus, may be involved in host-parasite interactions (Coakley et al., 2016).

It is important to clarify here that we are not suggesting that all Group 1 VAPs from parasitic flatworms are involved in host-parasite interactions. Some Group 1 VAPs must have roles completely unrelated to parasitism as free-living flatworms also express an impressive diversity of secretory VAPs. Moreover, not all secretory proteins are actually secreted (Nielsen, 2012). That being said, secretory proteins expressed by parasitic flatworms by their very nature are more likely to exist in a host-parasite interface than intracellular proteins. This study greatly expands our knowledge of candidate genes to explore for possible roles in interacting with/evading host immune responses.

By comparing average branch lengths between our Group 1 and Group 2 VAP CRISP domain trees, our study also supports the prediction that Group 1 VAPs, on average, have greater rates of amino acid substitutions than Group 2 VAPs. This finding is significant because gene families involved in host-parasite interactions have been shown to evolve quickly and be under positive selection (Jiggins, Hurst, & Yang, 2002; Zhu & Gao, 2017). As branch lengths are a measure of molecular divergence, longer branch lengths among Group 1 VAPs than Group 2 VAPs are indicative of faster rates of molecular evolution consistent with positive selection (Bonhomme et al., 2010). Although our study does not correct for the effect of phylogeny on comparisons of branch length estimates between Group 1 and Group 2 VAPs and lacks information about the root of the VAP CRISP domain tree, Philippsen, Wilson, & DeMarco, 2015 found accelerated rates of molecular evolution in Group 1 VAPs in schistosomes.

Also consistent with the hypothesis that Group 1 VAP divergence is driven by host-specific selection, Chalmers and colleagues originally found Group 1 VAP relationships to form lingeage-specific clades, which suggests that Group 1 VAPs have undergone expansions within lineages in response to reciprocal selective pressures with host immunoproteins. In addition, Costábile, Koziol, Tort, Iriarte, & Castillo, 2018 found evidence of species-specific VAP gene duplications in the tapeworm *Mesocestoides corti*, which they suggest may help explain the wide diversity of its intermediate hosts, and Wang, Zhu, & Cai, 2017 found at least 20 tandem VAP duplications in *Schistosoma mansoni*. Our study, however, can only offer limited support to this hypothesis because we focus on transcriptomic, rather than genomic, data. Interestingly, all VAP trees, including those estimated from the CRISP domain only and those estimated from full protein sequences, contain a remarkable number of polytomies, leaving relationships within each VAP group poorly resolved. That being said, all Group 1 VAP clades with at least 50% support contain VAPs from flatworms that express similar parasitic strategies (Figures 1.5, 1.6); monophyletic Group 1 VAPs were recovered from free-living flatworms, ectoparasites, tapeworms, and digeneans.

Interestingly, all sequences in the largest Group 1 VAP clade are from free-living flatworms, which on average express more VAPs than any parasitic flatworms. This observation is possibly explained by genome reductions in parasites (Jackson, 2015), but the diversity of VAPs in free-living flatworms is curious and potentially highlights the ancestral importance of VAPs for functions unrelated to parasitism. We are unaware of any studies investigating VAP function in free-living flatworms, however, free-living flatworms are voracious predators with many possessing a variety of mechanisms, including VAPs, to suppress prey (please see (von Reumont, Campbell, & Jenner, 2014)). This should be interpreted with caution, though, as parasitic flatworms, especially cestodes and digeneans, tend to have complex lifecyles involving several hosts, and we cannot untangle the effect of developmental stage on VAP expression here, though all previously unsampled blood flukes included in this study are believed to be adults as they were extracted from vertebrate hosts.

Although our study is an important step in understanding the role of this gene family in the evolution of parasitism in flatworms, we recognize we are limited by the uncertainty of relationships both among flatworm species and among VAP members. To enable more thorough investigations of VAP gene family expansions/contractions over time, we urge future studies focus on building a robust time-calibrated flatworm tree and explore more accurate alignments that incorporate better modeling of amino acids and ideally solved VAP structures. Nonetheless, our study highlights Group 1 VAPs from parasitic flatworms as a promising system to study the evolution of parasitism.

References

Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., & Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics, 33(21), 3387–3395.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research, 25(17), 3389–3402.

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., … Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature biotechnology, 37(4), 420.

Asojo, O. A., Darwiche, R., Gebremedhin, S., Smant, G., Lozano-Torres, J. L., Drurey, C., … Wilbers, R. H. (2018). *Heligmosomoides polygyrus* Venom Allergen-like Protein-4 (HpVAL-4) is a sterol binding protein. International journal for parasitology, 48(5), 359–369.

Bankers, L., & Neiman, M. (2017). De novo transcriptome characterization of a sterilizing trematode parasite (*Microphallus* sp.) from two species of New Zealand snails. G3: Genes, Genomes, Genetics, 7(3), 871–880.

Bendtsen, J. D., Nielsen, H., von Heijne, G., & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. Journal of molecular biology, 340(4), 783–795.

Bennett, H. M., Mok, H. P., Gkrania-Klotsas, E., Tsai, I. J., Stanley, E. J., Antoun, N. M., … others (2014). The genome of the sparganosis tapeworm *Spirometra erinaceieuropaei* isolated from the biopsy of a migrating brain lesion. Genome biology, 15(11), 510.

Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., Cerqueira, G. C., … others (2009). The genome of the blood fluke *Schistosoma mansoni*. Nature, 460(7253), 352.

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., & SanCristobal, M. (2010). Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics, 186(1), 241–262.

Breen, S., Williams, S. J., Outram, M., Kobe, B., & Solomon, P. S. (2017). Emerging insights into the functions of pathogenesis-related protein 1. Trends in plant science, 22(10), 871–879.

Brown, C. T., Howe, A., Zhang, Q., Pyrkosz, A. B., & Brom, T. H. (2012). A reference-free algorithm for computational normalization of shotgun sequencing data. arXiv preprint arXiv:1203.4802.

Buhot, C., Chenal, A., Sanson, A., Pouvelle-Moratille, S., Gelb, M. H., Ménez, A., … Maillère, B. (2004). Alteration of the tertiary structure of the major bee venom allergen Api m 1 by multiple mutations is concomitant with low IgE reactivity. Protein science, 13(11), 2970–2978.

Cantacessi, C., Campbell, B., Visser, A., Geldhof, P., Nolan, M., Nisbet, A. J., … others (2009). A portrait of the "SCP/TAPS" proteins of eukaryotes—developing a framework for fundamental research and biotechnological outcomes. Biotechnology advances, 27(4), 376–388.

Cantacessi, C., Hofmann, A., Young, N. D., Broder, U., Hall, R. S., Loukas, A., & Gasser, R. B. (2012). Insights into SCP/TAPS proteins of liver flukes based on large-scale bioinformatic analyses of sequence datasets. PloS one, 7(2), e31164.

Caraballo, L., Acevedo, N., & Zakzuk, J. (2018). *Ascariasis* as a model to study the helminth/allergy relationships. Parasite immunology, e12595.

Carrillo-Bustamante, P., Keşmir, C., & de Boer, R. J. (2015). A coevolutionary arms race between hosts and viruses drives polymorphism and polygenicity of NK cell receptors. Molecular biology and evolution, 32(8), 2149–2160.

Chalmers, I. W., & Hoffmann, K. F. (2012). Platyhelminth Venom Allergen-Like (VAL) proteins: revealing structural diversity, class-specific features and biological associations across the phylum. Parasitology, 139(10), 1231–1245.

Chalmers, I. W., McArdle, A. J., Coulson, R. M., Wagner, M. A., Schmid, R., Hirai, H., & 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(1), 89.

Coakley, G., Buck, A. H., & Maizels, R. M. (2016). Host parasite communications—messages from helminths for the immune system: parasite communication and cell-cell interactions. Molecular and biochemical parasitology, 208(1), 33–40.

Cooper, D., & Eleftherianos, I. (2016). Parasitic nematode immunomodulatory strategies: recent advances and perspectives. Pathogens, 5(3), 58.

Costábile, A., Koziol, U., Tort, J. F., Iriarte, A., & Castillo, E. (2018). Expansion of cap superfamily proteins in the genome of *Mesocestoides corti*: An extreme case of a general bilaterian trend. Gene Reports, 11, 110–120.

Darwiche, R., El Atab, O., Cottier, S., & Schneiter, R. (2018). The function of yeast CAP family proteins in lipid export, mating, and pathogen defense. FEBS letters, 592(8), 1304–1311.

Darwiche, R., Kelleher, A., Hudspeth, E. M., Schneiter, R., & Asojo, O. A. (2016). Structural and functional characterization of the CAP domain of pathogen-related yeast 1 (Pry1) protein. Scientific reports, 6, 28838.

Darwiche, R., Lugo, F., Drurey, C., Varossieau, K., Smant, G., Wilbers, R. H., … Asojo, O. A. (2018). Crystal structure of *Brugia malayi* venom allergen-like protein-1 (BmVAL-1), a vaccine candidate for lymphatic filariasis. International journal for parasitology, 48(5), 371–378.

Eizaguirre, C., Lenz, T. L., Kalbe, M., & Milinski, M. (2012). Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nature communications, 3, 621.

Farias, L. P., Rodrigues, D., Cunna, V., Rofatto, H. K., Faquim-Mauro, E. L., & Leite, L. C. (2012). *Schistosoma mansoni* venom allergen like proteins present differential allergic responses in a murine model of airway inflammation. PLoS neglected tropical diseases, 6(2), e1510.

Fernandes, R. S., Barbosa, T. C., Barbosa, M. M. F., Miyasato, P. A., Nakano, E., Leite, L. C. C., & Farias, L. P. (2017). Stage and tissue expression patterns of *Schistosoma mansoni* venom allergen-like proteins SmVAL 4, 13, 16 and 24. Parasites & vectors, 10(1), 223.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics, 28(23), 3150–3152.

Fukasawa, Y., Leung, R. K., Tsui, S. K., & Horton, P. (2011). Evolutionary sequence divergence predicts protein sub-cellular localization signals. In 2011 ieee international conference on systems biology (isb) (pp. 307–312).

Gibbs, G. M., & O'Bryan, M. K. (2007). Cysteine rich secretory proteins in reproduction and venom. Society of Reproduction and Fertility supplement, 65, 261–267.

Gibbs, G. M., Roelants, K., & 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(7), 865–897.

Gibbs, G. M., Scanlon, M. J., Swarbrick, J., Curtis, S., Gallant, E., Dulhunty, A. F., & 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. Journal of Biological Chemistry, 281(7), 4156–4163.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … others (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology, 29(7), 644.

Hahn, C., Fromm, B., & Bachmann, L. (2014). Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto-and endoparastic Neodermata. Genome biology and evolution, 6(5), 1105–1117.

Harnett, W. (2014). Secretory products of helminth parasites as immunomodulators. Molecular and biochemical parasitology, 195(2), 130–136.

Hewitson, J. P., Grainger, J. R., & Maizels, R. M. (2009). Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. Molecular and biochemical parasitology, 167(1), 1–11.

Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., & Berriman, M. (2017). WormBase ParaSite-a comprehensive resource for helminth genomics. Molecular and biochemical parasitology, 215, 2–10.

Huang, Y., Chen, W., Wang, X., Liu, H., Chen, Y., Guo, L., … others (2013). The carcinogenic liver fluke, *Clonorchis sinensis*: new assembly, reannotation and analysis of the genome and characterization of tissue transcriptomes. PloS one, 8(1), e54732.

Jackson, A. P. (2015). The evolution of parasite genomes and the origins of parasitism. Parasitology, 142(S1), S1–S5.

Jiggins, F. M., Hurst, G. D., & Yang, Z. (2002). Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. Molecular Biology and Evolution, 19(8), 1341–1349.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772–780.

Kelleher, A., Darwiche, R., Rezende, W. C., Farias, L. P., Leite, L. C., Schneiter, R., & Asojo, O. A. (2014). *Schistosoma mansoni* venom allergen-like protein 4 (SmVAL4) is a novel lipid-binding SCP/TAPS protein that lacks the prototypical CAP motifs. Acta Crystallographica Section D: Biological Crystallography, 70(8), 2186–2196.

Leontovyč, R., Young, N. D., Korhonen, P. K., Hall, R. S., Tan, P., Mikeš, L., … Gasser, R. B. (2016). Comparative transcriptomic exploration reveals unique molecular adaptations of neuropathogenic *Trichobilharzia* to invade and parasitize its avian definitive host. PLoS neglected tropical diseases, 10(2), e0004406.

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13), 1658–1659.

Liu, F., Zhou, Y., Wang, Z., Lu, G., Zheng, H., Brindley, P., … others (2009). *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. Nature, 460, 345–351.

Lozano-Torres, J. L., Wilbers, R. H., Warmerdam, S., Finkers-Tomczak, A., Diaz-Granados, A., van Schaik, C. C., … others (2014). Apoplastic venom allergen-like proteins of cyst nematodes modulate the activation of basal plant innate immunity by cell surface receptors. PLoS pathogens, 10(12), e1004569.

McNulty, S. N., Tort, J. F., Rinaldi, G., Fischer, K., Rosa, B. A., Smircich, P., … others (2017). Genomes of *Fasciola hepatica* from the Americas reveal colonization with Neorickettsia endobacteria related to the agents of Potomac horse and human Sennetsu fevers. PLoS genetics, 13(1), e1006537.

Monsalve, R. I., Lu, G., et al. (1999). Expressions of recombinant venom allergen, antigen 5 of yellowjacket (*Vespula vulgaris*) and paper wasp (*Polistes annularis*), in bacteria or yeast. Protein expression and purification, 16(3), 410–416.

Nielsen, H. (2012). SignalP Frequently Asked Questions. `http://www.cbs.dtu.dk/services/SignalP/faq.php`. (Accessed: 2019-07-15)

Nielsen, H. (2017). Predicting secretory proteins with SignalP. In Protein function prediction (pp. 59–73). Springer.

Nielsen, H., Engelbrecht, J., Brunak, S., & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein engineering, 10(1), 1–6.

Obbard, D. J., Jiggins, F. M., Halligan, D. L., & Little, T. J. (2006). Natural selection drives extremely rapid evolution in antiviral RNAi genes. Current biology, 16(6), 580–585.

Orélis-Ribeiro, R., Arias, C. R., Halanych, K. M., Cribb, T. H., & Bullard, S. A. (2014). Diversity and ancestry of flatworms infecting blood of nontetrapod craniates "fishes". In Advances in parasitology (Vol. 85, pp. 1–64). Elsevier.

Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A., & Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. European journal of cell biology.

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., … others (2010). Antagonistic coevolution accelerates molecular evolution. Nature, 464(7286), 275.

Pearce, E., & Sher, A. (1987). Mechanisms of immune evasion in schistosomiasis. Contributions to microbiology and immunology, 8, 219–232.

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. Current protocols in bioinformatics, 42(1), 3–1.

Philippsen, G. S., Wilson, R. A., & DeMarco, R. (2015). Accelerated evolution of schistosome genes coding for proteins located at the host—parasite interface. Genome biology and evolution, 7(2), 431–443.

R Core Team. (2019). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Robb, S. M., Ross, E., & Alvarado, A. S. (2007). SmedGD: the Schmidtea mediterranea genome database. Nucleic acids research, 36(suppl_1), D599–D606.

Roberts, J. R., Platt, T. R., Orélis-Ribeiro, R., & Bullard, S. A. (2016). New genus of blood fluke (Digenea: Schistosomatoidea) from Malaysian freshwater turtles (Geoemydidae) and its phylogenetic position within Schistosomatoidea. Journal of Parasitology, 102(4), 451–463.

Rofatto, H. K., Parker-Manuel, S. J., Barbosa, T. C., Tararam, C. A., Wilson, R. A., Leite, L. C., & Farias, L. P. (2012). Tissue expression patterns of *Schistosoma mansoni* Venom Allergen-Like proteins 6 and 7. International journal for parasitology, 42(7), 613–620.

Rozanski, A., Moon, H., Brandl, H., Martín-Durán, J. M., Grohme, M. A., Hüttner, K., … Rink, J. C. (2018). PlanMine 3.0—improvements to a mineable resource of flatworm biology and biodiversity. Nucleic acids research, 47(D1), D812–D820.

Salzet, M., Capron, A., & Stefano, G. (2000). Molecular Crosstalk in Host–Parasite Relationships:: Schistosome–and Leech–Host Interactions. Parasitology Today, 16(12), 536–540.

Schneiter, R., & Di Pietro, A. (2013). The CAP protein superfamily: function in sterol export and fungal virulence. Biomolecular concepts, 4(5), 519–525.

Spann, N. J., & Glass, C. K. (2013). Sterols and oxysterols in immune cell function. Nature immunology, 14(9), 893.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312–1313.

Talbot, B., Balvín, O., Vonhof, M. J., Broders, H. G., Fenton, B., & Keyghobadi, N. (2017). Host association and selection on salivary protein genes in bed bugs and related blood-feeding ectoparasites. Royal Society open science, 4(6), 170446.

Tsai, I. J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K. L., … others (2013). The genomes of four tapeworm species reveal adaptations to parasitism. Nature, 496(7443), 57.

von Reumont, B., Campbell, L., & Jenner, R. (2014). Quo vadis venomics? A roadmap to neglected venomous invertebrates. Toxins, 6(12), 3488–3551.

Wang, S., Wang, S., Luo, Y., Xiao, L., Luo, X., Gao, S., … others (2016). Comparative genomics reveals adaptive evolution of Asian tapeworm in switching to a new intermediate host. Nature communications, 7, 12845.

Wang, S., Zhu, X.-q., & Cai, X. (2017). Gene duplication analysis reveals no ancient whole genome duplication but extensive small-scale duplications during genome evolution and adaptation of *Schistosoma mansoni*. Frontiers in Cellular and Infection Microbiology, 7, 412.

Wasik, K., Gurtowski, J., Zhou, X., Ramos, O. M., Delás, M. J., Battistoni, G., … others (2015). Genome and transcriptome of the regeneration-competent flatworm, *Macrostomum lignano*. Proceedings of the National Academy of Sciences, 112(40), 12462–12467.

Weinstein, S. B., & Kuris, A. M. (2016). Independent origins of parasitism in Animalia. Biology letters, 12(7).

Whelan, N. V., Kocot, K. M., Moroz, L. L., & Halanych, K. M. (2015). Error, signal, and the placement of Ctenophora sister to all other animals. Proceedings of the National Academy of Sciences, 112(18), 5773–5778.

White, J. D. (2009). Select_Contigs.pl. `https://github.com/chrishah/phylog/blob/master/scripts-external/select_contigs.pl`. (Accessed: 2019-07-15)

Windsor, D. A. (1998). Controversies in parasitology, most of the species on earth are parasites. International journal for parasitology, 28(12), 1939–1941.

Young, N. D., Jex, A. R., Li, B., Liu, S., Yang, L., Xiong, Z., … others (2012). Whole-genome sequence of *Schistosoma haematobium*. Nature genetics, 44(2), 221.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution, 8(1), 28–36.

Zhu, S., & Gao, B. (2017). Positive selection in cathelicidin host defense peptides: adaptation to exogenous pathogens or endogenous receptors? Heredity, 118(5), 453.

Appendices

Appendix A

Supplementary Tables

Table A.1: Organism collection data for the previously unsampled non-schistosome blood fluke transcriptomes reported in this study

Taxa	Genus	species	Host	Host Genus	Host species	Site in host	Locality
Aporocotylidae	*Acipensericola*	*petersoni*	American paddlefish	*Polyodon*	*spathula*	heart	Tennessee River
Aporocotylidae	*Cardicola*	*currani*	Red drum	*Sciaenops*	*ocellatus*	heart	Gulf of Mexico
Aporocotylidae	*Cardicola*	*palmeri*	Black drum	*Pogonias*	*cromis*	heart	Gulf of Mexico
Aporocotylidae	*Elaphrobates*	*euzeti*	Northern red snapper	*Lutjanus*	*campechanus*	heart	Gulf of Mexico
Aporocotylidae	*Elopicola*	*bristowi*	Hawaiian ladyfish	*Elops*	*hawaiensis*	viscera wash	South China Sea
Aporocotylidae	*Elopicola*	*franksi*	Atlantic tarpon	*Megalops*	*atlanticus*	viscera wash	Gulf of Mexico
Aporocotylidae	*Elopicola*	*nolancribbi*	Northern ladyfish/ tenpounder	*Elops*	*saurus*	viscera wash	Gulf of Mexico
Aporocotylidae	*Littorellicola*	*billhawkinsi*	Florida pompano	*Trachinotus*	*carolinus*	heart	Gulf of Mexico
Aporocotylidae	*Myliobaticola*	sp. nov.	Caribbean electric ray	*Narcine*	*bancroftii*	heart	Gulf of Mexico
Aporocotylidae	*Myliobaticola*	*richardheardi*	Atlantic stingray	*Dasyatis*	*sabina*	heart	Gulf of Mexico
Aporocotylidae	gen. nov.	sp. nov.	Thai yellow catfish?	*Pangasius*	*siamensis*	body cavity	Mekong River
Aporocotylidae	*Nomasanguinicola*	*canthoensis*	Broadhead catfish	*Clarias*	*macrocephalus*	branchial vessels	Mekong River
Aporocotylidae	*Phthinomita*	sp.	Orange-spotted spinefoot	*Siganus*	*cf.guttatus*	heart	South China Sea
Aporocotylidae	*Psettarium*	*anthicum*	Cobia	*Rachycentron*	*canadum*	heart	South China Sea
Aporocotylidae	*Selachohemecus*	*olsoni*	Atlantic sharpnose shark	*Rhizoprionodon*	*terranovae*	heart	Gulf of Mexico
Spirochiidae	*Coeuritrema*	*platti*	Chinese softshell turtle	*Pelodiscus*	*sinensis*	blood	Da Rang River
Spirochiidae	*Hapalorhynchus*	*foliorchis*	Common snapping turtle	*Chelydra*	*serpentina*	mesentery	E.W. Shell, Tallapoosa River, AL
Spirochiidae	*Spirorchis*	*haematobius*	Common snapping turtle	*Chelydra*	*serpentina*	heart	Canoe Lake, Coosa River, AL
Spirochiidae	*Spirorchis*	*picta*	Pond slider turtle	*Trachemys*	*scripta*	mesentery	E.W. Shell, Tallapoosa River, AL
Spirochiidae	*Spirorchis*	*scripta*	Pond slider turtle	*Trachemys*	*scripta*	vasculature of the eye	E. W. Shell Aquaculture Station, Tallapoosa River, Auburn, AL (32°38’57.88”N, 85°29’4.81”W)
Spirochiidae	*Vasotrema*	sp.	Spiny softshell turtle	*Apalone*	*spinifera*	mesentery	Perry Lakes Oxbow, Cahaba River, Marion, AL (32°41’50.91”N, 87°14’30.39”W)

Table A.2: Venom allergen-like protein filtering summary by bioinformatics step.

Step	Description	Count	File Name[1]
0	Total contigs	2632801	`transcriptomes.tgz`
1	VAL homologs	1456	`s01_VALhomologs.fasta`
2	Predicted VAL open reading frames	1908	`s02_longest_orfs.pep`
3	Putative VALs	1461	`s03_putativeVALs.pep`
4	Complete putative VALs (pVAL)	849	`s04_putativeVALs_complete.pep`
5	pVAL with CRISP domain	803	`s05_putativeVALs_complete_CRISP.pep`
6	Unique pVAL with CRISP domain	624	`s06_putativeVALs_complete_CRISP_unique.pep`
7	New unique pVAL with CRISP domain	557	`s07_putativeVALs_complete_CRISP_unique_new.pep`
8	VALs	474	`s08_VAL_Sipley_474.fasta`
9	VALs with predicted signal peptide	187	`s09_VAL_Sipley_474_signalP.fasta`

---

[1] All scripts used to generate these files and the files themselves are available at `https://github.com/Sipley/MS-thesis`

Table A.3: Characterization of all novel flatworm venom allergen-like proteins predicted in this study

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP9_Nomasanguinicola_canthoensis_trematoda_aporocotylidae_bloodFluke_contig11946	197	yes	no	no
VAP14_Phthinomita_spp_trematoda_aporocotylidae_bloodFluke_contig11164	192	yes	no	no
VAP15_Phthinomita_spp_trematoda_aporocotylidae_bloodFluke_contig11165	322	yes	no	no
VAP16_Phthinomita_spp_trematoda_aporocotylidae_bloodFluke_contig17821	148	yes	no	no
VAP19_Planaria_torva_turbellaria_contig2110	365	yes	no	no
VAP49_Acipensericola_petersoni_trematoda_aporocotylidae_bloodFluke_contig70976	247	yes	no	no
VAP66_Polycelis_tenuis_turbellaria_contig9016	224	yes	no	no
VAP68_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig2076	502	yes	no	no
VAP84_Polycelis_tenuis_turbellaria_contig22419	428	yes	no	no
VAP103_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig5995	313	yes	no	no
VAP109_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig7850	290	yes	no	no
VAP113_Schistocephalus_solidus_cestoda_schistocephalidae_contig6999	414	yes	no	no
VAP118_Schistosoma_haematobium_trematoda_schistosomatidae_bloodFluke_contig418	412	yes	no	no
VAP124_Schistosoma_haematobium_trematoda_schistosomatidae_bloodFluke_contig7983	400	yes	no	no
VAP125_Schistosoma_haematobium_trematoda_schistosomatidae_bloodFluke_contig8399	366	yes	no	no
VAP127_Schistosoma_haematobium_trematoda_schistosomatidae_bloodFluke_contig11539	205	yes	no	no
VAP133_Schistosoma_japonicum_trematoda_schistosomatidae_bloodFluke_contig866	354	yes	no	no
VAP146_Schistosoma_japonicum_trematoda_schistosomatidae_bloodFluke_contig11149	174	yes	no	no
VAP153_Schistosoma_mansoni_trematoda_schistosomatidae_bloodFluke_contig162	231	yes	no	no
VAP156_Schistosoma_mansoni_trematoda_schistosomatidae_bloodFluke_contig3091	148	yes	no	no
VAP159_Schistosoma_mansoni_trematoda_schistosomatidae_bloodFluke_contig4763	143	yes	no	no
VAP171_Schistosoma_mansoni_trematoda_schistosomatidae_bloodFluke_contig6585	197	yes	no	no
VAP180_Schistosoma_mansoni_trematoda_schistosomatidae_bloodFluke_contig10047	272	yes	no	no
VAP184_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig10327	220	yes	no	no
VAP189_Acipensericola_petersoni_trematoda_aporocotylidae_bloodFluke_contig70982	155	yes	no	no
VAP195_Schmidtea_mediterranea_turbellaria_contig9933	157	yes	no	no
VAP222_Schmidtea_mediterranea_turbellaria_contig24635	144	yes	no	no
VAP232_Schmidtea_polychroa_turbellaria_contig2087	238	yes	no	no
VAP240_Schmidtea_polychroa_turbellaria_contig8122	225	yes	no	no
VAP246_Schmidtea_polychroa_turbellaria_contig16155	354	yes	no	no
VAP249_Schmidtea_polychroa_turbellaria_contig17169	135	yes	no	no
VAP250_Schmidtea_polychroa_turbellaria_contig17170	226	yes	no	no

Continued on next page

Table A.3 – *Continued from previous page*

ID	Length	Group 1	Group 2	SignalP
VAP279_Spirorchis_picta_trematoda_spirochiidae_bloodFluke_contig6649	257	yes	no	no
VAP281_Spirorchis_picta_trematoda_spirochiidae_bloodFluke_contig17849	193	yes	no	no
VAP282_Spirorchis_picta_trematoda_spirochiidae_bloodFluke_contig18193	200	yes	no	no
VAP289_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig18447	188	yes	no	no
VAP290_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig18450	164	yes	no	no
VAP304_Taenia_asiatica_cestoda_taeniidae_contig1031	236	yes	no	no
VAP312_Taenia_saginata_cestoda_taeniidae_contig1404	467	yes	no	no
VAP318_Taenia_saginata_cestoda_taeniidae_contig4756	264	yes	no	no
VAP327_Taenia_saginata_cestoda_taeniidae_contig10595	216	yes	no	no
VAP329_Taenia_saginata_cestoda_taeniidae_contig10600	266	yes	no	no
VAP331_Taenia_saginata_cestoda_taeniidae_contig11572	345	yes	no	no
VAP345_Taenia_solium_cestoda_taeniidae_contig11004	186	yes	no	no
VAP346_Taenia_solium_cestoda_taeniidae_contig11208	264	yes	no	no
VAP349_Taenia_solium_cestoda_taeniidae_contig12296	242	yes	no	no
VAP358_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig2969	194	yes	no	no
VAP367_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig10852	113	yes	no	no
VAP379_Vasotrema_spp_trematoda_spirochiidae_bloodFluke_contig3718	195	yes	no	no
VAP385_Vasotrema_spp_trematoda_spirochiidae_bloodFluke_contig11881	136	yes	no	no
VAP386_Acipensericola_petersoni_trematoda_aporocotylidae_bloodFluke_contig70990	255	yes	no	no
VAP406_Dendrocoelum_lacteum_turbellaria_contig23647	180	yes	no	no
VAP408_Dendrocoelum_lacteum_turbellaria_contig25449	311	yes	no	no
VAP409_Dendrocoelum_lacteum_turbellaria_contig26450	157	yes	no	no
VAP410_Dendrocoelum_lacteum_turbellaria_contig26451	144	yes	no	no
VAP417_Acipensericola_petersoni_trematoda_aporocotylidae_bloodFluke_contig71007	145	yes	no	no
VAP422_Dendrocoelum_lacteum_turbellaria_contig55280	139	yes	no	no
VAP423_Dendrocoelum_lacteum_turbellaria_contig55281	149	yes	no	no
VAP432_Echinococcus_multilocularis_cestoda_taeniidae_contig5189	230	yes	no	no
VAP434_Echinococcus_multilocularis_cestoda_taeniidae_contig5687	251	yes	no	no
VAP447_Elopicola_bristowi_trematoda_aporocotylidae_bloodFluke_contig28210	195	yes	no	no
VAP455_Fasciola_hepatica_trematoda_echinostomatidae_NBF_contig2423	158	yes	no	no
VAP460_Fasciola_hepatica_trematoda_echinostomatidae_NBF_contig7506	176	yes	no	no
VAP463_Fasciola_hepatica_trematoda_echinostomatidae_NBF_contig11316	269	yes	no	no

Continued on next page

Table A.3 – *Continued from previous page*

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP466_Gyrodactylus_salaris_monogenea_contig1917	283	yes	no	no
VAP469_Gyrodactylus_salaris_monogenea_contig2794	635	yes	no	no
VAP474_Gyrodactylus_salaris_monogenea_contig4115	147	yes	no	no
VAP478_Gyrodactylus_salaris_monogenea_contig4554	641	yes	no	no
VAP481_Gyrodactylus_salaris_monogenea_contig9135	539	yes	no	no
VAP483_Gyrodactylus_salaris_monogenea_contig10666	138	yes	no	no
VAP489_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig8099	598	yes	no	no
VAP494_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig22364	129	yes	no	no
VAP512_Hymenolepis_diminuta_cestoda_hymenolepididae_contig5577	160	yes	no	no
VAP519_Hymenolepis_microstoma_cestoda_hymenolepididae_contig211	316	yes	no	no
VAP520_Hymenolepis_microstoma_cestoda_hymenolepididae_contig1738	229	yes	no	no
VAP527_Hymenolepis_microstoma_cestoda_hymenolepididae_contig7392	260	yes	no	no
VAP528_Hymenolepis_microstoma_cestoda_hymenolepididae_contig8387	250	yes	no	no
VAP529_Hymenolepis_microstoma_cestoda_hymenolepididae_contig8388	215	yes	no	no
VAP535_Hymenolepis_microstoma_cestoda_hymenolepididae_contig10036	145	yes	no	no
VAP541_Littorellicola_billhawkinsi_trematoda_aporocotylidae_bloodFluke_contig12418	204	yes	no	no
VAP579_Mesocestoides_corti_cestoda_mesocestoididae_contig2005	127	yes	no	no
VAP583_Mesocestoides_corti_cestoda_mesocestoididae_contig2761	238	yes	no	no
VAP584_Mesocestoides_corti_cestoda_mesocestoididae_contig3733	119	yes	no	no
VAP594_Mesocestoides_corti_cestoda_mesocestoididae_contig8730	190	yes	no	no
VAP596_Mesocestoides_corti_cestoda_mesocestoididae_contig8882	162	yes	no	no
VAP599_Mesocestoides_corti_cestoda_mesocestoididae_contig9076	143	yes	no	no
VAP600_Mesocestoides_corti_cestoda_mesocestoididae_contig9399	159	yes	no	no
VAP602_Mesocestoides_corti_cestoda_mesocestoididae_contig9401	252	yes	no	no
VAP4_nGen_nsp_trematoda_aporocotylidae_bloodFluke_contig48357	177	yes	no	yes
VAP10_Paragonimus_westermani_trematoda_troglotrematidae_NBF_contig6407	180	yes	no	yes
VAP20_Planaria_torva_turbellaria_contig8108	227	yes	no	yes
VAP22_Planaria_torva_turbellaria_contig9240	198	yes	no	yes
VAP27_Planaria_torva_turbellaria_contig16714	200	yes	no	yes
VAP30_Planaria_torva_turbellaria_contig21661	223	yes	no	yes
VAP31_Planaria_torva_turbellaria_contig21678	199	yes	no	yes
VAP36_Planaria_torva_turbellaria_contig24776	199	yes	no	yes

Continued on next page

Table A.3 – *Continued from previous page*

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP38_Planaria_torva_turbellaria_contig26043	325	yes	no	yes
VAP39_Planaria_torva_turbellaria_contig32587	200	yes	no	yes
VAP40_Planaria_torva_turbellaria_contig33042	221	yes	no	yes
VAP41_Planaria_torva_turbellaria_contig33043	207	yes	no	yes
VAP43_Polycelis_nigra_turbellaria_contig249	214	yes	no	yes
VAP44_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig795	190	yes	no	yes
VAP45_Polycelis_nigra_turbellaria_contig3867	214	yes	no	yes
VAP47_Polycelis_nigra_turbellaria_contig8413	244	yes	no	yes
VAP48_Polycelis_nigra_turbellaria_contig10064	323	yes	no	yes
VAP54_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig1051	317	yes	no	yes
VAP55_Polycelis_nigra_turbellaria_contig20709	202	yes	no	yes
VAP56_Polycelis_nigra_turbellaria_contig20730	227	yes	no	yes
VAP57_Polycelis_nigra_turbellaria_contig31401	691	yes	no	yes
VAP60_Polycelis_nigra_turbellaria_contig34878	225	yes	no	yes
VAP61_Polycelis_nigra_turbellaria_contig35623	192	yes	no	yes
VAP64_Polycelis_nigra_turbellaria_contig40397	224	yes	no	yes
VAP65_Polycelis_nigra_turbellaria_contig44158	221	yes	no	yes
VAP67_Polycelis_tenuis_turbellaria_contig10466	219	yes	no	yes
VAP69_Polycelis_tenuis_turbellaria_contig11254	221	yes	no	yes
VAP70_Polycelis_tenuis_turbellaria_contig11269	202	yes	no	yes
VAP71_Polycelis_tenuis_turbellaria_contig11635	214	yes	no	yes
VAP74_Polycelis_tenuis_turbellaria_contig14475	191	yes	no	yes
VAP82_Polycelis_tenuis_turbellaria_contig21498	211	yes	no	yes
VAP85_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig4632	233	yes	no	yes
VAP86_Polycelis_tenuis_turbellaria_contig23773	323	yes	no	yes
VAP87_Polycelis_tenuis_turbellaria_contig27876	257	yes	no	yes
VAP89_Polycelis_tenuis_turbellaria_contig33558	231	yes	no	yes
VAP90_Polycelis_tenuis_turbellaria_contig33559	226	yes	no	yes
VAP95_Polycelis_tenuis_turbellaria_contig46930	192	yes	no	yes
VAP101_Psettarium_anthicum_trematoda_aporocotylidae_bloodFluke_contig55244	198	yes	no	yes
VAP112_Schistocephalus_solidus_cestoda_schistocephalidae_contig5442	282	yes	no	yes
VAP129_Schistosoma_haematobium_trematoda_schistosomatidae_bloodFluke_contig12456	232	yes	no	yes

Continued on next page

Table A.3 – *Continued from previous page*

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP130_Schistosoma_haematobium_trematoda_schistosomatidae_bloodFluke_contig12736	196	yes	no	yes
VAP134_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig8914	196	yes	no	yes
VAP142_Schistosoma_japonicum_trematoda_schistosomatidae_bloodFluke_contig8130	185	yes	no	yes
VAP143_Schistosoma_japonicum_trematoda_schistosomatidae_bloodFluke_contig10544	173	yes	no	yes
VAP148_Schistosoma_japonicum_trematoda_schistosomatidae_bloodFluke_contig12396	191	yes	no	yes
VAP151_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig9240	251	yes	no	yes
VAP175_Schistosoma_mansoni_trematoda_schistosomatidae_bloodFluke_contig8251	236	yes	no	yes
VAP183_Schistosoma_mansoni_trematoda_schistosomatidae_bloodFluke_contig11629	149	yes	no	yes
VAP187_Schmidtea_mediterranea_turbellaria_contig4340	204	yes	no	yes
VAP188_Schmidtea_mediterranea_turbellaria_contig5147	178	yes	no	yes
VAP205_Schmidtea_mediterranea_turbellaria_contig14888	200	yes	no	yes
VAP217_Schmidtea_mediterranea_turbellaria_contig23044	201	yes	no	yes
VAP221_Schmidtea_mediterranea_turbellaria_contig24468	217	yes	no	yes
VAP224_Schmidtea_mediterranea_turbellaria_contig25826	150	yes	no	yes
VAP227_Schmidtea_mediterranea_turbellaria_contig27976	196	yes	no	yes
VAP231_Schmidtea_polychroa_turbellaria_contig1320	200	yes	no	yes
VAP233_Schmidtea_polychroa_turbellaria_contig2225	243	yes	no	yes
VAP234_Schmidtea_polychroa_turbellaria_contig2371	201	yes	no	yes
VAP235_Schmidtea_polychroa_turbellaria_contig3940	201	yes	no	yes
VAP236_Schmidtea_polychroa_turbellaria_contig4339	199	yes	no	yes
VAP237_Schmidtea_polychroa_turbellaria_contig7120	244	yes	no	yes
VAP239_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig7963	228	yes	no	yes
VAP245_Schmidtea_polychroa_turbellaria_contig15886	270	yes	no	yes
VAP251_Schmidtea_polychroa_turbellaria_contig17379	324	yes	no	yes
VAP253_Schmidtea_polychroa_turbellaria_contig19813	216	yes	no	yes
VAP256_Schmidtea_polychroa_turbellaria_contig28186	607	yes	no	yes
VAP264_Selachohemecus_olsoni_trematoda_aporocotylidae_bloodFluke_contig74927	222	yes	no	yes
VAP265_Spirometra_erinaceieuropaei_cestoda_diphylobothidae_contig7658	160	yes	no	yes
VAP266_Spirometra_erinaceieuropaei_cestoda_diphylobothidae_contig11319	167	yes	no	yes
VAP280_Spirorchis_picta_trematoda_spirochiidae_bloodFluke_contig14159	223	yes	no	yes
VAP292_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig29289	193	yes	no	yes
VAP302_Taenia_asiatica_cestoda_taeniidae_contig344	232	yes	no	yes

Continued on next page

Table A.3 – *Continued from previous page*

ID	Length	Group 1	Group 2	SignalP
VAP303_Taenia_asiatica_cestoda_taeniidae_contig486	165	yes	no	yes
VAP305_Taenia_asiatica_cestoda_taeniidae_contig2081	230	yes	no	yes
VAP310_Taenia_asiatica_cestoda_taeniidae_contig5331	284	yes	no	yes
VAP311_Taenia_asiatica_cestoda_taeniidae_contig6171	204	yes	no	yes
VAP313_Taenia_saginata_cestoda_taeniidae_contig1661	213	yes	no	yes
VAP323_Taenia_saginata_cestoda_taeniidae_contig7593	333	yes	no	yes
VAP324_Taenia_saginata_cestoda_taeniidae_contig8913	241	yes	no	yes
VAP325_Taenia_saginata_cestoda_taeniidae_contig10000	207	yes	no	yes
VAP326_Taenia_saginata_cestoda_taeniidae_contig10133	213	yes	no	yes
VAP328_Taenia_saginata_cestoda_taeniidae_contig10596	197	yes	no	yes
VAP330_Taenia_saginata_cestoda_taeniidae_contig10602	336	yes	no	yes
VAP332_Taenia_saginata_cestoda_taeniidae_contig12893	233	yes	no	yes
VAP334_Taenia_solium_cestoda_taeniidae_contig2829	229	yes	no	yes
VAP339_Taenia_solium_cestoda_taeniidae_contig5724	160	yes	no	yes
VAP341_Taenia_solium_cestoda_taeniidae_contig7398	166	yes	no	yes
VAP343_Taenia_solium_cestoda_taeniidae_contig8338	197	yes	no	yes
VAP348_Taenia_solium_cestoda_taeniidae_contig12089	224	yes	no	yes
VAP351_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig969	140	yes	no	yes
VAP352_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig1528	297	yes	no	yes
VAP353_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig1768	217	yes	no	yes
VAP354_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig2083	177	yes	no	yes
VAP355_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig2084	156	yes	no	yes
VAP356_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig2085	163	yes	no	yes
VAP357_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig2968	214	yes	no	yes
VAP361_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig4105	164	yes	no	yes
VAP362_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig6156	246	yes	no	yes
VAP363_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig6743	154	yes	no	yes
VAP375_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig17768	169	yes	no	yes
VAP380_Vasotrema_spp_trematoda_spirochiidae_bloodFluke_contig5871	188	yes	no	yes
VAP383_Vasotrema_spp_trematoda_spirochiidae_bloodFluke_contig11792	182	yes	no	yes
VAP384_Vasotrema_spp_trematoda_spirochiidae_bloodFluke_contig11794	144	yes	no	yes
VAP388_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig54031	163	yes	no	yes

Continued on next page

Table A.3 – *Continued from previous page*

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP389_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig54032	193	yes	no	yes
VAP390_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig54035	193	yes	no	yes
VAP393_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig105712	189	yes	no	yes
VAP394_Dendrocoelum_lacteum_turbellaria_contig5116	217	yes	no	yes
VAP395_Dendrocoelum_lacteum_turbellaria_contig12388	203	yes	no	yes
VAP396_Dendrocoelum_lacteum_turbellaria_contig13693	202	yes	no	yes
VAP399_Dendrocoelum_lacteum_turbellaria_contig16778	215	yes	no	yes
VAP400_Dendrocoelum_lacteum_turbellaria_contig17850	222	yes	no	yes
VAP401_Dendrocoelum_lacteum_turbellaria_contig18304	221	yes	no	yes
VAP402_Dendrocoelum_lacteum_turbellaria_contig19466	223	yes	no	yes
VAP403_Dendrocoelum_lacteum_turbellaria_contig20292	325	yes	no	yes
VAP404_Dendrocoelum_lacteum_turbellaria_contig20900	215	yes	no	yes
VAP407_Dendrocoelum_lacteum_turbellaria_contig24199	241	yes	no	yes
VAP412_Dendrocoelum_lacteum_turbellaria_contig30496	359	yes	no	yes
VAP424_Echinococcus_multilocularis_cestoda_taeniidae_contig618	231	yes	no	yes
VAP427_Echinococcus_multilocularis_cestoda_taeniidae_contig2194	288	yes	no	yes
VAP433_Echinococcus_multilocularis_cestoda_taeniidae_contig5453	210	yes	no	yes
VAP437_Echinococcus_multilocularis_cestoda_taeniidae_contig6349	211	yes	no	yes
VAP438_Echinococcus_multilocularis_cestoda_taeniidae_contig9194	215	yes	no	yes
VAP439_Echinococcus_multilocularis_cestoda_taeniidae_contig9783	221	yes	no	yes
VAP441_Echinococcus_multilocularis_cestoda_taeniidae_contig10518	200	yes	no	yes
VAP458_Cardicola_currani_trematoda_aporocotylidae_bloodFluke_contig16180	193	yes	no	yes
VAP462_Fasciola_hepatica_trematoda_echinostomatidae_NBF_contig11093	220	yes	no	yes
VAP464_Fasciola_hepatica_trematoda_echinostomatidae_NBF_contig11941	204	yes	no	yes
VAP468_Gyrodactylus_salaris_monogenea_contig2384	183	yes	no	yes
VAP471_Gyrodactylus_salaris_monogenea_contig3433	179	yes	no	yes
VAP473_Gyrodactylus_salaris_monogenea_contig4050	219	yes	no	yes
VAP475_Gyrodactylus_salaris_monogenea_contig4204	165	yes	no	yes
VAP476_Gyrodactylus_salaris_monogenea_contig4312	184	yes	no	yes
VAP482_Gyrodactylus_salaris_monogenea_contig9138	161	yes	no	yes
VAP491_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig19739	235	yes	no	yes
VAP492_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig19740	241	yes	no	yes

Continued on next page

Table A.3 – *Continued from previous page*

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP493_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig22363	197	yes	no	yes
VAP495_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig22370	197	yes	no	yes
VAP496_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig22373	137	yes	no	yes
VAP497_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig22374	195	yes	no	yes
VAP498_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig22375	146	yes	no	yes
VAP499_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig22377	204	yes	no	yes
VAP501_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig22378	143	yes	no	yes
VAP502_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig22380	202	yes	no	yes
VAP508_Hymenolepis_diminuta_cestoda_hymenolepididae_contig699	247	yes	no	yes
VAP509_Hymenolepis_diminuta_cestoda_hymenolepididae_contig700	248	yes	no	yes
VAP510_Hymenolepis_diminuta_cestoda_hymenolepididae_contig3694	194	yes	no	yes
VAP515_Hymenolepis_diminuta_cestoda_hymenolepididae_contig7967	226	yes	no	yes
VAP516_Hymenolepis_diminuta_cestoda_hymenolepididae_contig8581	329	yes	no	yes
VAP517_Hymenolepis_diminuta_cestoda_hymenolepididae_contig10418	215	yes	no	yes
VAP522_Hymenolepis_microstoma_cestoda_hymenolepididae_contig4345	203	yes	no	yes
VAP523_Hymenolepis_microstoma_cestoda_hymenolepididae_contig5408	229	yes	no	yes
VAP524_Hymenolepis_microstoma_cestoda_hymenolepididae_contig5409	269	yes	no	yes
VAP525_Hymenolepis_microstoma_cestoda_hymenolepididae_contig6933	227	yes	no	yes
VAP526_Hymenolepis_microstoma_cestoda_hymenolepididae_contig6934	319	yes	no	yes
VAP530_Hymenolepis_microstoma_cestoda_hymenolepididae_contig8389	223	yes	no	yes
VAP531_Hymenolepis_microstoma_cestoda_hymenolepididae_contig8395	198	yes	no	yes
VAP533_Hymenolepis_microstoma_cestoda_hymenolepididae_contig9163	232	yes	no	yes
VAP534_Hymenolepis_microstoma_cestoda_hymenolepididae_contig10004	195	yes	no	yes
VAP536_Hymenolepis_microstoma_cestoda_hymenolepididae_contig10201	190	yes	no	yes
VAP539_Hymenolepis_microstoma_cestoda_hymenolepididae_contig11835	196	yes	no	yes
VAP543_Littorellicola_billhawkinsi_trematoda_aporocotylidae_bloodFluke_contig19766	198	yes	no	yes
VAP556_Macrostomum_lignano_turbellaria_contig18726	396	yes	no	yes
VAP562_Macrostomum_lignano_turbellaria_contig25068	398	yes	no	yes
VAP563_Macrostomum_lignano_turbellaria_contig28003	824	yes	no	yes
VAP569_Cardicola_palmeri_trematoda_aporocotylidae_bloodFluke_contig20136	193	yes	no	yes
VAP570_Macrostomum_lignano_turbellaria_contig35936	257	yes	no	yes
VAP580_Mesocestoides_corti_cestoda_mesocestoididae_contig2405	222	yes	no	yes

Continued on next page

Table A.3 – *Continued from previous page*

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP586_Mesocestoides_corti_cestoda_mesocestoididae_contig4360	235	yes	no	yes
VAP587_Mesocestoides_corti_cestoda_mesocestoididae_contig5237	242	yes	no	yes
VAP588_Mesocestoides_corti_cestoda_mesocestoididae_contig5540	237	yes	no	yes
VAP589_Mesocestoides_corti_cestoda_mesocestoididae_contig6212	189	yes	no	yes
VAP591_Mesocestoides_corti_cestoda_mesocestoididae_contig6499	164	yes	no	yes
VAP593_Mesocestoides_corti_cestoda_mesocestoididae_contig8729	164	yes	no	yes
VAP595_Mesocestoides_corti_cestoda_mesocestoididae_contig8881	164	yes	no	yes
VAP597_Mesocestoides_corti_cestoda_mesocestoididae_contig9074	204	yes	no	yes
VAP598_Mesocestoides_corti_cestoda_mesocestoididae_contig9075	174	yes	no	yes
VAP601_Mesocestoides_corti_cestoda_mesocestoididae_contig9400	214	yes	no	yes
VAP607_Microphallus_livelyi_trematoda_microphallidae_cercaria_NBF_contig6021	205	yes	no	yes
VAP609_Microphallus_livelyi_trematoda_microphallidae_cercaria_NBF_contig12409	211	yes	no	yes
VAP611_Microphallus_livelyi_trematoda_microphallidae_cercaria_NBF_contig14777	206	yes	no	yes
VAP613_Myliobaticola_nsp_trematoda_aporocotylidae_bloodFluke_contig12117	260	yes	no	yes
VAP614_Myliobaticola_nsp_trematoda_aporocotylidae_bloodFluke_contig14779	236	yes	no	yes
VAP619_Myliobaticola_nsp_trematoda_aporocotylidae_bloodFluke_contig23280	215	yes	no	yes
VAP620_Myliobaticola_nsp_trematoda_aporocotylidae_bloodFluke_contig23284	192	yes	no	yes
VAP2_nGen_nsp_trematoda_aporocotylidae_bloodFluke_contig46364	246	no	yes	no
VAP3_nGen_nsp_trematoda_aporocotylidae_bloodFluke_contig46372	288	no	yes	no
VAP5_nGen_nsp_trematoda_aporocotylidae_bloodFluke_contig59972	391	no	yes	no
VAP7_nGen_nsp_trematoda_aporocotylidae_bloodFluke_contig59976	237	no	yes	no
VAP8_nGen_nsp_trematoda_aporocotylidae_bloodFluke_contig59976	186	no	yes	no
VAP17_Planaria_torva_turbellaria_contig1003	171	no	yes	no
VAP18_Planaria_torva_turbellaria_contig1004	312	no	yes	no
VAP21_Planaria_torva_turbellaria_contig8164	411	no	yes	no
VAP23_Planaria_torva_turbellaria_contig9917	431	no	yes	no
VAP24_Planaria_torva_turbellaria_contig11406	264	no	yes	no
VAP25_Planaria_torva_turbellaria_contig12790	156	no	yes	no
VAP26_Planaria_torva_turbellaria_contig15938	156	no	yes	no
VAP28_Planaria_torva_turbellaria_contig17078	218	no	yes	no
VAP29_Planaria_torva_turbellaria_contig18866	262	no	yes	no
VAP32_Planaria_torva_turbellaria_contig22706	250	no	yes	no

Continued on next page

Table A.3 – *Continued from previous page*

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP33_Planaria_torva_turbellaria_contig22707	258	no	yes	no
VAP34_Planaria_torva_turbellaria_contig23403	429	no	yes	no
VAP35_Planaria_torva_turbellaria_contig23581	381	no	yes	no
VAP37_Planaria_torva_turbellaria_contig25642	315	no	yes	no
VAP42_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig723	164	no	yes	no
VAP46_Polycelis_nigra_turbellaria_contig7052	156	no	yes	no
VAP50_Polycelis_nigra_turbellaria_contig16108	283	no	yes	no
VAP51_Polycelis_nigra_turbellaria_contig16109	415	no	yes	no
VAP52_Polycelis_nigra_turbellaria_contig16110	249	no	yes	no
VAP53_Polycelis_nigra_turbellaria_contig16795	250	no	yes	no
VAP58_Polycelis_nigra_turbellaria_contig33331	315	no	yes	no
VAP59_Polycelis_nigra_turbellaria_contig33729	379	no	yes	no
VAP62_Polycelis_nigra_turbellaria_contig35710	256	no	yes	no
VAP63_Polycelis_nigra_turbellaria_contig37160	220	no	yes	no
VAP72_Polycelis_tenuis_turbellaria_contig12347	156	no	yes	no
VAP73_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig3102	225	no	yes	no
VAP75_Polycelis_tenuis_turbellaria_contig14704	220	no	yes	no
VAP76_Polycelis_tenuis_turbellaria_contig15453	165	no	yes	no
VAP77_Polycelis_tenuis_turbellaria_contig15454	315	no	yes	no
VAP78_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig4555	150	no	yes	no
VAP79_Polycelis_tenuis_turbellaria_contig15853	256	no	yes	no
VAP80_Polycelis_tenuis_turbellaria_contig17101	379	no	yes	no
VAP83_Polycelis_tenuis_turbellaria_contig22209	147	no	yes	no
VAP88_Polycelis_tenuis_turbellaria_contig29577	384	no	yes	no
VAP91_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig4753	450	no	yes	no
VAP92_Polycelis_tenuis_turbellaria_contig40808	424	no	yes	no
VAP93_Polycelis_tenuis_turbellaria_contig41083	841	no	yes	no
VAP94_Polycelis_tenuis_turbellaria_contig41086	841	no	yes	no
VAP99_Psettarium_anthicum_trematoda_aporocotylidae_bloodFluke_contig42549	193	no	yes	no
VAP100_Psettarium_anthicum_trematoda_aporocotylidae_bloodFluke_contig42550	195	no	yes	no
VAP102_Psettarium_anthicum_trematoda_aporocotylidae_bloodFluke_contig58915	446	no	yes	no
VAP104_Psettarium_anthicum_trematoda_aporocotylidae_bloodFluke_contig81112	262	no	yes	no

Continued on next page

Table A.3 – *Continued from previous page*

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP105_Psettarium_anthicum_trematoda_aporocotylidae_bloodFluke_contig81113	270	no	yes	no
VAP106_Psettarium_anthicum_trematoda_aporocotylidae_bloodFluke_contig86731	177	no	yes	no
VAP107_Psettarium_anthicum_trematoda_aporocotylidae_bloodFluke_contig86733	392	no	yes	no
VAP114_Schistocephalus_solidus_cestoda_schistocephalidae_contig7559	287	no	yes	no
VAP115_Clonorchis_sinensis_trematoda_opisthorchiidae_NBF_contig8668	228	no	yes	no
VAP116_Schistocephalus_solidus_cestoda_schistocephalidae_contig10294	169	no	yes	no
VAP122_Schistosoma_haematobium_trematoda_schistosomatidae_bloodFluke_contig2295	177	no	yes	no
VAP123_Schistosoma_haematobium_trematoda_schistosomatidae_bloodFluke_contig3375	187	no	yes	no
VAP131_Schistosoma_haematobium_trematoda_schistosomatidae_bloodFluke_contig12738	190	no	yes	no
VAP138_Schistosoma_japonicum_trematoda_schistosomatidae_bloodFluke_contig6408	387	no	yes	no
VAP166_Schistosoma_mansoni_trematoda_schistosomatidae_bloodFluke_contig4969	238	no	yes	no
VAP202_Schmidtea_mediterranea_turbellaria_contig13414	228	no	yes	no
VAP230_Schmidtea_polychroa_turbellaria_contig27	430	no	yes	no
VAP238_Schmidtea_polychroa_turbellaria_contig7394	245	no	yes	no
VAP241_Schmidtea_polychroa_turbellaria_contig8803	672	no	yes	no
VAP242_Schmidtea_polychroa_turbellaria_contig11930	379	no	yes	no
VAP243_Schmidtea_polychroa_turbellaria_contig13106	270	no	yes	no
VAP244_Schmidtea_polychroa_turbellaria_contig15107	157	no	yes	no
VAP247_Schmidtea_polychroa_turbellaria_contig16244	171	no	yes	no
VAP248_Schmidtea_polychroa_turbellaria_contig16245	310	no	yes	no
VAP252_Schmidtea_polychroa_turbellaria_contig19714	342	no	yes	no
VAP254_Schmidtea_polychroa_turbellaria_contig21189	423	no	yes	no
VAP255_Schmidtea_polychroa_turbellaria_contig22145	222	no	yes	no
VAP257_Schmidtea_polychroa_turbellaria_contig40183	262	no	yes	no
VAP258_Schmidtea_polychroa_turbellaria_contig40184	265	no	yes	no
VAP259_Schmidtea_polychroa_turbellaria_contig41108	265	no	yes	no
VAP260_Selachohemecus_olsoni_trematoda_aporocotylidae_bloodFluke_contig32663	392	no	yes	no
VAP262_Selachohemecus_olsoni_trematoda_aporocotylidae_bloodFluke_contig61765	428	no	yes	no
VAP268_Spirometra_erinaceieuropaei_cestoda_diphylobothidae_contig25285	220	no	yes	no
VAP270_Spirometra_erinaceieuropaei_cestoda_diphylobothidae_contig28551	155	no	yes	no
VAP272_Spirorchis_haematobius_trematoda_spirochiidae_bloodFluke_contig9884	171	no	yes	no
VAP273_Spirorchis_haematobius_trematoda_spirochiidae_bloodFluke_contig32842	453	no	yes	no

Continued on next page

Table A.3 – *Continued from previous page*

ID	**Length**	**Group 1**	**Group 2**	**SignalP**
VAP274_Spirorchis_haematobius_trematoda_spirochiidae_bloodFluke_contig33627	178	no	yes	no
VAP275_Spirorchis_haematobius_trematoda_spirochiidae_bloodFluke_contig33628	173	no	yes	no
VAP276_Spirorchis_haematobius_trematoda_spirochiidae_bloodFluke_contig44354	217	no	yes	no
VAP278_Spirorchis_picta_trematoda_spirochiidae_bloodFluke_contig5248	453	no	yes	no
VAP283_Spirorchis_picta_trematoda_spirochiidae_bloodFluke_contig19965	173	no	yes	no
VAP284_Spirorchis_picta_trematoda_spirochiidae_bloodFluke_contig26867	217	no	yes	no
VAP286_Spirorchis_picta_trematoda_spirochiidae_bloodFluke_contig46215	413	no	yes	no
VAP288_Spirorchis_picta_trematoda_spirochiidae_bloodFluke_contig47652	261	no	yes	no
VAP291_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig26130	171	no	yes	no
VAP293_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig30923	173	no	yes	no
VAP294_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig37357	413	no	yes	no
VAP295_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig41156	157	no	yes	no
VAP296_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig44296	323	no	yes	no
VAP297_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig44302	453	no	yes	no
VAP298_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig44303	373	no	yes	no
VAP300_Spirorchis_scripta_trematoda_spirochiidae_bloodFluke_contig59318	217	no	yes	no
VAP301_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig30163	340	no	yes	no
VAP307_Taenia_asiatica_cestoda_taeniidae_contig3707	342	no	yes	no
VAP309_Taenia_asiatica_cestoda_taeniidae_contig4560	272	no	yes	no
VAP315_Taenia_saginata_cestoda_taeniidae_contig3680	669	no	yes	no
VAP317_Taenia_saginata_cestoda_taeniidae_contig3681	418	no	yes	no
VAP319_Taenia_saginata_cestoda_taeniidae_contig4891	798	no	yes	no
VAP320_Taenia_saginata_cestoda_taeniidae_contig4892	288	no	yes	no
VAP321_Taenia_saginata_cestoda_taeniidae_contig4956	191	no	yes	no
VAP322_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig30167	413	no	yes	no
VAP335_Taenia_solium_cestoda_taeniidae_contig3127	313	no	yes	no
VAP340_Taenia_solium_cestoda_taeniidae_contig6880	267	no	yes	no
VAP342_Taenia_solium_cestoda_taeniidae_contig7667	169	no	yes	no
VAP344_Taenia_solium_cestoda_taeniidae_contig10846	176	no	yes	no
VAP347_Taenia_solium_cestoda_taeniidae_contig11854	167	no	yes	no
VAP359_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig42843	258	no	yes	no
VAP364_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig6910	138	no	yes	no

Continued on next page

Table A.3 – *Continued from previous page*

ID	Length	Group 1	Group 2	SignalP
VAP368_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig48391	233	no	yes	no
VAP370_Trichobilharzia_regenti_trematoda_schistosomatidae_bloodFluke_contig11830	166	no	yes	no
VAP377_Vasotrema_spp_trematoda_spirochiidae_bloodFluke_contig76	448	no	yes	no
VAP378_Vasotrema_spp_trematoda_spirochiidae_bloodFluke_contig2164	173	no	yes	no
VAP381_Vasotrema_spp_trematoda_spirochiidae_bloodFluke_contig10596	411	no	yes	no
VAP382_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig48393	225	no	yes	no
VAP387_Vasotrema_spp_trematoda_spirochiidae_bloodFluke_contig18030	217	no	yes	no
VAP391_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig93544	232	no	yes	no
VAP392_Coeuritrema_platti_trematoda_spirochiidae_bloodFluke_contig93546	217	no	yes	no
VAP397_Dendrocoelum_lacteum_turbellaria_contig15824	319	no	yes	no
VAP398_Dendrocoelum_lacteum_turbellaria_contig16686	156	no	yes	no
VAP405_Dendrocoelum_lacteum_turbellaria_contig22337	427	no	yes	no
VAP411_Dendrocoelum_lacteum_turbellaria_contig30366	221	no	yes	no
VAP413_Dendrocoelum_lacteum_turbellaria_contig33318	230	no	yes	no
VAP414_Dendrocoelum_lacteum_turbellaria_contig36660	421	no	yes	no
VAP415_Dendrocoelum_lacteum_turbellaria_contig38403	411	no	yes	no
VAP416_Dendrocoelum_lacteum_turbellaria_contig38404	309	no	yes	no
VAP418_Dendrocoelum_lacteum_turbellaria_contig45560	260	no	yes	no
VAP419_Acipensericola_petersoni_trematoda_aporocotylidae_bloodFluke_contig76015	235	no	yes	no
VAP420_Dendrocoelum_lacteum_turbellaria_contig54631	392	no	yes	no
VAP421_Dendrocoelum_lacteum_turbellaria_contig54632	386	no	yes	no
VAP425_Acipensericola_petersoni_trematoda_aporocotylidae_bloodFluke_contig76018	240	no	yes	no
VAP428_Echinococcus_multilocularis_cestoda_taeniidae_contig2809	191	no	yes	no
VAP430_Echinococcus_multilocularis_cestoda_taeniidae_contig2834	288	no	yes	no
VAP431_Echinococcus_multilocularis_cestoda_taeniidae_contig2835	382	no	yes	no
VAP435_Echinococcus_multilocularis_cestoda_taeniidae_contig6015	167	no	yes	no
VAP436_Echinococcus_multilocularis_cestoda_taeniidae_contig6016	167	no	yes	no
VAP442_Acipensericola_petersoni_trematoda_aporocotylidae_bloodFluke_contig79159	196	no	yes	no
VAP444_Elaphrobates_euzeti_trematoda_aporocotylidae_bloodFluke_contig50115	209	no	yes	no
VAP445_Elaphrobates_euzeti_trematoda_aporocotylidae_bloodFluke_contig62175	395	no	yes	no
VAP446_Elaphrobates_euzeti_trematoda_aporocotylidae_bloodFluke_contig62176	287	no	yes	no
VAP448_Elopicola_franksi_trematoda_aporocotylidae_bloodFluke_contig36122	174	no	yes	no

Continued on next page

Table A.3 – *Continued from previous page*

ID	Length	Group 1	Group 2	SignalP
VAP449_Acipensericola_petersoni_trematoda_aporocotylidae_bloodFluke_contig87294	162	no	yes	no
VAP450_Elopicola_franksi_trematoda_aporocotylidae_bloodFluke_contig36125	393	no	yes	no
VAP452_Elopicola_franksi_trematoda_aporocotylidae_bloodFluke_contig36134	379	no	yes	no
VAP453_Elopicola_franksi_trematoda_aporocotylidae_bloodFluke_contig36135	160	no	yes	no
VAP454_Elopicola_nolancribbi_trematoda_aporocotylidae_bloodFluke_contig124690	391	no	yes	no
VAP465_Gyrodactylus_salaris_monogenea_contig1901	809	no	yes	no
VAP467_Gyrodactylus_salaris_monogenea_contig2013	324	no	yes	no
VAP477_Gyrodactylus_salaris_monogenea_contig4429	517	no	yes	no
VAP479_Gyrodactylus_salaris_monogenea_contig4732	159	no	yes	no
VAP480_Gyrodactylus_salaris_monogenea_contig7069	185	no	yes	no
VAP484_Gyrodactylus_salaris_monogenea_contig13574	221	no	yes	no
VAP485_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig6001	223	no	yes	no
VAP486_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig7673	225	no	yes	no
VAP487_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig7674	238	no	yes	no
VAP500_Cardicola_currani_trematoda_aporocotylidae_bloodFluke_contig43541	274	no	yes	no
VAP503_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig24070	336	no	yes	no
VAP504_Cardicola_currani_trematoda_aporocotylidae_bloodFluke_contig43542	266	no	yes	no
VAP505_Hapalorhynchus_foliorchis_trematoda_spirochiidae_bloodFluke_contig37935	150	no	yes	no
VAP506_Cardicola_currani_trematoda_aporocotylidae_bloodFluke_contig43543	227	no	yes	no
VAP507_Hymenolepis_diminuta_cestoda_hymenolepididae_contig551	429	no	yes	no
VAP511_Hymenolepis_diminuta_cestoda_hymenolepididae_contig4111	224	no	yes	no
VAP513_Hymenolepis_diminuta_cestoda_hymenolepididae_contig7560	164	no	yes	no
VAP514_Cardicola_currani_trematoda_aporocotylidae_bloodFluke_contig61660	160	no	yes	no
VAP521_Hymenolepis_microstoma_cestoda_hymenolepididae_contig3433	426	no	yes	no
VAP532_Hymenolepis_microstoma_cestoda_hymenolepididae_contig8881	186	no	yes	no
VAP537_Hymenolepis_microstoma_cestoda_hymenolepididae_contig10683	159	no	yes	no
VAP538_Hymenolepis_microstoma_cestoda_hymenolepididae_contig11502	185	no	yes	no
VAP540_Littorellicola_billhawkinsi_trematoda_aporocotylidae_bloodFluke_contig9483	318	no	yes	no
VAP542_Littorellicola_billhawkinsi_trematoda_aporocotylidae_bloodFluke_contig13412	192	no	yes	no
VAP545_Littorellicola_billhawkinsi_trematoda_aporocotylidae_bloodFluke_contig22849	192	no	yes	no
VAP550_Macrostomum_lignano_turbellaria_contig11701	960	no	yes	no
VAP551_Macrostomum_lignano_turbellaria_contig12528	237	no	yes	no

Continued on next page

Table A.3 – *Continued from previous page*

ID	Length	Group 1	Group 2	SignalP
VAP552_Macrostomum_lignano_turbellaria_contig13191	239	no	yes	no
VAP554_Macrostomum_lignano_turbellaria_contig16708	482	no	yes	no
VAP557_Macrostomum_lignano_turbellaria_contig19908	237	no	yes	no
VAP564_Macrostomum_lignano_turbellaria_contig28599	149	no	yes	no
VAP565_Macrostomum_lignano_turbellaria_contig29800	180	no	yes	no
VAP566_Macrostomum_lignano_turbellaria_contig32042	402	no	yes	no
VAP571_Macrostomum_lignano_turbellaria_contig36071	260	no	yes	no
VAP572_Macrostomum_lignano_turbellaria_contig36322	187	no	yes	no
VAP573_Cardicola_palmeri_trematoda_aporocotylidae_bloodFluke_contig40889	260	no	yes	no
VAP575_Macrostomum_lignano_turbellaria_contig50500	941	no	yes	no
VAP576_Macrostomum_lignano_turbellaria_contig51744	320	no	yes	no
VAP577_Mesocestoides_corti_cestoda_mesocestoididae_contig132	230	no	yes	no
VAP578_Mesocestoides_corti_cestoda_mesocestoididae_contig1312	452	no	yes	no
VAP603_Microphallus_livelyi_trematoda_microphallidae_cercaria_NBF_contig1974	181	no	yes	no
VAP604_Microphallus_livelyi_trematoda_microphallidae_cercaria_NBF_contig2557	228	no	yes	no
VAP605_Cardicola_palmeri_trematoda_aporocotylidae_bloodFluke_contig40895	268	no	yes	no
VAP606_Microphallus_livelyi_trematoda_microphallidae_cercaria_NBF_contig3125	165	no	yes	no
VAP608_Microphallus_livelyi_trematoda_microphallidae_cercaria_NBF_contig10229	413	no	yes	no
VAP610_Microphallus_livelyi_trematoda_microphallidae_cercaria_NBF_contig13997	125	no	yes	no
VAP612_Myliobaticola_nsp_trematoda_aporocotylidae_bloodFluke_contig9884	174	no	yes	no
VAP615_Myliobaticola_nsp_trematoda_aporocotylidae_bloodFluke_contig15400	186	no	yes	no
VAP616_Myliobaticola_nsp_trematoda_aporocotylidae_bloodFluke_contig15401	506	no	yes	no
VAP618_Myliobaticola_nsp_trematoda_aporocotylidae_bloodFluke_contig17055	428	no	yes	no
VAP622_nGen_nsp_trematoda_aporocotylidae_bloodFluke_contig27205	162	no	yes	no
VAP443_Elaphrobates_euzeti_trematoda_aporocotylidae_bloodFluke_contig50113	210	no	yes	yes
VAP456_Fasciola_hepatica_trematoda_echinostomatidae_NBF_contig4594	157	no	yes	yes

Appendix B

Files

All files of interest to this manuscript, including all those referenced in Supplementary Table A.2, are publicly available at `https://github.com/Sipley/MS-thesis`. The transcriptomes are too large to store on GitHub, but instructions for accessing them are available in the `README.md`. If the link does not work for you, this manuscript is not published yet, but if you email me at `Breanna.Sipley@gmail.com` with your GitHub Username (if applicable), I would be happy to discuss granting you access to the private repository.

Appendix C

Code

All code of interest to this manuscript, including step-by-step instructions for running through the bioinformatics pipeline employed and figures generated in this study, are publicly available at `https://github.com/Sipley/MS-thesis`. If the link does not work for you, this manuscript is not published yet, but if you email me at `Breanna.Sipley@gmail.com` with your GitHub Username (if applicable), I would be happy to discuss granting you access to the private repository.