2017

Platyhelminth drug targets: Identification, annotation, and validation

Nicolas James Wheeler
Iowa State University

Platyhelminth drug targets: Identification, annotation, and validation

by

Nicolas James Wheeler

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Genetics

Program of Study Committee:
Tim A. Day, Major Professor
Michael Kimber
Mary Heather West Greenlee
Steve Carlson
Steven Whitham

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

DEDICATION

I would like to dedicate this dissertation to my wife Deanna.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I give my sincere thanks to several people who have contributed to this work and my development as a scientist. First and foremost, I thank Dr. Tim Day, who recruited me to ISU and created a platform by which I could freely explore my interests within the realm of neglected tropical diseases. Second, I am thankful to my committee members, in particular Dr. Michael Kimber, who helped my development as both a scientist and as a teacher. I am especially thankful to Dr. Mostafa Zamanian, whose novel work in Dr. Day's lab prepared the way for my advancement, and who has remained a friend and colleague for many years. Finally, I thank my fellow students in Drs. Day and Kimber's laboratories - Sreemoyee Acharya and Hiruni Hettiarachchi. They are much better students than I, and they often made up for my shortcomings. Finally, I would like to thank my reliable editors and friends, Travis and Erin.

ABSTRACT

The following dissertation describes work performed to probe the biology of flatworms, with a focus on anthelmintic development and target identification and validation. A wide range of molecular and computational approaches were used to achieve this goal. In particular, I show how free-living planaria can be utilized as a model for its parasitic relatives, for answering biological questions and for anthelmintic screening techniques. Through a variety of approaches, I also contribute to the ongoing study and annotation of a large group of flatworm-specific G protein-coupled receptors, the Platyhelminth-Specific Rhodopsin-Like Orphan Family. This family is specific to the flatworms, is the largest clade of GPCRs and the largest taxonomically-restricted gene family in the entire phylum, and a PROF representative may be preferentially expressed in the neural tissue of planaria. Finally, using the PROF as a case study, I comment on the task of functionally annotating GPCRs from parasitic worms and suggest a more wholistic and rigorous approach. The entirety of this dissertation is then discussed, and the results are reinterpreted through a lens that focuses on anthelmintic discovery and development.

CHAPTER 1. OVERVIEW

This dissertation describes work performed to enhance the understanding of a large phylum of invertebrates that houses several parasites of veterinary and public health importance, Platyhelminthes. Particularly, it is basic research organized around two separate paths for drug discovery: 1) utilizing free-living planaria as models in place of parasites and 2) focusing on popular putative drug targets for further annotation, characterization, and validation.

This first chapter gives background on the organisms of interest, the diseases they cause, and the current knowledge of drugs and drug targets in platyhelminths, also known as flatworms. It also provides a general discussion on the state of drug development for parasitic platyhelminth infections and describes four conceptual paths to anthelmintic development. Following are three manuscripts that were produced through the course of the work for this dissertation. Finally, these manuscripts will be connected and expanded upon in a concluding chapter, and a suggestion of the next logical steps to take is included. Several appendices describe other related contributions to published and not-yet-published work that was performed outside of the scope of this dissertation.

1.1 Parasitic platyhelminths and the diseases they cause

Of the seventeen Neglected Tropical Diseases (NTDs) prioritized by the World Health Organization (WHO), four (schistosomiasis, foodborne trematodiases, taeniasis/cysticercosis, and echinococcosis) are caused by taxa of the Platyhelminthes. Together, these diseases cause significant morbidity that severely inhibits the advancement and flourishing of developing countries, particularly in sub-Saharan Africa (Table 1.1).

Table 1.1: Neglected platyhelminthiases and their associated morbidity. Morbidity of platyhelminthiases measured in disability-adjusted life years (DALYs), or healthy years lost due to disability or early death.

Disease	Genus	Class	DALYs (thousands)
Schistosomiasis	*Schistosoma*	Trematoda	3062.8
Foodborne Trematodiases	*Clonorchis*, *Fasciola*, *Opisthorchis*, *Paragonimus*	Trematoda	3634.8
Taeniasis/Cysticercosis	*Taenia*	Cestoda	341.2
Echinococcosis	*Echinococcus*	Cestoda	181.7

Parasitic flatworms reside in a small subphylum of Platyhelminthes called Neodermata (Figure 1.1). These worms are characterized by their complex life cycles that traverse through an intermediate host, either a freshwater invertebrate such as a snail or crustacean for the Trematoda (*Schistosoma, Clonorchis, Fasciola, Paragonimus, Opisthorchis*) or an ungulate for the Cestoda (*Taenia, Echinococcus*). The worms come to reside in humans by burrowing through the skin or being ingested in food. Once parasites reach adulthood inside their definitive host, the worms lay eggs that are excreted from the host through a variety of routes depending on the species. The hatched eggs yield forms infective to the intermediate host, and the life cycle is restarted.

Platyhelminth NTDs are targeted by the WHO for prevention and control in endemic areas by so-called preventive chemotherapy (PC). As a public health concern, instead of treating sporadic cases of platyhelminth NTDs after diagnosis, public institutions organize community and regional-level administration of anthelmintics to at-risk individuals. PC helps to prevent individual morbidity, but it also slows transmission of the infections by killing the worms before they reach fecundity.

Mass drug administration (MDA) and community level PC have been incredibly effective [222]. However, widespread use of only a few drugs creates fear of the eventual selection of resistant strains. Anthelmintic resistance is a serious problem in veterinary medicine (see [149] for instance), strengthening fears that it will also become a public health concern. For this reason, increased research of anthelmintic resistance, alternative strategies for platyhelminthiasis control, new chemotherapeutics, and the basic biology of parasitic platyhelminths is necessary.

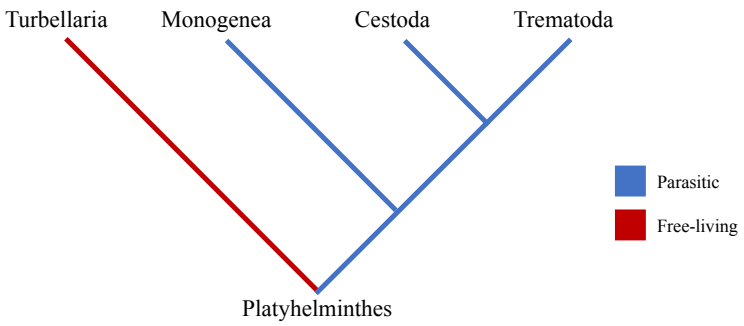

Figure 1.1: Phylogeny of Platyhelminthes. The parasitic Neodermata (blue) and the free-living Turbellaria, which includes planaria (red). Monogenea are ectopic parasites that do not infect humans or animals of veterinary importance.

1.1.1 *Schistosoma* and schistosomiasis

Schistosomiasis is the most widespread and morbid of the platyhelminthiases. It is conservatively estimated that 230 million people are currently infected, and it is endemic in much of sub-Saharan Africa, which experiences 93% of the global disease burden [201, 84]. Schistosomiasis is caused by infection by flatworms of *Schistosoma* spp. Human infections primarily result from *S. mansoni, S. japonicum*, and *S. haematobium*. Disease etiology and pathophysiology differ by species. Urogenital schistosomiasis is caused by *S. haematobium* and is associated with increased risk of bladder cancer and fibrosis of the bladder and kidney. Hepatic/intestinal schistosomiasis is caused by *S. mansoni* and *S. japonicum* and is associated cirrhosis of the liver. Unfortunately, symptoms are particularly severe in children, especially with repeated infection, and they can result in growth stunting and reduced educational attainment [223].

The morbidity associated with both types of schistosomiasis is not primarily caused by the adult flukes themselves, but instead by the chronic deposition of eggs by the adult female worms. The genus *Schistosoma* is unique to the rest of the Neodermata in that it has evolved sex differentiation (Figure 1.1). Upon maturity to fecundity, male and female schistosomes pair in the mesenteric vasculature and can remain paired for the rest of their lives. Paired *S. mansoni* living in the blood stream can lay up to 300 eggs per day [74]. These eggs circulate throughout the vasculature until they pass through the intestinal wall into the gut lumen and

are excreted, or until they become lodged in the host tissue. Eggs embedded in tissue cause chronic inflammation, resulting the pathologies associated with the disease.

As mentioned, schistosomes have complex life cycles that go through an aquatic gastropod mollusc intermediate host. After eggs are excreted in the feces or urine of the definitive host, they hatch in fresh water as miracidia. Miracidia penetrate the snail host and develop to sporocysts, where they clonally reproduce from mother sporocysts to daughters. Infective cercariae are later released from snails back into the water, where they can find a definitive human host (though *S. japonicum* can reside in other mammals as well) and burrow through the skin into the vasculature. After entering, cercariae shed their tails and develop to juvenile schistosomulae in the host's blood. Schistosomulae circulate through the blood to the portal blood of the liver, where they become sexually mature and form male/female pairs.

1.1.2 Others

The four foodborne trematodiases are caused by the consumption of food that harbors the eggs of parasitic trematodes, typically raw fish, crustaceans, or vegetables (Table 1.1). These diseases are indirect zoonoses do not depend on infection humans to complete the life cycle, in contrast to schistosomes, for which humans are the most suitable definitive host. Instead, foodborne trematodes cause pathology in humans when they are inadvertently eaten. Generally, these diseases are asymptomatic but become severe when worm load becomes high in the organs in which they live. For clonorchiasis and opisthorchiasis, adults worms live in the smaller bile ducts of the liver where they can cause inflammation. Like *S. haematobium, Chlonorchis sinensis and Opisthorchis viverrini* are carcinogenic and can cause mortality through bile duct cancer. *Fasciola* spp. also reside in the liver, but these prefer the larger bile ducts and gall bladder. Though noncarcinogenic, *Fasciola* can cause similar inflammation and fibrosis to that seen by the other liver flukes.

Paragonimus spp. are classified as lung flukes, according to their preferred organ of habitation. Clinical manifestations of lung fluke are diverse and can be mistaken for tuberculosis, which confounds diagnosis because tuberculosis is often coendemic with paragonimiasis [113]. Worms can also travel ectopically, and symptoms become particularly dangerous when the

flukes make it to the brain. Because of their residence in the lungs, lung fluke eggs are typically expelled through the sputum rather than the urine or feces, though they are occasionally found there as well [101].

Finally, zoonotic tapeworms of the genuses *Taenia* and *Echinococcus* are also foodborne, but tapeworms are from the class Cestoda rather than the related Trematoda. These can be acquired by consuming raw or undercooked beef or pork in the case of *Taenia* spp., or by the inadvertent consumption of *Echinococcus* eggs that have been excreted from dogs, foxes, and other Canidae, depending on the species. *Taenia* spp. develop to adults in the human small intestine, where they can grow up to several meters in length. Though often asymptomatic, larger tapeworms can cause digestive and stomach problems, but the most dangerous manifestation of the disease is caused by the distribution of the larval cysts throughout the body. In muscle, these cysts can cause pain and swelling, but cysts can also travel to the brain where they can lead to adult-onset seizures. Similarly, *Echinococcus granulosas* also forms larval cysts that distribute throughout the body. These can burst, releasing cystic fluid and causing severe inflammatory reactions. In contrast, zoonotic *Echinococcus multilocularis* cannot fully mature to cysts in humans and can asymptomatically live in humans for more than a decade. However, as the immature cysts grow, they becomes tumour-like and can metastasize to other organs and tissues, becoming fatal if left untreated.

1.2 Current drugs and their targets

Each platyhelminthiasis described above has an efficacious chemotherapeutic to treat it. Indeed, schistosomiasis, chlonorchiasis, and opisthorciasis have been targeted by MDA strategies [164]. Other foodborne trematodiases and tapeworm diseases, on the other hand, are treated after clinical diagnosis.

1.2.1 Praziquantel

Praziquantel (PZQ) is an efficacious small molecule that resolves infections caused by trematode adults and is used to combat schistosomiasis and foodborne trematodiases. Because it results in the death egg-laying adults, PZQ is especially amenable for MDA programs by ab-

lating the morbidity of the disease while also decreasing the rate of transmission through the excretion of trematode eggs. PZQ was initially developed through a collaboration between the pharmaceutical companies Bayer and Merck, and PZQ MDA programs progress through strategic public-private partnerships, where industry partners provide the drug while public bureaucracies oversee the administration.

PZQ was first developed for veterinary use in the late 1970s and was subsequently converted for use as an anthelmintic for humans [170]. PZQ was discovered and characterized through screening of whole animal infection models [11, 93]. Though decades old, the direct target for PZQ is still unknown. Disruption of calcium homeostasis, and the resulting membrane blebbing, and tegumental damage that the disruption creates, has long been known to be a significant result of treatment [12], but the target of the mechanism of action is elusive. Evidence from both schistosomes and free-living planaria indicate that voltage-operated calcium channels play a significant role in PZQ's effects [104, 144], but recent evidence also suggests that aminergic G protein-coupled receptors might also be important [37].

Despite these unknowns, PZQ continues to be remarkably efficacious today, and tens of millions are treated with it every year. The WHO's most recent data indicate that over 53 million people were treated with PZQ in 2015, resulting in 28% coverage for those requiring PC for schistosomiasis [227]. New models have suggested that continuing to increase coverage (and expanding it to treatment of adults instead of just school-aged children) could result in disease elimination [126]. Thus, the WHO has amended their MDA guidelines [228]. However, as PZQ is the sole anti-schistosomal used for these campaigns and the treatment of choice for several foodborne trematodiases, the selection for resistance is a lingering fear. Indeed, resistance can be reproducibly selected for in the lab [60], and lower than expected PZQ efficacy has been repeatedly reported in puncta throughout Africa [91, 48].

1.2.2 Oxamniquine

Oxamniquine (OXAM) was developed by Pfizer in the 1960s, before the arise of PZQ [173] and its eventual usurpation of OXAM. OXAM kills both juvenile and adult worms, but it is only efficacious against *S. mansoni*. For this reason, it was leveraged for MDA in Brazil until

2010, when OXAM resistant had become widespread [212]. Though resistance contributed to the discontinuation of OXAM, OXAM had also been quickly superseded by PZQ because of its higher efficacy, its ability to treat all types of schistosomiasis, and its cheaper synthesis and administration.

OXAM resistance has been a useful case study for how anthelmintic resistance arises, and it stimulated prediction that PZQ resistance may one day become a problem. It was reported early that OXAM was metabolized by a *S. mansoni* sulfotransferase [163], and that the metabolite inhibited DNA and RNA synthesis by associating with worm DNA [162]. Thus, it was speculated that mutations to the sulfotransferase was the cause of resistance, and this was recently confirmed [216].

1.2.3 Benzimidazoles

Two benzimadazoles (BZs) are used for the treatment of platyhelminthiases: triclabendazole (TCBZ) for liver fluke and albendazole (ALB) for tapeworm infections [101]. Like other anthelmintics, BZs were first developed for use in veterinary medicine, but these were discovered in *in vitro* phenotypic screens rather than screens of infected animals [32]. TCBZ was first used in veterinary practice in 1983 and in human medicine in 1989, after a liver fluke outbreak in Iran [231]. ALB followed a similar trajectory: developed in the UK and first marketed in 1977 as a veterinary anthelmintic, and then approved for human use 10 years later [151].

BZs all target β-tubulin, and their use is more prevalent against nematode parasites than flatworm parasites. Nevertheless, its mechanism of action remains the same for all species for which it is utilized. By blocking the polymerization of tubulin, worms are unable to uptake glucose and synthesize ATP, causing slow death. Like in nematodes, benzimidazole treatment failure and resistance is widely reported in liver flukes [210, 29], but not yet in tapeworms [56].

1.2.4 The state of drug development for platyhelminthiases

The newest commercially available anthelmintic against platyhelminthiases is now almost 15 years old, and it is only used in Egypt and lowly efficacious [212, 232]. There are currently no antischistosomals in clinical trials [212, 193], and there are none in the pipeline for any of

the flatworm infections. Some have argued that, at least for schistosomiasis, new drugs might not even be necessary, and instead focus should be given to strengthening public-private partnerships and improving MDA infrastructure [148, 193]. Furthermore, even if new drugs were developed, it seems unlikely that they would provide more utility than PZQ, as was recently concluded for two anti-malarials that were shown to have efficacy against *S. haematobium* [100]. Nevertheless, the fear of PZQ resistance and its lack of efficacy against juvenile worms drives continued drug discovery efforts.

There are four main trajectories for anthelmintic development (Table 1.2). As noted, most anthelmintics were discovered and developed through screens against infected animals models like mice, rabbits, hamsters, or pigs. While the benzimidazoles arose from screens against parasites *in vitro*, they are one success story among tens of millions of failures [66]. Currently, for both chemical and reverse genetic screens, there are only three significant phenotypes available for most medium-throughput measurement - motility, death, and morphology. There is hope that additional and more sensitive phenotypic assays will increase the discovery rate, but this is yet to be realized. In contrast, free-living flatworms have a more diverse range of phenotypic assays, and some platyhelminth parasitologists hope to establish free-living planaria as a reliable screening model for platyhelminth parasites (see below). However, even use of the robust and pliable *C. elegans* has failed to elucidate any new nematicides [66], and it is unclear if planaria are a better model for parasitic flatworms than *C. elegans* is for roundworms. Finally, it was at one time prophesied that the genomic age would enable mechanistic-based drug development in which putative targets, identified through basic research or *in silico* prioritization, would be heterologously expressed and screened [67]. Unfortunately, this also has not come to fruition, though it is still an active area of pursuit [38].

1.3 Anthelmintic discovery and development in an academic setting

While screening chemical libraries against infected model organisms has been the most productive way to develop anthelmintics, the cost of these approaches is very limiting. Thus, academic researchers who are interested in contributing to anthelmintic discovery and development are compelled to look to alternative strategies (Table 1.2). These may not be as suited to

Table 1.2: Trajectories for anthelmintic discovery.

Trajectory	Technique	Relative Cost	Throughput	Proximity to Translation	Positive Results
1. Infected model organism	Chemical screen	$$$$	*	****	Lead compound
2. Parasite *in vitro*	Chemical screen Genetic screen	$$$	*	***	Lead compound Biological conclusions
3. Planaria *in vivo*	Chemical screen Genetic screen	$$	***	**	Lead compound Biological conclusions
4. Drug target heterologous expression	Chemical screen	$	****	**	Lead compound

provide leads that are easily translated to medicine, but they do provide worthwhile byproducts that may not be directly related to anthelmintic development.

For example, while screens of parasites or free-living models may not result in a lead compound, initial hits can be directed toward a track that is more interested in the biology of the organism and the biological mechanism of the chemical itself. Likewise, reverse genetic screens are useful for functional annotation of genes, because they are more interested in the target than the chemical. This is to say, misses in screens against parasites or free-living models is not necessarily a complete waste of time and money; valuable data can be gleaned from these experiments. In practice, the same cannot be said for the mechanistic-based screens in heterologous expression systems. These may provide leads, but they are much less likely to provide important data about the biology of the worm. However, these screens are typically the ultimate experiment at the end a long pipeline of target identification, annotation, and validation, all of which are worthy pursuits in and of themselves.

Thus, in an academic setting, researchers have the choice of which approach to take based on the general aims of the laboratory, as well as the availability of a labor force and specialized equipment or knowledge. This dissertation focuses on contributions that were made for two of the four approaches: the development of free-living models for *in vitro screens* and the construction of pipelines that prioritize putative drug targets. In addition to discussing the translational capacity for these contributions, focus is also given to the biological conclusions that resulted from this work.

1.3.1 Platyhelminth models for *in vitro* screens

Though *C. elegans* as a screening model for nematicides has yet to cultivate any new nematicides, there remains momentum to continue its use [26]. In addition to screening, both classical and innovative comparative studies between *C. elegans* and parasitic nematodes pervade the recent literature, and this pattern will continue [194, 161, 235].

In contrast to the nematodes, the use of free-living flatworms as models for their parasitic phyla-mates is relatively recent [44, 39, 37, 36]. This is partly historical - *C. elegans* has been a popular eukaryotic model for decades, and it was the first multicellular organism to have its genome solved [28]. Thus, its utilization as a model for parasites was only natural. Planaria, on the other hand, are not nearly as popular, and the first planarian draft genome of *Schmidtea mediterranea* was not published until much later [176]. Even then, the genome was so fragmented that it was onerous to use, and it required frequent updating [178]. Additional transcriptomes were added to aid gene discovery and annotation [115, 172, 6, 3, 20], but there was disagreement on the best species of planaria to use [40, 155]. Although *S. mediterranea* was the first free-living flatworm to have a sequenced genome and database of ESTs, they were in many ways unusable. Further, *S. mediterranea* is somewhat difficult to culture in the lab and not commercially available.

Compared to parasitic platyhelminths, planaria are easier to culture in the lab and are amenable to reverse genetics and other molecular tools. Furthermore, because of their rich history in the fields of regeneration and stem cell biology, there is a greater breadth of phenotypes available for assay. The recent discovery of neoblasts, the totipotent stem cells that give planaria their unsurpassed regenerative capacity, in schistosomes [45, 219] and tapeworms [24] has served to further strengthen the idea of planaria as a model for parasites. Toxicological [76] and reverse genetic screens [179, 221] of planaria have been reported, but whether or not this model can generate to new anthelmintic leads remains to be seen.

1.3.2 Approaches for target prioritization and validation

In addition to these planarian resources, there are now dozens of platyhelminth parasite genomes available for comparative analyses [214, 165, 86]. Comparative and subtractive genomics will be important techniques to identify attractive drug targets in these pathogens and prioritize those that should be screened through heterologous expression (Table 1.2) [207, 131]. Crucial to this endeavor is the precise prediction and annotation of the millions of genes included in these datasets. Toward this end, automated approaches can perform *ab initio* and extrinsic gene finding and annotation [19, 34, 80]. Other more directed approaches give particular attention to superfamilies of genes that are conceptualized as attractive drug targets. For example, most parasitic flatworm genome publications include glosses of G protein-coupled receptors (GPCRs), ion channels, kinases, and others [214, 18].

GPCRs in particular are often highlighted not only because of their therapeutic potential, but because they are readily identifiable in well assembled genomes. Open-reading frames can be scanned *in silico*, and hydrophobicity plots can be constructed for each predicted gene in the genome. Because of their canonical and invariable 7 transmembrane (TM) domain structure, GPCRs can be unbiasedly annotated with few false positives (there are non-GPCR proteins with a flipped 7TM topology, but so far these seem to be constrained to arthropods [16]). This approach has been used in several GPCR identification projects for worms throughout the phylum, including *S. mediterranea, S. mansoni, S. haematobium*, and four different tapeworms [214, 236, 33]. Once the entire GPCR complement is identified, downstream analyses can engage in more precise functional annotation, phylogenetic inference, and, eventually, experimentation and ligand validation.

The power of comparative genomics for GPCR identification, annotation, and prioritization is increased with each new flatworm representative. With enough data from across the phylum, focused analysis of specific gene families can reveal interesting phenomena, such as:

- the contraction or expansion of subfamilies,
- the proliferation of taxonomically-restricted and lineage-specific subfamilies,

- correlations between binary presence/absence of subfamilies with organismal life history traits, and
- divergence between host and parasite GPCR complements.

Broad comparisons such as these serve to provide the foundation for hypothesis-driven research, and they are valuable datasets for a community interested in GPCRs as putative drug targets. A phylum-wide briefing of the GPCR complements of each flatworm with a sequenced genome would be a welcome addition to the conglomerate of data at hand for validation as drug targets.

1.3.3 A GPCR target case study

Until recently, the genomes of *S. mansoni* and *S. mediterranea* were the only platyhelminth genomes available. Comparisons of these two datasets identified several interesting gene families for further study, but, with only two flatworm genomes available, many of the inferences listed above could only be made with heavy caveats. From these preliminary analyses, the most compelling subfamily of GPCRs was the Platyhelminth-Specific Rhodopsin-Like Orphan Family (PROF) [236]. This subfamily was the largest GPCR clade in both flatworms, and its members showed no sequence similarity to any other protein in the entire metazoan database. PROFs retain the canonical features of GPCRs - 7 TM domains, predicted intracellular C-terminus and extracellular N-terminus, and the E/DRY motif on TM3), However, at this time, several questions remained, including:

- the extent to which the PROF is found in the rest of the phylum,
- the veracity of the family's lineage-specific expansion,
- the biological functions of the PROF members, and
- the "druggability" of the PROF receptors [153].

At the time, many of these questions could not be answered, but the recent expansion in flatworm genomic data allows for a renewed look at this compelling family.

1.4 Dissertation Organization

The following chapters describe work done to increase the applicability of the third and fourth anthelmintic discovery methods described in Table 1.2. Specifically, chapter 2 describes the sequencing of the transcriptome of *Dugesia tigrina*, a planaria that is proposed as a model for parasitic platyhelminths. Chapter 3 describes comparative work that sought to further describe and annotate the PROFs and conjecture upon their druggability. Chapter 4 is an opinion piece that discusses the importance of accuracy and precision in our gene annotation protocols, and is uses the history of the PROF annotation as a case study to demonstrate where annotation efforts can go wrong and mislead. Finally, chapter 5 discusses how chapters 2-4 fit into the general framework of anthelmintic discovery, and it provides reasonable next steps for how this work can be extended.

CHAPTER 2. FUNCTIONAL ANALYSIS OF DUGESIA TIGRINA TRANSCRIPTOME SEEDS PIPELINE FOR ANTHELMINTIC TARGET DISCOVERY

A paper published in the Journal *Parasites and Vectors*

Nicolas J Wheeler[1,2,3], Prince N Agbedanu[1], Michael J Kimber[1], Paula A Ribeiro[4], Tim A Day[1,2], Mostafa Zamanian[4,5]

2.1 Abstract

Neglected diseases caused by infection by parasitic flatworms (platyhelminthes) impose a massive hindrance to the continued progress of the developing world. While basic research on flatworm parasites has increased over the past decade, researchers have yet to broadly adopt or establish a free-living model as a complementary approach to studying these parasites. Planaria are free-living flatworms best known for their robust regenerative capacities and have recently attracted greater attention as model organisms for their parasitic relatives. Here, we report the high-coverage sequencing and *de novo* assembly of the transcriptome of *Dugesia tigrina*, a planarian widely utilized in the laboratory. The transcriptome of this model organism was annotated and used to seed a pipeline for the rational prioritization and validation of putative drug targets. Following extensive orthology analysis, a small number of targets conserved between parasitic and free-living flatworms were comparatively interrogated. As proof of principle, it is shown how RNAi and pharmacology in the more convenient planarian model system

[1]Department of Biomedical Sciences, Iowa State University, Ames, IA USA
[2]Interdepartmental Genetics and Genomics, Iowa State University, Ames, IA USA
[3]Primary researcher and author
[4]Institute of Parasitology, McGill University, Ste. Anne de Bellevue, Quebec, Canada
[5]Author for correspondence

can both inform parasite biology and serve as an efficient screening tool for the identification of lucrative anthelmintic targets.

2.2 Author Summary

The phylum Platyhelminthes includes a number of important human pathogens that inflict a devastating global health burden. Free-living members of this phylum are best known for their robust regenerative capacities and have more recently been forwarded as model organisms for their parasitic relatives. To further this paradigm, a *de novo* transcriptome was assembled for the planarian *Dugesia tigrina* across a set of dynamic conditions, followed by annotation and extensive orthology analysis with respect to trematode (blood fluke) and cestode (tapeworm) sequence data. A drug target prioritization scheme was implemented, where the investigation of select targets in planaria can act as a convenient and tractable first-pass screening platform for evaluating the druggability of corresponding parasite homologs. As proof of principle, three planarian-parasite ortholog groups were probed with a mixture of RNAi and pharmacology experiments. The results illustrate that this high quality *D. tigrina* transcriptome can serve as a beneficial substrate for anthelmintic target identification, screening and prioritization.

2.3 Introduction

The Platyhelminthes (flatworms) comprise a diverse phylum of medically and economically important species. Trematodes (flukes) and cestodes (tapeworms) are the etiological agents of several Neglected Tropical Diseases (NTDs) that disproportionately devastate the health and economic prospects of the poor across much of the developing world. Schistosomes infect over 220 million in sub-Saharan Africa alone, and 600-800 million live at risk of infection worldwide [226]. Echinococcosis and cysticercosis, while less prevalent than schistosomiasis, are zoonotic parasitic diseases of great public health importance. These neglected diseases inflict significant morbidity and mortality, accounting for upwards of 280,000 deaths and an annual loss of between 3.5 - 70 million disability-adjusted life years (DALYs) [217, 140].

The prioritization of flatworm-associated NTDs by the World Health Organization underscores the urgency of efforts to control infection and to develop new anthelmintic treatments [211]. The threat of drug resistance further calls attention to the need for novel pipelines for drug target validation and drug discovery [220, 69]. Against this backdrop, free-living flatworms represent a new and potentially powerful screening model for parasite drug discovery efforts [41]. Planaria of the free-living Turbellaria are widely interrogated in the realm of stem cell biology due to their remarkable regenerative abilities [192]. In comparison to their parasitic counterparts, planarians are much more amenable to modern genetic protocols and their culture and maintenance within the laboratory is relatively cheap and simple. Many behavioral, biochemical, and morphological phenotypes have also been described for planaria, enabling straightforward inferences of function from combined reverse genetic, pharmacological, and phenotypic analyses [238, 59, 169].

In the case of schistosomes, it is necessary to maintain active populations of freshwater snails as intermediate hosts, manage periodic shedding of the infective cercariae, induce transformation to schistosomula or allow for penetrance into a definitive host (usually mice). The process is difficult, time consuming, moderately dangerous, and, for many labs, cost prohibitive. These concerns underpin efforts to extend the utility of planarian biology to the study of nearly-related parasites [43], mirroring the important role that *Caenorhabditis elegans* has played in furthering our understanding of the biology of parasitic nematodes [70].

A number of planarian species see use in the laboratory, with varying modes of reproduction, regenerative potential, and genome ploidy. *Schmidtea mediterranea* is among the most widely studied species. Clonal lines of *S. mediterranea* have been propagated to mitigate genetic heterogeneity, and both genomic and transcriptomic data have been published for this stable diploid [177, 3, 64, 172, 199]. Other notable planarian species include *Dugesia tigrina* and *Dugesia japonica*[143]. Genome assembly and analyses are partly complicated in these species due to their mixoploid genomes and the presence of large numbers of transposable elements [65]. No significant sequence resources yet exist for *D. tigrina*, despite the convenient commercial availability of this species and its broad adoption in the fields of regeneration, pharmacology and learning and memory [107, 154, 168].

The emergence of a comprehensive sequence resource for *D. tigrina* will open avenues for more precise biological manipulation of these planaria. RNA-Seq provides a powerful platform for producing a high-coverage transcriptome, without the complications of whole genome assembly. The selection of *D. tigrina* for this undertaking presents a reasonable trade-off, whereby some level of genetic heterogeneity is accepted for the greater ease of procuring, maintaining and scaling colonies, in comparison to clonally-derived laboratory strains. This is ultimately beneficial in the scope of our aims and isn't of consequence to the viability of the planarian-parasite model paradigm. Genetic variation within this sexual strain is minor with respect to the accepted genetic distance between planaria and the flatworm parasites for which they are to serve as models. Although computationally intensive in the absence of a reference genome, a high-quality *de novo* transcriptome assembly would allow for closer examination of our overarching hypothesis that *D. tigrina* could provide a shortcut to identifying potential drug targets in the phylum.

2.4 Results and Discussion

2.4.1 *De novo* transcriptome assembly

To help improve the odds of capturing transcripts in a more comprehensive manner, RNA was isolated from *D. tigrina* across a set of dynamic conditions. Planaria were passaged through a feed-starve cycle under different conditions prior to RNA extraction (Figure 2.1A). Worms were left untreated, cut transversely, and cut tranversely while incubated in the presence of the biogenic amine serotonin (5-hydroxytryptamine: 5HT). The aim of bisecting planaria was to illicit activation of potentially dormant regeneration-associated transcripts [98]. Serotonin was included due to its abundance and wide distribution in flatworm nervous systems, as well as the fundamental role it plays in parasite neuromuscular signaling [158].

Total RNA was extracted and assessed for quality in preparation for Illumina paired-end (2x100 bp) RNA-Seq. Read sets were combined for adapter-trimming, quality control and *de novo* assembly using two independent pipelines. Trinity [72] was used alongside a multiple k-mer (k = 21, 25, 29 and 33) Velvet/Oases [237], [191] pipeline to produce initial assemblies, as

depicted in Figure 2.1B. The Trinity assembly was filtered at a low abundance threshold after transcript abundance estimation vis RSEM [121]. The final Trinity (Assembly-T) and Velvet (Assembly-V) assemblies exhibit similar statistical profiles, with a comparable total transcript count, mean transcript length, N50, and transcript length range (Figure 2.1B). Both assemblies compare very favorably to other published planarian assemblies, due in part to the large read count and the computationally expensive incorporation of all available reads [3, 64, 143].

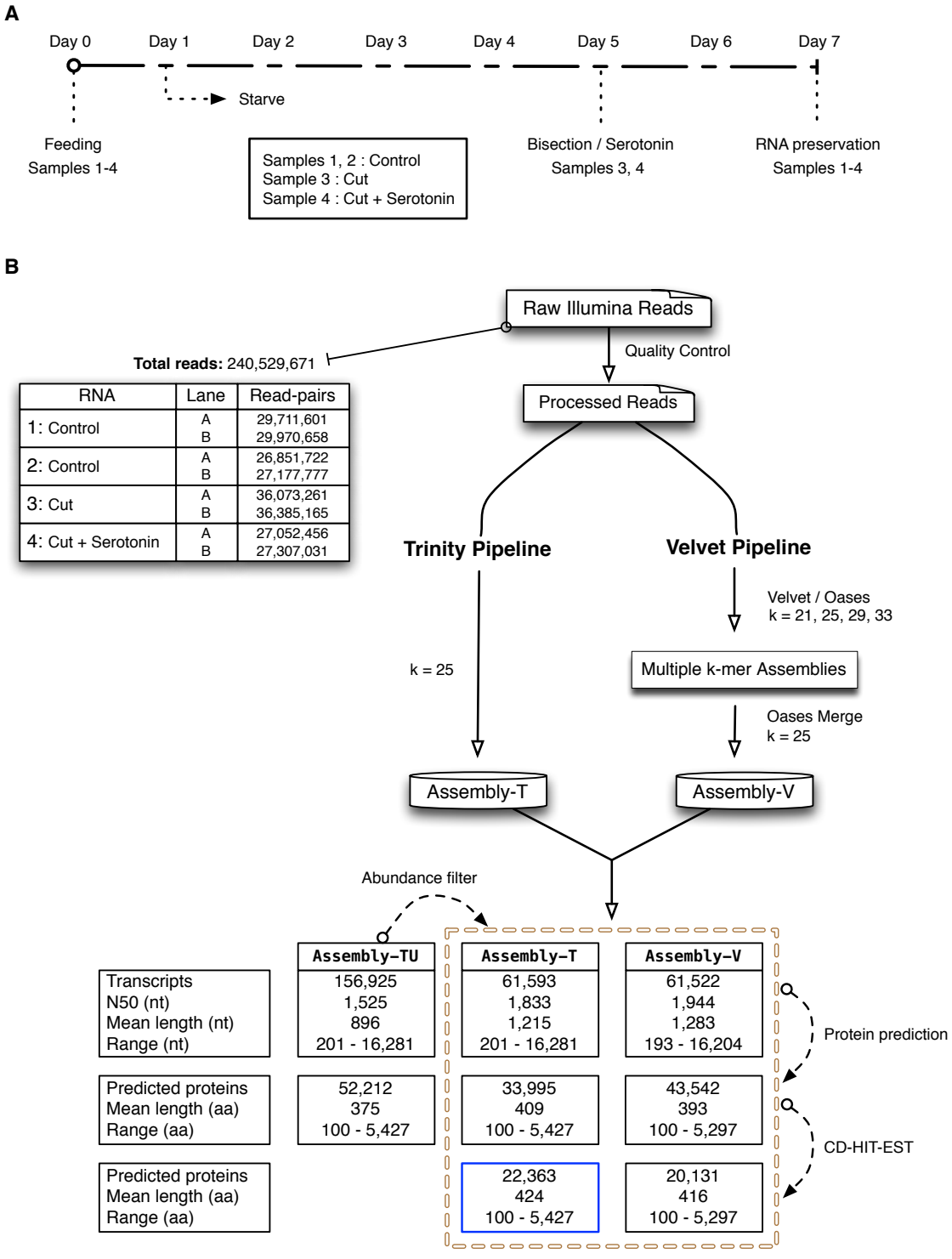

Figure 2.1: RNA-Seq workflow. A) Planarian feed-starve cycle and RNA extraction timeline. B) Processing of raw reads and parallel *de novo* transcriptome assembly using the Trinity and Velvet/Oases pipelines. The table depicts relevant statistics for the transcriptomes and predicted proteomes associated with each pipeline through various post-assembly filtering stages.

2.4.2 Protein prediction and transcriptome annotation

Predicted proteomes were created for each assembly using Transdecoder (http://transdecoder.sourceforge.net/) to evaluate the coding potential of open reading frames based on codon usage. After subtraction of redundant proteins, the Assembly-T proteome had a marginally larger unique protein count and was therefore used for all subsequent analysis. The 22,363 predicted proteins from this dataset were used in blastx queries against the NCBI nr database [30]. 16,467 sequences had at least one significant hit (E-value < 0.001) and the top 20 hits were retained for each sequence. We then applied Gene Ontology (GO) annotations to the *de novo* transcriptome using the Blast2GO pipeline [46]. The mapping of sequence-specific blast results to GO identifiers resulted in functional annotation of all 16,467 proteins with significant blast hits, accounting for 73% of the predicted proteome. This was complemented by InterProScan domain mapping, resulting in the identification of at least one domain for 18,051 protein sequences [95].

Annotated predicted proteins were categorized according to their involvements in various biological processes at different hierarchy levels. Figure 2.2 depicts this categorization from more general level 2 categories through more specific level 6 categories, with many proteins binned into multiple categories. Separately, to further gauge coverage of core pathways, Kegg pathway mapping was carried out [97]. As a representative example, over 95% of the reference glycolosis pathway enzymes were identified (data not shown). The very small numbers of unmapped enzymes in these and related metabolic pathways could plausibly also result partly from fundamental biological differences, as opposed to gaps in our dataset.

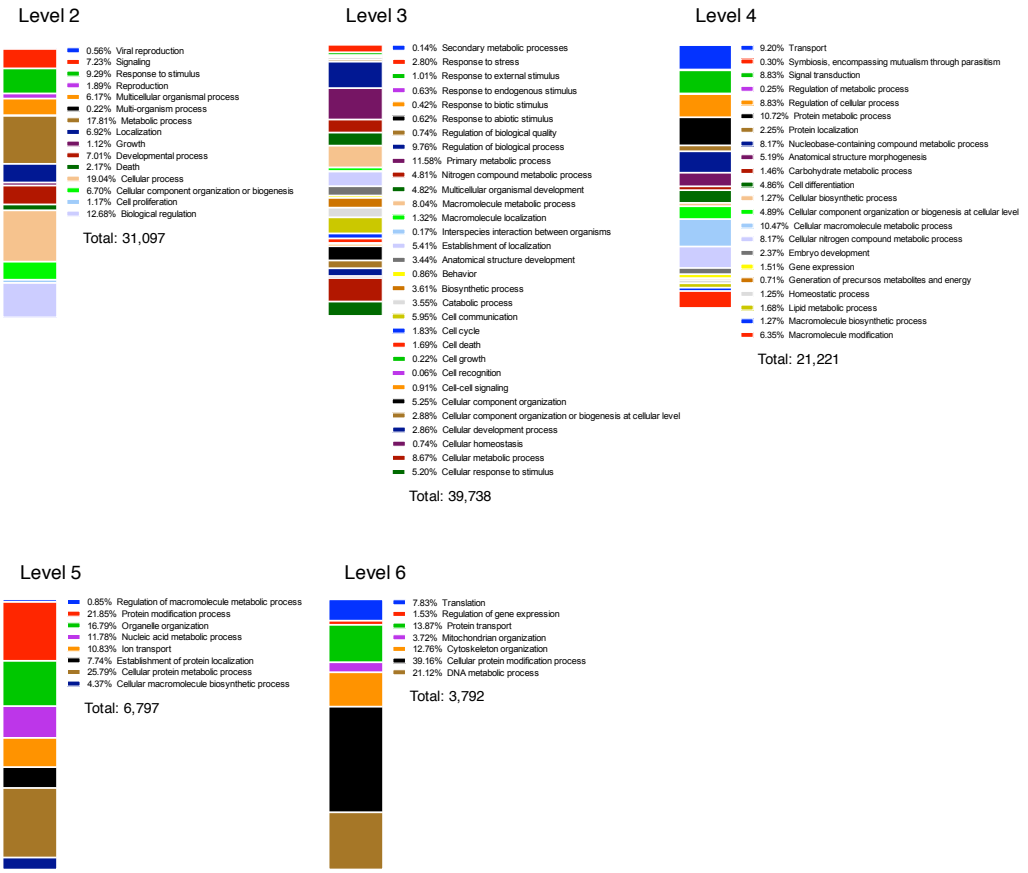

Figure 2.2: Comprehensive annotation of RNA-Seq transcripts. GO annotations reveal the wide range of biological process associated with the *D. tigrina* predicted proteome. Level 2 categories provide the most general functional annotation, with the largest fraction of proteins assigned to cellular and metabolic processes. The majority of signal transduction related proteins are captured in Level 4 annotations, and Level 6 annotations contain predictions for smaller numbers of proteins implicated in more specialized pathways (e.g. protein modification and DNA metabolism).

2.4.3 Identification of differentially expressed transcripts

While the inclusion of different treatment conditions was primarily aimed at increasing transcript capture, it also presents an opportunity to identify transcripts that are significantly upregulated or downregulated with respect to these conditions. Previous investigators have carried out experiments to identify regeneration-associated genes in *S. mediterranea* by performing RNAi screens to perturb normal regeneration [169]. While many transcripts show greater than

four-fold differences in expression between control and cut worms (e.g. RNA-dependent RNA polymerase, RNA helicases, reverse transcriptase), these data do not lend themselves to facile implications of molecular mechanisms of regeneration and are provided for further examination and investigation (see Additional Materials).

2.4.4 Identification of planarian-parasite orthologs

The sequencing and assembly of the *D. tigrina* transcriptome allows for an important genetic comparison between this free-living worm and its pathogenic relatives. Proteinortho was used to identify orthologous protein sequence groups shared between and among *D. tigrina* and the parasitic species *Schistosoma mansoni* and *Echinococcosis multilocularis* [118]. This program employs an efficient reciprocal best alignment algorithm, yielding a very conservative but reliable subset of likely ortholog groups using the predicted proteomes of a set of species. A total of 3,179 orthologs were identified for the *D. tigrina* - *S. mansoni* pairing, contrasted with a more expansive pairwise homology (BLASTp) search which identifies over 10,000 significant (E-value > 0.01) hits. Overall, 2,693 sequences were identified as belonging to ortholog groups that spanned all three species (Figure 2.3).

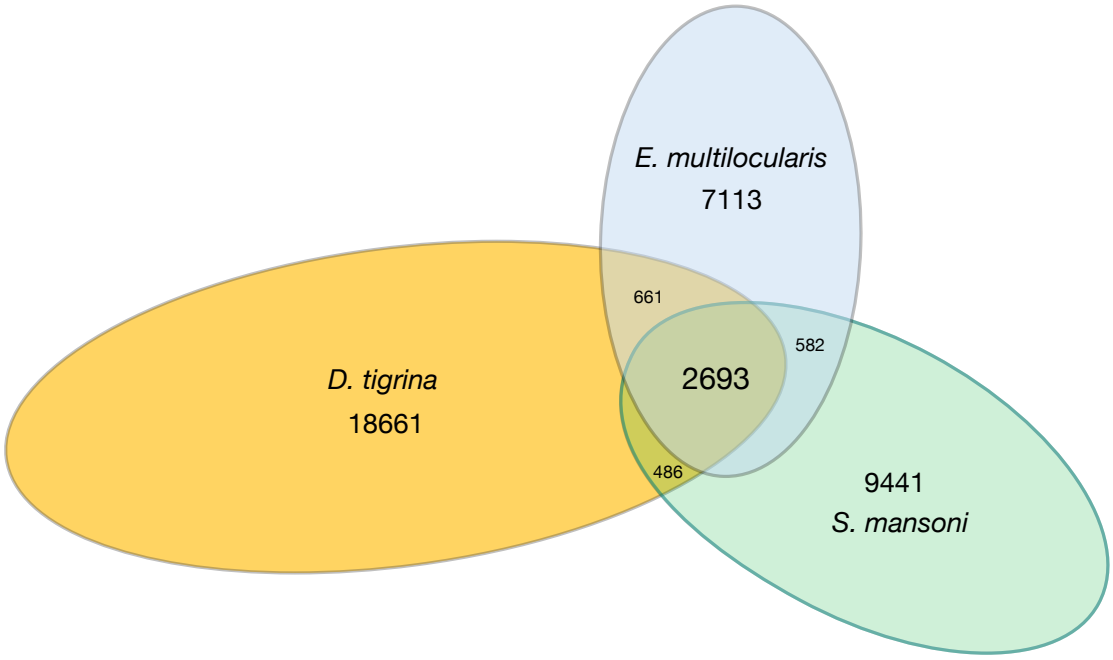

Figure 2.3: Protein-level orthology among free-living and parasitic flatworms. Orthology analysis via Proteinortho identified 2,693 orthologues shared among all 3 flatworms, and several hundred between flatworm pairs. In this high-stringency approach, the majority of *D. tigrina* transcripts were not identified as *bona fide* orthologs, despite substantial sequence homology.

To better visualize these relationships, a Circos diagram was created that mapped the chromosomal arrangements of orthologous genes for the selected parasites (Figure 2.4) [110]. Given that a stand-alone transcriptome lacks this spatial information, *D. tigrina* transcripts were arbitrarily ordered to allow for the mapping of planarian transcripts to the parasite genomes. Figure 2.4A and Figure 2.4B show pairwise individual sequence comparisons between *D. tigrina* and the two parasites. The ideograms highlight the genomic locations of identified parasite orthologs, and are surrounded by heatmaps that display the percent sequence similarity shared for each planarian-parasite ortholog pair, as well as for each parasite sequence and its nearest-matching human homolog, identified with BLASTp searches against the RefSeq human proteome.

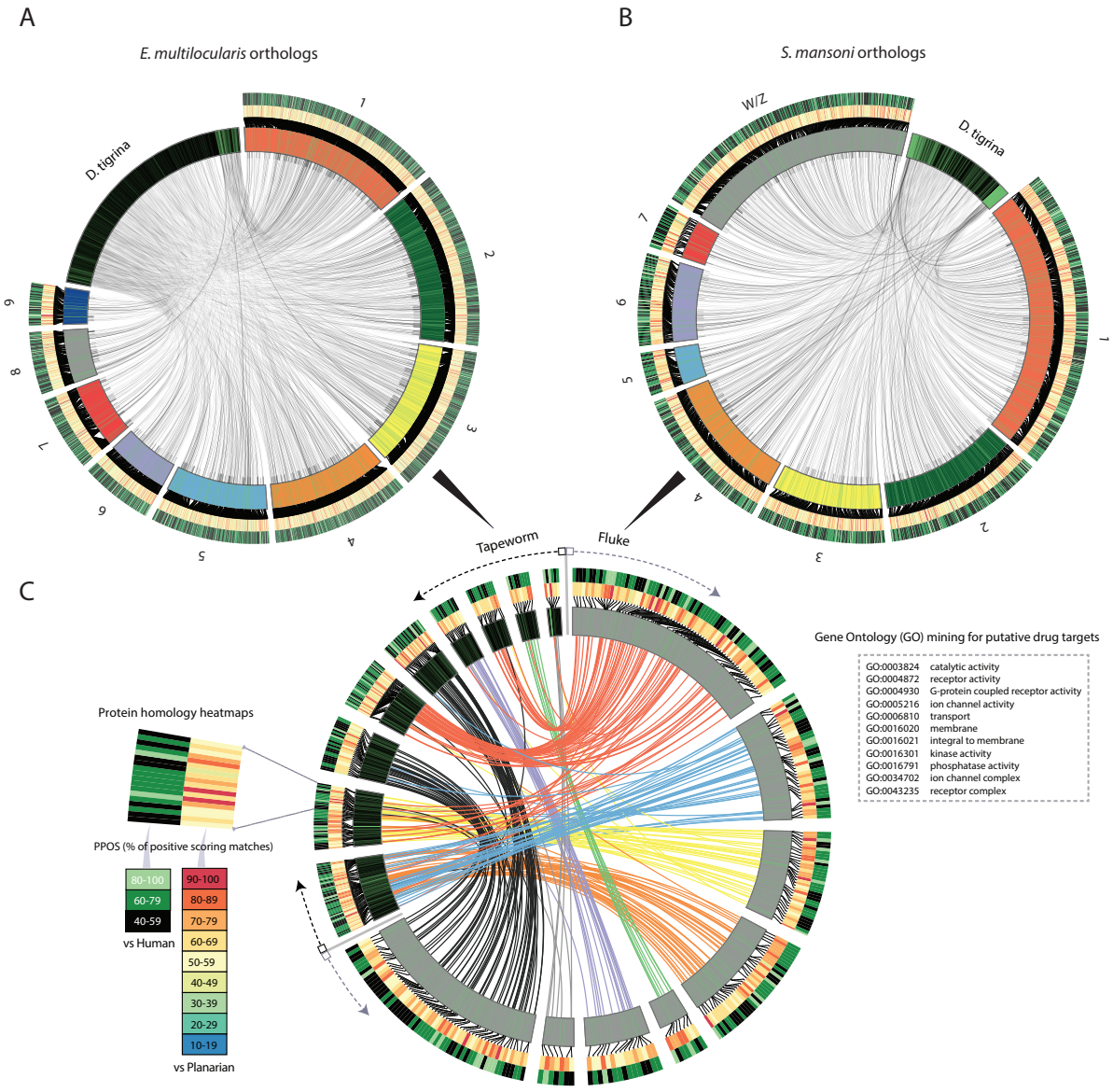

Figure 2.4: Mapping of orthology relationships between *D. tigrina* and the genomes of pathogenic flatworms. A and B) Circos diagrams depicting ortholog pairs between arbitrarily arranged *D. tigrina* transcriptome and the genomes of the tapeworm *E. multilocularis* and the blood fluke *S. mansoni*. C) Ideograms are shown for *E. multilocularis* and *S. mansoni* chromosomes. Physical ortholog links reveal synteny between these parasites for putative drug targets. Links are shown only where there exists a *D. tigrina* ortholog. Drug targets were extracted by mining the *D. tigrina* predicted proteome for GO terms displayed in the box on the right. The inner heat map shows the percent similarity (ppos) between parasite and planarian ortholog protein pairs, and the outer heat map shows similarity between a given parasite protein and its nearest human homolog (RefSeq).

2.4.5 Drug target prioritization

Ortholog groups with high sequence conservation through the phylum represent potential broad-spectrum therapeutic targets, and these can plausibly be interrogated using *D. tigrina* as a more tractable free-living model. Ideally, we want to prioritize planarian protein targets that share very high sequence similarity with both fluke and tapeworm homologs, and that exhibit lower levels of sequence similarity with any identifiable host (human) proteins. Figure 2.4C applies this selection logic towards exploitation of the available sequence data. Here, links between *S. mansoni* and *E. multilocularis* reveal synteny for those orthologous gene pairs that share a highly similar *D. tigrina* ortholog and which represent lucrative anthelmintic targets. The final links are restricted to a set of 441 putative drug targets, filtered from the initial set of 2,693 ortholog groups.

To estimate the druggability of protein targets, we utilized GO annotations that are most often associated with established drug target classes, which are found at DrugPort. Highly-associated GO terms, manually supplemented, were collected into a list and used to extract specific sequences from the annotated ortholog dataset. Within this set, we looked to identify a handful of targets as proof of concept for our model paradigm. Specifically, three targets were chosen that showed high sequence similarity between free-living and parasitic flatworms and that we could potentially pharmacologically manipulate or inhibit with commercially available chemicals. These targets were actin-related protein complex 2/3 subunit 2 (ARPC2), succinate dehydrogenase (SDH1), and NADH dehydrogenase (ubiquinone) flavoprotein 2 (NDUFV2).

2.4.6 Comparative chemical screen of targets

To carry out a comparative first-pass phenotypic screen, chemical inhibitors for each target were used to treat newly transformed *S. mansoni* schistosomula and planaria across a range of concentrations. It should be noted that these chemical inhibitors were first developed to act on mammalian proteins, and the specificity of each interaction is therefore unknown. However, there is significant sequence conservation between protein domains in these proof of concept targets and their human homologs (Additional Materials Figure 2.11 - Figure 2.13), suggesting

a high likelihood of a conserved mode of action. For example, *Dugesia* ARPC2 shares 58% sequence identity with its human counterpart. While it can be hypothesized that CK-666, which locks this complex in an inactive state [81], performs this action in flatworms as well, it can not be necessarily inferred.

Motility phenotypes were measured in terms of body contractions (bends per second) for *S. mansoni* and average velocity (mm per second) for *D. tigrina*. As shown in Figure 2.5, the dose response curves for each chemical elicited a similar phenotypic response profile for both *S. mansoni* and *D. tigrina*. CK-666 (ARPC2 inhibitor) and 3-nitropropionic acid (3-NPA; SDH1 inhibitor) caused dose-dependent decreases in motility in both worms. In contrast, rotenone (NADH dehydrogenase inhibitor) did not alter either worm's motility. This further evidences the notion that pharmacological manipulation of highly conserved flatworm molecules in planaria can be predictive of phenotypic outcomes in schistosomes. One target that brought about a phenotypic effect in the pharmacological screen, ARPC2, and one that showed no apparent effect, NDUFV2, were then further examined with RNA interference (RNAi).

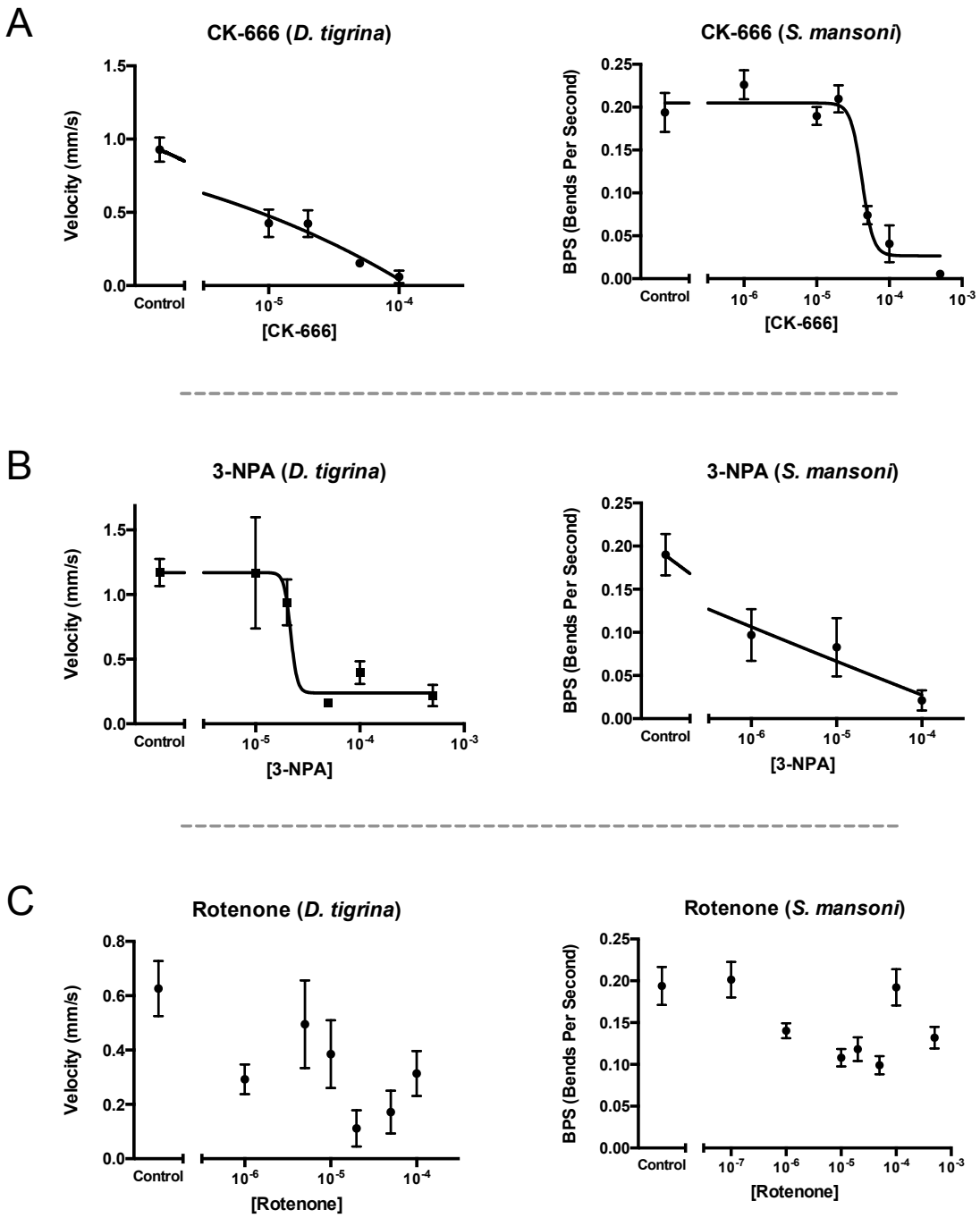

Figure 2.5: Comparative effects of pharmacological inhibition on motility. A-C) Pharmacological inhibition of three putative targets leads to correlative motility phenotype in free-living (*D. tigrina*) and parasitic flatworms (*S. mansoni*). Chemical inhibitors of ARPC2 (A) and SDH1 (B) caused a dose-dependent decrease in motility in both species as measured by contractions per second (schistosomula) or millimeters of translational movement per second (planaria). Chemical inhibitor of NDUFV2 (C) did not have any dose-dependent effects in either organism. Nonlinear regression is fit to a four-parameter variable slope model; log(inhibitor) vs. response. Bars represent SEM.

2.4.7 RNAi in planaria is predictive of phenotypes in parasites

Complementary to the pharmacological screen, the expression of *ARPC2* and *NDUFV2* was suppressed using RNAi. After three dsRNA feedings dispersed over a 7 day timeline, semi-quantitative PCR was used to confirm near-complete transcript knock-down (Figure 2.14 - Figure 2.15). Phenotypic analyses were carried out on *D. tigrina* by monitoring motility and regeneration as commonly assayed outcomes of gene suppression in planaria. Motility was not significantly altered for either experimental group (Figure 2.6A), although this might be expected for NDUFV2 considering chemical inhibition resulted in no motility phenotype.

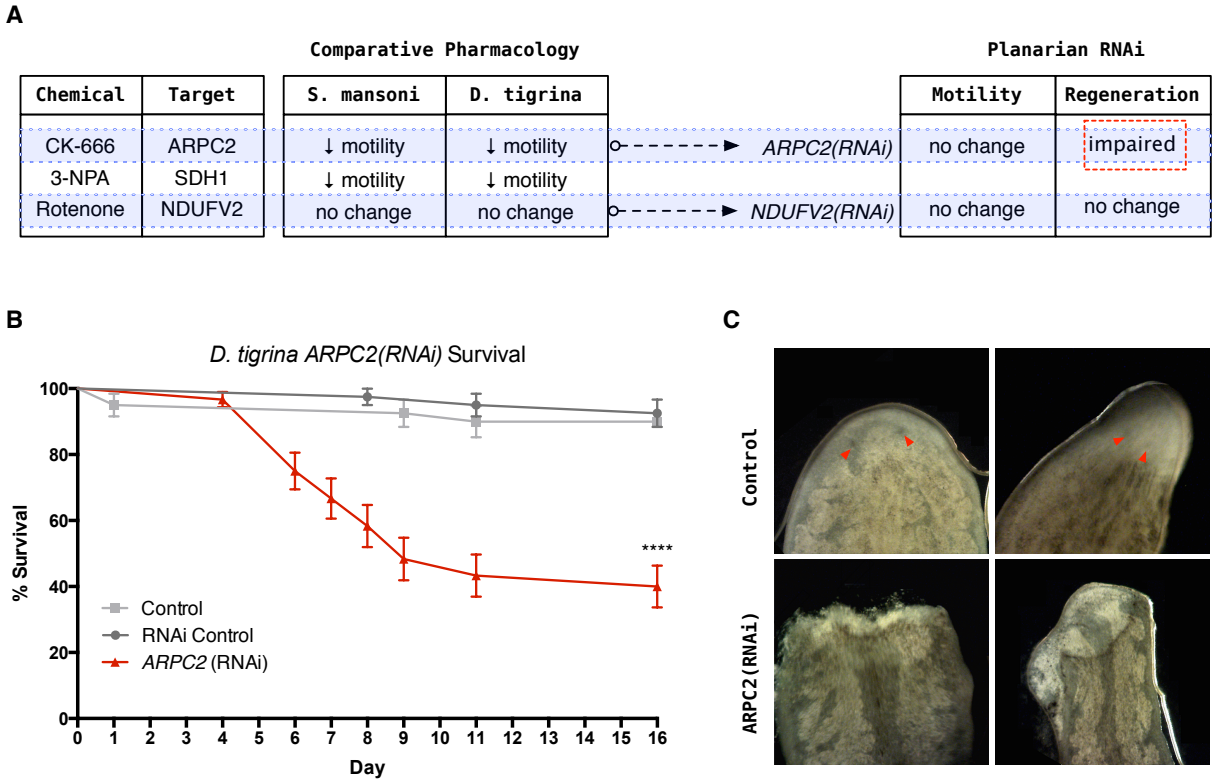

Figure 2.6: RNAi phenotypes correlate to pharmacology screen. A) Comparative pharmacology led to the selection of *ARPC2* (increased motility) and *NDUFV2* (no noticeable phenotype) as proof of principle targets for RNAi-mediated knock-down. RNAi of both targets did not bring about any significant changes in planarian motility. However, *ARPC2* suppression was lethal to regenerating *D. tigrina*. These observations correlate to the pharmacology screen where no phenotype was observed with application of NDUFV2 inhibitor, but stark motility phenotypes were observed with ARPC2 chemical inhibitor. B) Survival curves show significantly decreased ($P < 0.0001$; Log-rank Mantel-Cox test) rates of survival for *ARPC2(RNAi)* cut worms in comparison to control cut worms. C) Prior to death, caudal fragments of *ARPC2(RNAi)* cut worms showed impaired sealing of the initial wound and improper blastema formation. Proper eye spot formation can be observed in control worms (red arrows) and is absent in the *ARPC2* suppressed worms.

To assay regeneration, worms were bisected above the pharynx and each half was maintained in a separate well. Cephalic and caudal regeneration was observed over the course of 2-3 weeks. No developmental phenotype was observed for *Dtig-NDUFV2(RNAi)* worms, however,

Dtig-ARPC2(RNAi) worms showed aberrant regeneration in comparison to control worms, consisting of a range of specific outcomes that included stalled or slowed regeneration, blastema malformation, an inability to seal the wound, and eventual death (Figure 2.6B and Figure 2.6C). In these assays, readily identifiable phenotypes were observed in planaria that were predictive of visible phenotypes in the comparative chemical screen.

This proof of principle was limited to a handful of putative targets with commercially available inhibitors, but the pipeline can readily be scaled to larger numbers of targets. In the proposed scheme, highly conserved planarian-parasite orthologs are first interrogated with RNAi in planaria. The detection of aberrant phenotypes relating to planarian motility, morphology, or regeneration can serve as an efficient and high-throughput filter for target druggability in flatworm parasites. Parasite orthologs can then be investigated using available techniques such as RNAi and heterologous expression for drug target validation and functional characterization. This approach sidesteps the often prohibitive costs and technical challenges of carrying out large high-throughput screens in transient parasite life stages.

2.5 Conclusion

This work further promotes the adoption of planaria, and in particular *D. tigrina*, as a model screening organism for candidate drug targets in parasites. We provide a high-coverage annotated *de novo* transcriptome as a substrate for such efforts. The identification of ortholog groups that extend to planaria, blood flukes, and tapeworms, allows for the rational prioritization of likely broad-spectrum drug targets that can be readily screened in *D. tigrina*. We outline a pathway for the high-throughput evaluation of putative drug targets in planaria as a prelude to validation and more extensive characterization in parasitic flatworms. We further show how such screens can be predictive of biological phenotypes in parasites.

This study builds on other recent studies that have shown the utility of the planarian system in understanding parasite biology. For example, the antischistosomal praziquantel has been shown to lead to changes in planarian regenerative polarity through the action of voltage-operated calcium channel (VOCC) β subunits [144]. In this conserved signaling pathway, regenerative polarity in planaria acts as a phenotypic correlate of drug efficacy and worm paralysis

in schistosomes. This was followed by a more comprehensive investigation of the phenotypic correlates of manipulating signal transduction pathways in the planarian *D. japonica* and the parasite *S. mansoni*, as a predictive tool for the discovery of antischistosomal agents [50]. The striking identification of adult stem cells in *S. mansoni* that resemble planarian neoblasts further strengthens the notion of fundamental biological conservation between free-living and parasitic flatworms [219]. The annotated *D. tigrina* sequence resource, along with the orthology-based prioritization of putative drug targets, can act as a valuable substrate to help catalyze a low cost and scalable *in vivo* pipeline for anthelmintic drug discovery.

2.6 Methods

2.6.1 Planarian culture and RNA isolation

D. tigrina (Ward's Natural Science, Rochester, NY) colonies were maintained in aerated spring water. Planarians were starved for one week prior to RNA isolation. Five animals were randomly selected per experimental condition on day 7. Each group was washed repeatedly with spring water and tissue grinding was carried out using mortar and pestle in the presence of liquid nitrogen. A hybrid TRIzol (Invitrogen)/ RNeasy (Qiagen) protocol was used to isolate total RNA from ground tissue, whereby supernatants from the chloroform phase separation were combined with an equal volume of 100% ethanol and loaded into RNeasy columns for purification. Total RNA quality and concentration was assessed with an Agilent Bioanalyzer 2100. RNA integrity number (RIN) proved to be a poor benchmark of RNA quality, as the *Dugesia* 28S rRNA subunit is evidently converted into fragments that co-migrate with 18S rRNA to produce a triple-peak, giving the misleading appearance of RNA degradation. All samples yielded at least 1 ug/ul of RNA when eluted in 40 ul of H_20, with an OD A260/A280 of ~2.1 and OD A260/A230 of ~2.2.

2.6.2 Library preparation and Illumina paired-end RNA-seq

Illumina HiSeq 2000 paired-end (2x100 bp) library preparation and sequencing was carried out at the McGill University and Genome Quebec Innovation Center. The four RNA samples

were multiplexed across two sequencing lanes with an average fragment size of $\sim$ 350 bp, corresponding to an average insert size of $\sim$ 224 bp. The sequencer run yielded $\sim$ 30 million paired-end reads per sample (241 million total paired-end reads) with an average Phred quality score of 37.

2.6.3 *De novo* transcriptome assembly

Adapter sequences were trimmed and reads were passed through a sliding window quality filter (window size = 4, minimum average quality score = 25) using Trimmomatic 0.22 [22]. Paired-end reads and singletons $\geq$ 50 bp in length were retained. Overlapping paired-end reads were identified and merged using FLASH [132] with an expected insert size of 220 bp. Quality control and read collapsing led to a total filtered pool of 165 million paired-end reads and 55 million singletons. Surviving reads were combined and fed into the Trinity [72] pipeline for *de novo* assembly, performed on the GLUMEQ Guillimin supercomputer maintained by McGill University. Assembly optimization and runs were carried out on a 1 TB ScaleMP node that allows for a virtualized shared large memory environment required by the OpenMP standard. Final assembly was carried out with a minimum k-mer coverage of 2 and the default k-mer size of 25. Complex graphs that proved unresolvable within a 6 hour window were manually excised to allow the assembly to proceed. Separately, an available Python script was used to feed the same read pool into the Velvet [237] pipeline and to generate multiple k-mer assemblies (k = 21, 25, 29 and 33) for merging with Oases (k = 25) [191]. The minimum contig or transcript length for both assembly pipelines was set to 200 nt. The statistical software R [208] was used to generate and evaluate assembly statistics. Further bioinformatic analysis was restricted to the Trinity transcriptome.

2.6.4 Transcriptome filtering and annotation

Filtered paired-end reads were mapped to the Trinity transcriptome with Bowtie [114]. Abundance estimation with RSEM [121] was used to select for transcripts that accounted for at least 1% of the per-component (IsoPct) expression and that met a TPM cutoff of 1. Open reading frames (ORFs) with coding potential were predicted from the final transcriptome using log-

likelihood scores based on codon usage with Transdecoder (http://transdecoder.sourceforge.net/). The resulting predicted proteome was further filtered with CD-HIT-EST [123] at a threshold of 0.95. BLAST2GO [46] was used to functionally annotate the *Dugesia* predicted proteome and to assign GO terms to predicted proteins.

2.6.5 Differential expression analysis

Condition-specific abundance estimation was carried out with Bowtie and RSEM using the final filtered transcriptome. Existing Trinity scripts and the R/Bioconductor packages DESeq and edgeR of the statistical programming language R were used to identify differentially expressed transcripts. A threshold e-value of 10^3 and minimum four-fold expression changes were used to select and cluster transcripts as either up or down-regulated. Transcript sets were then mapped to previous annotations, where available.

2.6.6 Orthology analysis and drug target prioritization

Proteinortho [118] was used to detect ortholog groups between and among the *D. tigrina*, *S. mansoni* and *E. multilocularis* predicted proteomes. The predicted parasite proteins were downloaded from GeneDB [127]. The visualization tool Circos [110] was utilized to organize and display the orthologous relationships among these species in the context of Turbellarode, Trematoda, and Cestoda synteny. Provided that a transcriptome does not imply any spatial or chromosomal arrangement, the *D. tigrina* transcripts were arbitrarily arranged on a pseudo-chromosome - designated as an ideogram - to enable visualization. GFF files obtained from GeneDB (*S. mansoni*) or personal communication (*E. multilocularis*) containing gene coordinate data for the two parasitic species were parsed for necessary sequence features using a set of in-house Python scripts and used to draw orthologous "links" between ideograms. Heatmap data was created by running command-line BLAST [30]. Similarity calculations were carried out with *D. tigrina* orthologs as queries against their corresponding parasitic orthologs, as well as with parasitic orthologs as queries against the RefSeq human proteome [166]. Sequence homology was used to select orthologs that displayed high sequence similarity among the three examined flatworm species, as well those sufficiently diverged from their nearest-identifiable

human homolog. These prioritized ortholog groups were mined with GO IDs, along with manually-selected IDs, to extract and highlight annotated sequences belonging to notoriously druggable protein families.

2.6.7 Parasite maintenance

Snails (*Biomphalaria glabrata*) infected with *S. mansoni* were provided by the Biomedical Research Institute (BRI) (Rockville, MD). Cercariae were shed from snails by light exposure and subsequently mechanically transformed to schistosomula *in vitro* per existing protocols [136]. Somules were cultured in modified Basch medium (containing 10% Fetal Bovine Serum) at 37°C in 5% CO_2 atmosphere.

2.6.8 Schistosomula assays

Three small molecule inhibitors (CK-666, ARPC2; 3-NPA, SDH1; rotenone, NDUFV2) were purchased from Sigma-Aldrich (St. Louis, MO). Newly transformed schistosomula were incubated with varying concentrations of each inhibitor for 30 minutes and recorded at 10x magnification for 5 minutes. The wrMTrck plugin of ImageJ [190] was used to track schistosomules and quantify motility in terms of contractile rate (body bends per second; BPS).

2.6.9 Planarian pharmacology and motility analysis

Dugesia individuals were placed in 35 mm dishes filled with 4 mL of media supplemented with inhibitor at varying concentrations or an equal amount of solvent control, and the dishes were placed on a light box in a dark room. After 30 minutes of incubation, worms were recorded for 5 minutes by EthoVision [146], and motility was quantified by dividing the parameter DistanceTraveled (mm) by TimeInZone (s). Down-sampling was set to 5 to ensure that small bending and twisting motions were not factored. Tracking profiles were visually diagnosed for errors and manually edited where required. Errors were most often attributable to to light reflections off of the surface of liquid media or imperfect arena definitions.

2.6.10 Planarian RNAi and regenerative assays

Total RNA was extracted from homogenized *D. tigrina* and converted to cDNA with Ambion's RetroScript RT kit. 600 bp sequences were PCR amplified using primers designed with the online implementation of Primer3 [106]. T7 cites were added using a two-step PCR protocol, and dsRNA was created with the Ambion MegaScript RNAi kit. dsRNA was added directly to homogenized liver paste according to prescribed methods [183] (10 umol/worm for one hour). Feedings were performed on Days 1, 3, and 5, and worms were bisected immediately above the pharynx on Day 6. Two worms from each experimental group were set aside for semi-quantitative RT-PCR performed with Ambion's QuantumRNA18S Internal Standards kit. Worms were observed for defects in regeneration over the full regeneration period (∼2-3 weeks).

2.7 Acknowledgments

Infected snails were provided by BRI via the NIAID schistosomiasis resource center under NIH-NIAID Contract No. HHSN272201000005I. *E. multilocularis* GFF files were kindly provided by M. Zarowiecki. Nicolas J Wheeler was supported by fellowship funds from the Office of Biotechnology, Iowa State University. We gratefully acknowledge Compute Canada/CLUMEQ for access to valuable computational resources.

2.8 Additional Materials

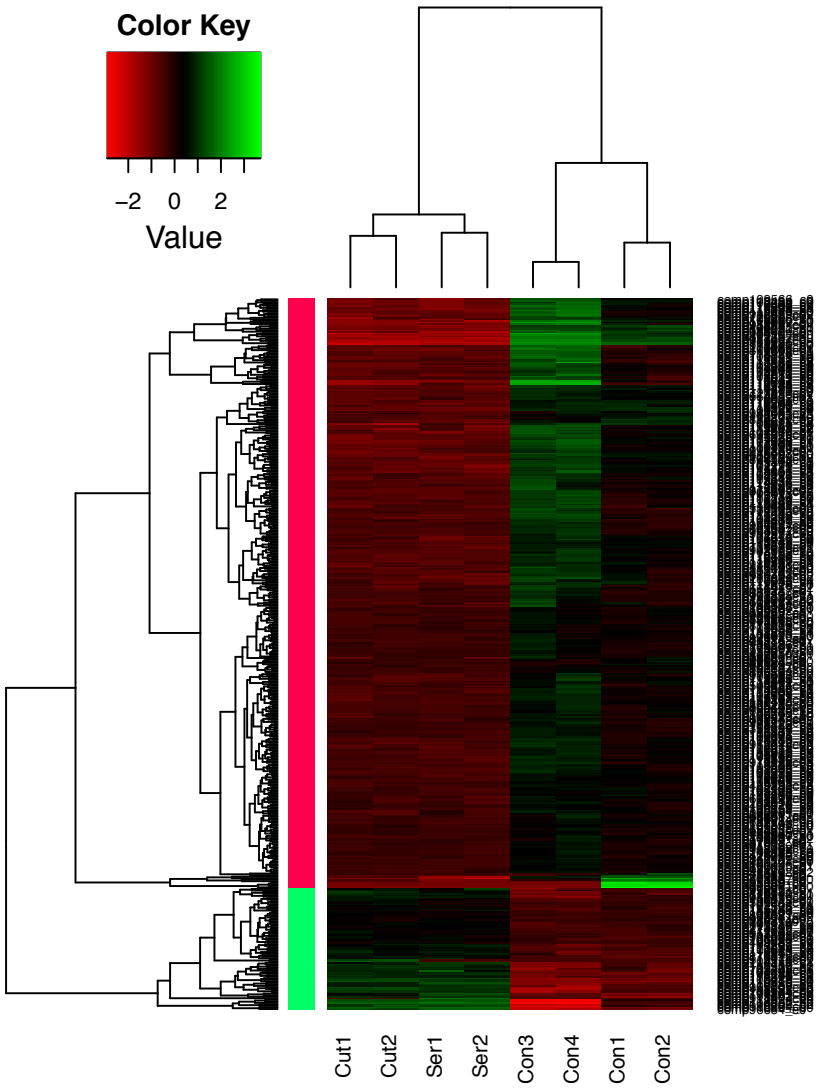

Figure 2.7: Differential Expression of control vs. cut animals. Green shows relatively increased expression, red shows relatively decreased expression.

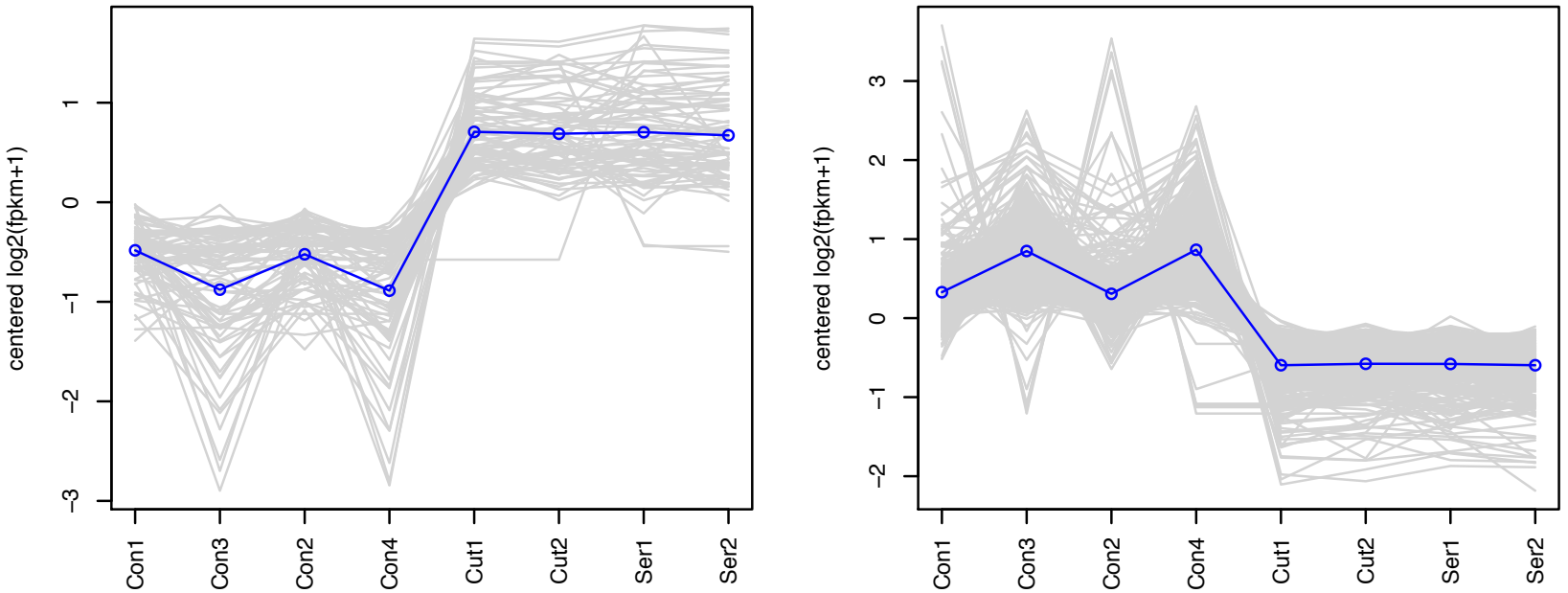

Figure 2.8: Differential Expression of control vs. cut animals. RNA-seq datasets show discrimination between different treatments.

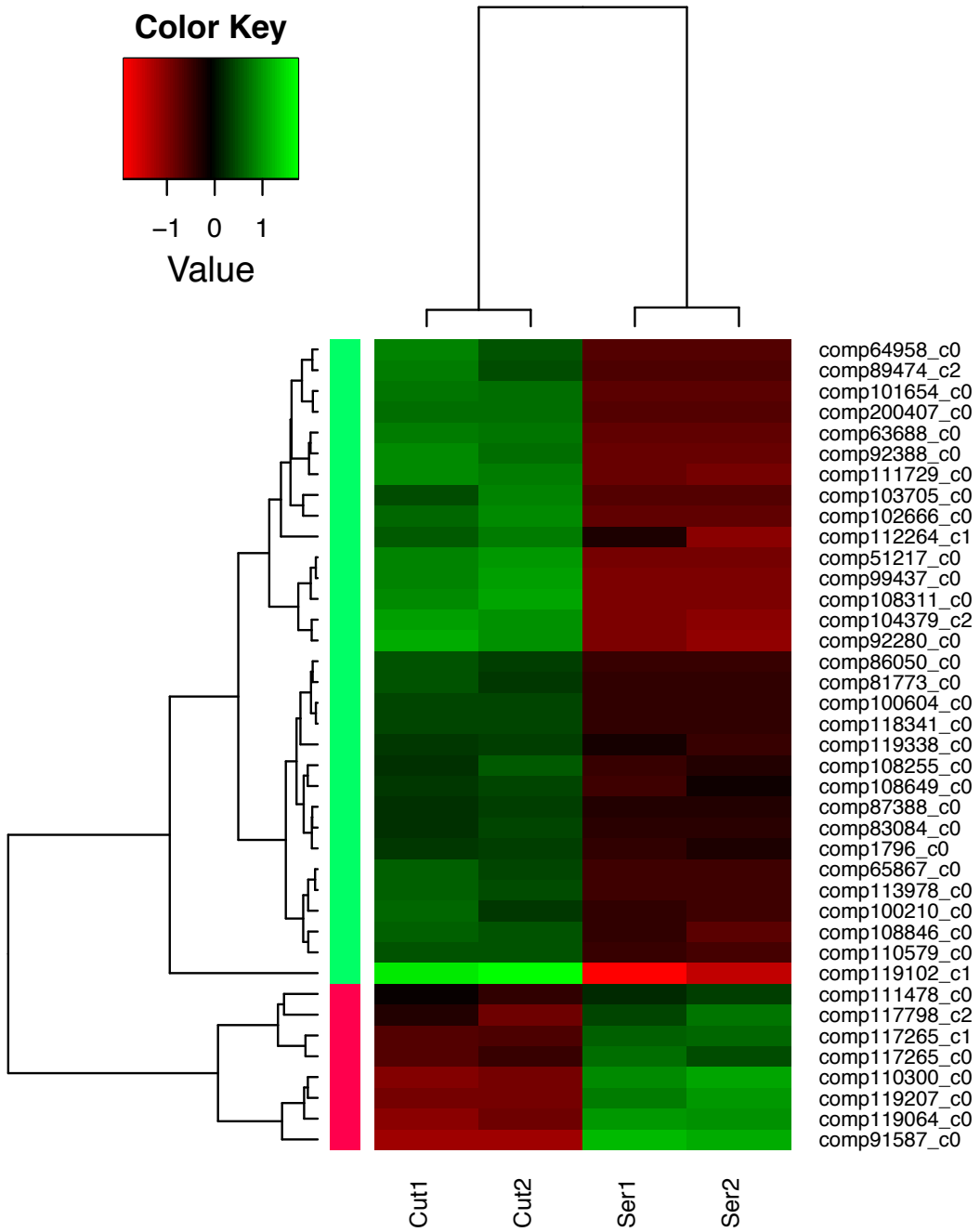

Figure 2.9: Differential Expression of cut (no serotonin) vs. cut (serotonin) animals. Green shows relatively increased expression, red shows relatively decreased expression.

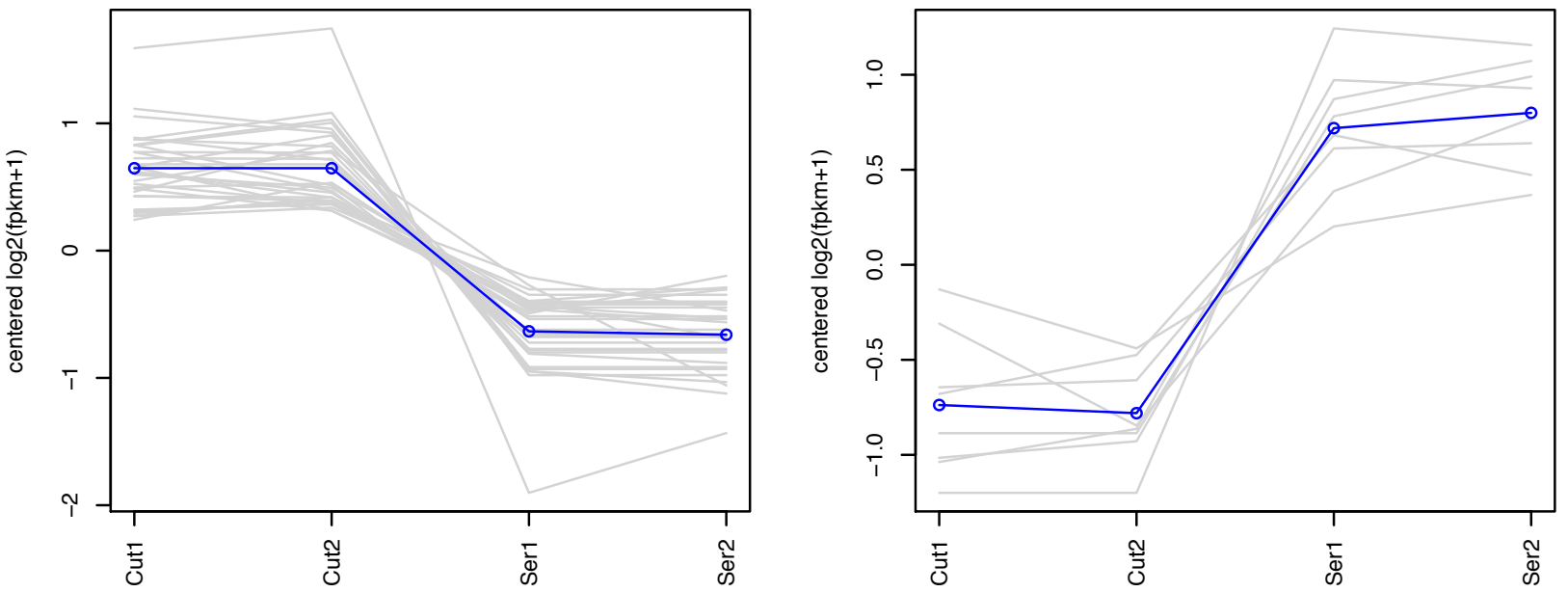

Figure 2.10: Differential Expression of cut (no serotonin) vs. cut (serotonin) animals. RNA-seq datasets show discrimination between different treatments.

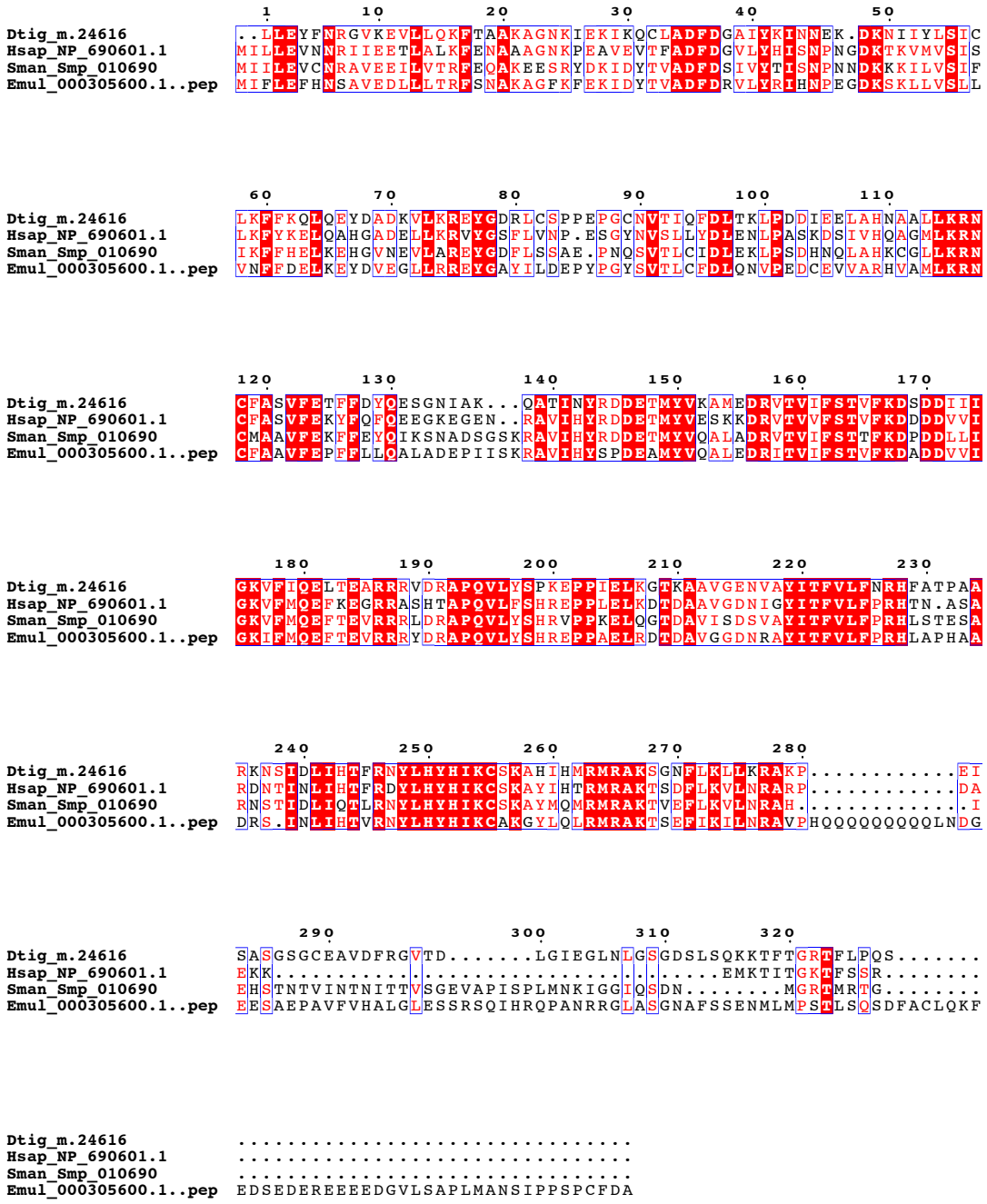

Figure 2.11: Flatworm-Human target homology. Sequence alignment ARPC2. Each *D. tigrina* target sequence was aligned against its nearest *S. mansoni*, *E. multilocularis*, and human homolog.

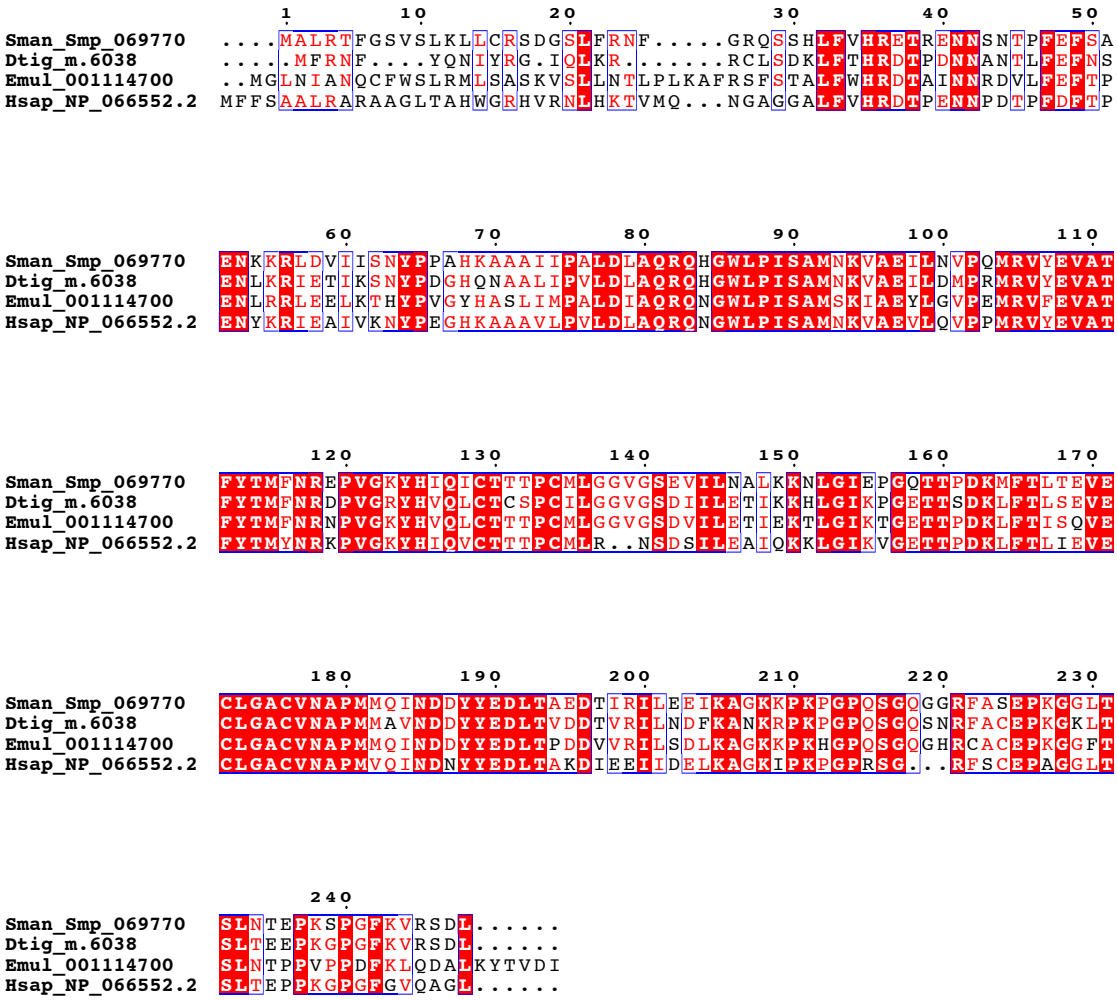

Figure 2.12: Flatworm-Human target homology. Sequence alignment SDH1. Each *D. tigrina* target sequence was aligned against its nearest *S. mansoni, E. multilocularis,* and human homolog.

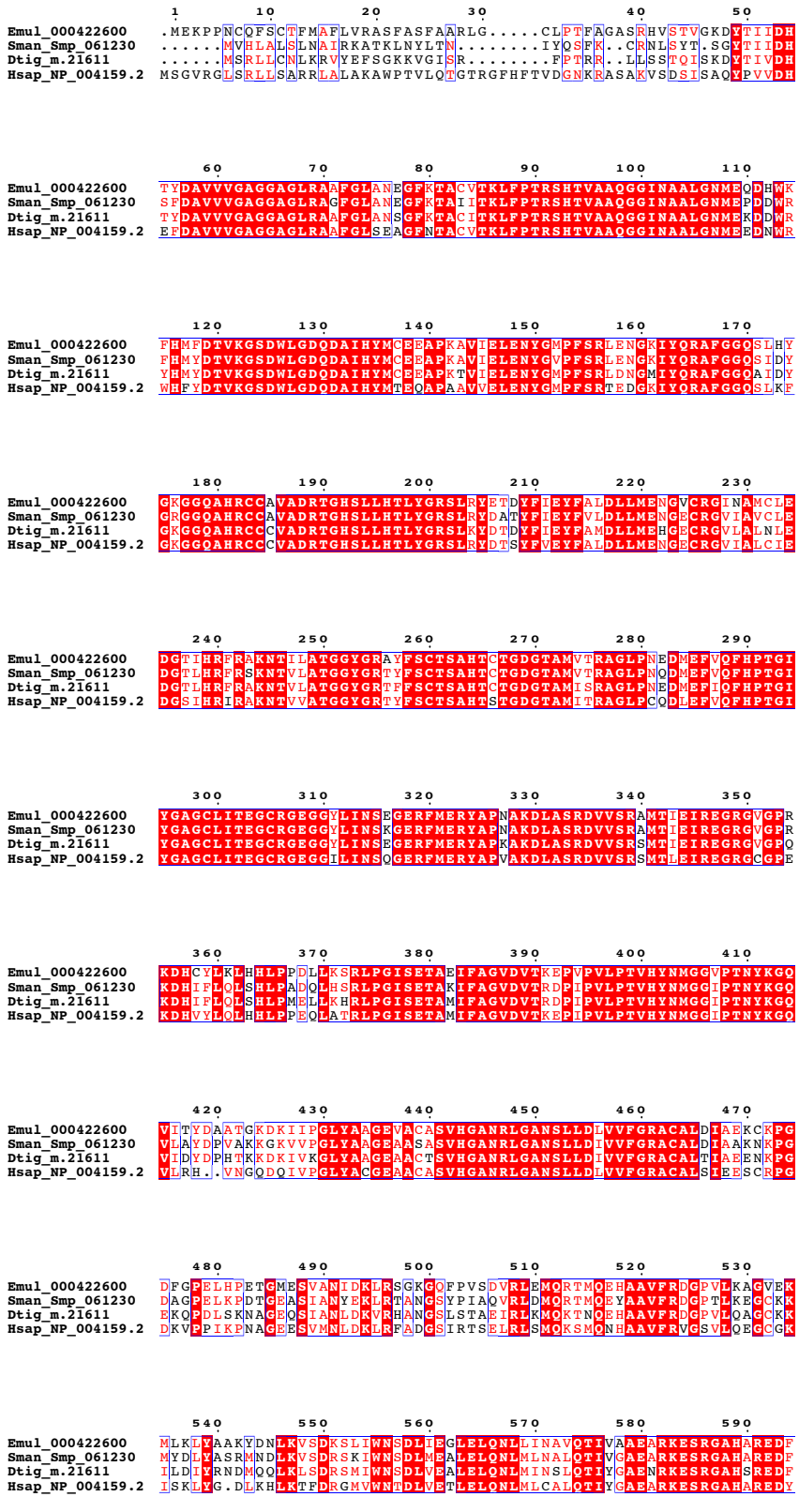

Figure 2.13: Flatworm-Human target homology. Sequence alignment NDUFV2. Each *D. tigrina* target sequence was aligned against its nearest *S. mansoni, E. multilocularis,* and human homolog.

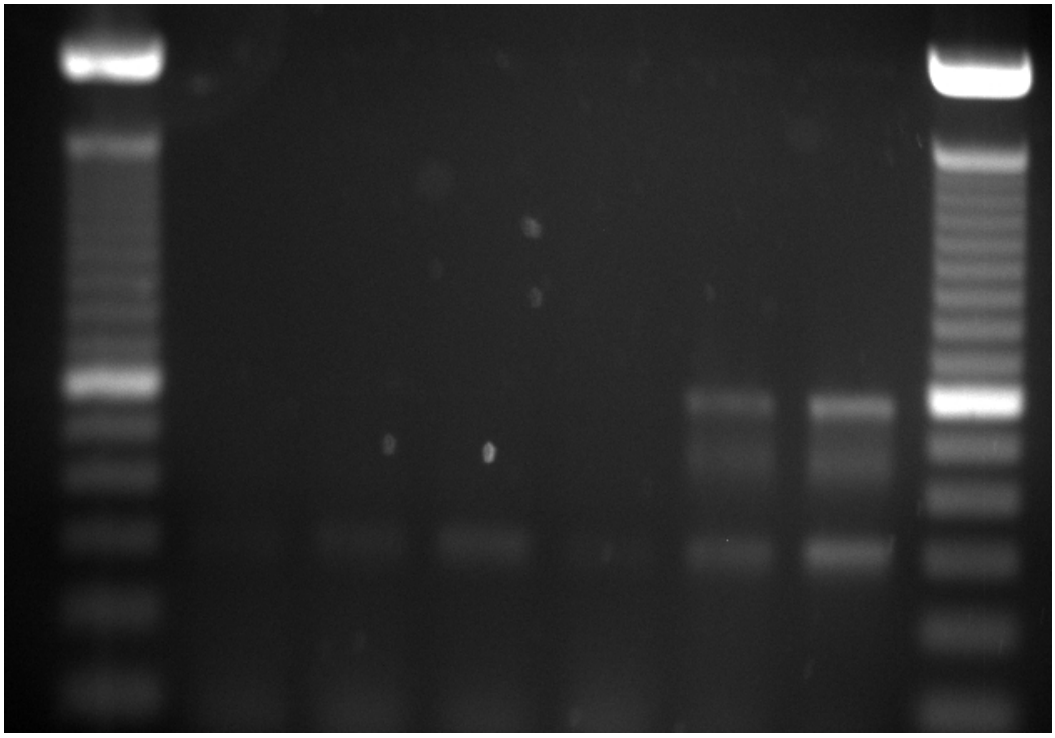

Figure 2.14: Semi-quantitative RT-PCR of RNAi experiments. Images of representative semi-quantitative RT-PCR gels. The first four non-ladder lanes are RNAi worms and the last two lanes are negative controls. The bottom bands are 18S Ribosomal RNA reference reference (300 bp), and the top bands are ARPC2 amplicons (600 bp).

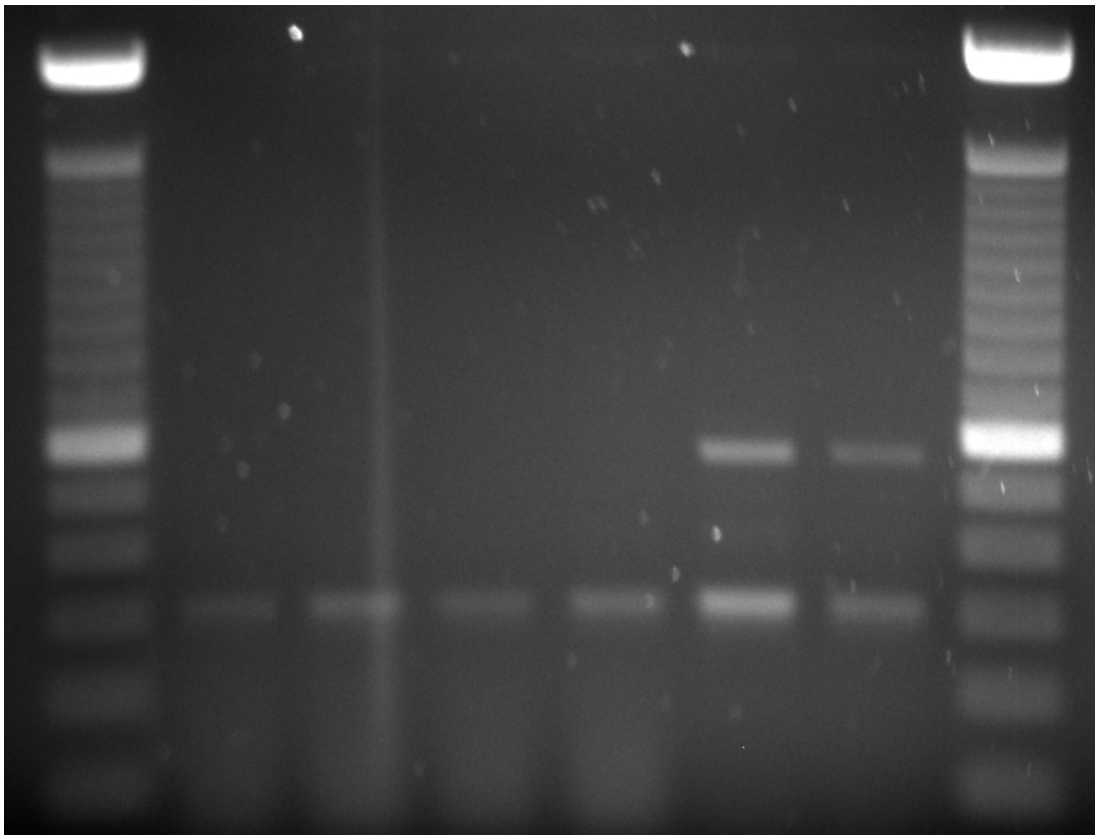

Figure 2.15: Semi-quantitative RT-PCR of RNAi experiments. Images of representative semi-quantitative RT-PCR gels. The first four non-ladder lanes are RNAi worms and the last two lanes are negative controls. The bottom bands are 18S Ribosomal RNA reference reference (300 bp), and the top bands are NDUFV2 amplicons (600 bp).

CHAPTER 3. PLATYHELMINTH-SPECIFIC RHODOPSIN-LIKE ORPHAN FAMILY: FAMILY EXPANSION, PHYLOGENY, AND IMPLICATIONS FOR DRUG TARGETING

A paper submitted, awaiting publication of seminal 50HGI manuscript

Nicolas J Wheeler[1,2,3] Mostafa Zamanian[4], Prince N Agbedanu[1], Michael J Kimber[1], Dennis Lavrov[5], Nancy Holyrod[6], Matthew Berriman[5], Tim A Day[1,2,7]

3.1 Abstract

G-protein coupled receptors, especially neuropeptide receptors, have been shown to be absolutely critical in platyhelminth biology and are the source of ongoing study as potential drug targets in parasitic flatworms. However, the high conservation of GPCRs between parasites and human hosts can complicate such inferences, as any chemical targeted to GPCRs is likely to cause host toxicity. It is hypothesized that this effect could be abrogated by focusing on phylum-specific proteins, paying particular attention to orphan proteins. Here, we describe the expansion of the recently discovered Platyhelminth-specific Rhodopsin-like Orphan Family to 25 flatworms, many of which have had genomes recently published via the 50 Helminth Genomes Initiative. These genomes were passed through the Compara comparative genomic pipeline, assigning each predicted gene to a conserved family. A combinatorial approach utilizing phylogenetics and Hidden Markov Models distinguished four PROF Compara families from

[1]Department of Biomedical Sciences, Iowa State University, Ames, IA USA
[2]Interdepartmental Genetics and Genomics, Iowa State University, Ames, IA USA
[3]Primary researcher and author
[4]Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, USA
[5]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA, USA
[6]Parasite Genomics, Wellcome Trust Sanger Institute, Hinxton, England
[7]Author for correspondence

the more then 200 GPCR families. A total of 238 predicted protein models were validated as PROFs, making it the largest clade of Class A (Rhodopsin-like) receptors. Additionally, the phylogeny Class A receptors shows the PROFs to be strongly related and most closely related to established neuropeptide receptors. This family is shown to be one of the largest phylum-specific families, contain three paraphyletic subfamilies, and have robust expression in the neurogenic tissue of juvenile and adult worms.

3.2 Introduction

Nearly 500 million people living in developing countries are parasitized by platyhelminths, causing upwards of 100,000 deaths per year and the loss of millions of healthy life-years. Four types of platyhelminthiases schistosomiasis (Schistosoma), foodborne trematodiases (Clonorchis, Opisthochis, Fasciola, Paragonimus), echinococcossis (Echinococcus), and taeniasis/cysticercosis (Taenia) have been designated by the World Health Organization (WHO) as neglected tropical diseases (NTDs) and subsequently emphasized in current and future research. Furthermore, infections in animals and livestock primarily trematodiases are equally significant and add to the burden experienced by developing countries, particularly in rural farming communities. While these diseases have often been conceptualized as diseases of only developing regions, cases of schistosomiasis have been reported in Europe [21], and climate change is hypothesized to further increase the incidence of flatworm infections across the European Union [31], further affirming the urgent need for additional NTD research.

In recent years, the WHO has initiated preventive chemotherapy in some endemic countries. In 2014, over 49 million at-risk children were treated with praziquantel, the efficacious anthelmintic used mainly for the prevention and control of schistosomiasis, and this number will continue to rise][228] . The other drug of choice, triclabendazole, is used for control of foodborne trematodiases. Unfortunately, the combination of the large burden and the widespread preventive efforts leads to the fear of drug resistance, which has been demonstrated in the laboratory [60] and reported in the field [71, 91].

New anthelmintics need to be developed and added to the repertoire of chemotherapeutics for the neglected platyhelminthiases. We and others have proposed the utilization of free-living

flatworms as a platform to identify, predict, and screen putative drug targets [36, 37, 44, 224]. The bottleneck for such a pipeline was the absence of necessary sequence information for both free-living and parasitic flatworms. Recent sequencing projects [165, 187, 214] have widened this bottleneck, and the path is clear for a new emphasis on identifying pan-phyletic drug molecules.

G protein-coupled receptors (GPCRs) are the mostly widely drugged protein class, targeted for all types of diseases, and their extreme functional and evolutionary diversity theoretically allows for specific targeting. Although traditionally perceived only as precise modulators, GPCRs in flatworms have been shown to be essential to proper motility, a key factor in successful parasitism [52, 53, 130, 159]. However, their general similarity to human receptors raises concern that toxicity would accompany drug targeting.

There has been renewed focus on orphan genes, or taxonomically restricted gene families that have no computationally identifiable annotation (Fang et al., 2015; Tang et al., 2012). So motivated, we have used bioinformatics to search for platyhelminth-specific gene families. Recently, comparative genomics and transcriptomics within a subset of the flatworms (Schmidtea (free-living), Girardia (free-living), Schistosoma (fluke), and Echinococcus (tapeworm)) uncovered a large family of GPCRs that showed no readily identifiable similarity to non-flatworm organisms [224, 236]. This Platyhelminth-specific Rhodopsin-like Orphan Family (PROF) was designated as a Class A GPCR subfamily and showed some similarity to other flatworm neuropeptide GPCRs. Other than these initial clues, no other functional information has been gleaned about these receptors.

At the time of their discovery, the PROFs appear to be greatly diverged from any mammalian or vertebrate ancestor, allowing the possibility of designing or discovering molecules that are selective, and even specific, for the parasite receptor family. However, only a handful of flatworms had sequenced genomes at this time, so the taxonomical-restriction of the PROFs could not be definitively determined. With the recent availability of over two dozen flatworm genomes and comparative genomic resources to classify phylum-specific gene families, we sought to search for PROFs in these new organisms. We have expanded the PROF family by identifying the PROFs in all flatworm genomic sources included in Sangers 50 Helminth

Genomes Initiative, which incorporated 25 different flatworms (unpublished, but available at www.parasite.wormbase.org). Additionally, we have inferred a comprehensive phylogeny to provide evolutionary information regarding distinctions between the three PROF subfamilies and fit the PROF within the scope of Class A GPCR evolution. Finally, we have conglomerated available life-cycle expression data from parasites and demonstrated neurogenic tissue-specific expression in free-living flatworms. We foresee these data providing invaluable information to aid and support deorphanization efforts and functional characterization, and the conclusions herein demanding further study as potential drug targets.

3.3 Results

3.3.1 PROF Gene Identification and Family Expansion

Our coarse GPCR analysis identified 200 putative GPCR gene families. We set three filters to parse putative PROF families from these 200 GPCR families. Each family was 1) Class A (Rhodopsin-like), 2) absent in taxa outside of the flatworms, and 3) had no readily identifiable, computationally-derived annotation. Importantly, in addition to dozens of nematodes and flatworms, several diverse animal outgroups, notably *Crassotrea gigas*, a mollusk and the most closely related outgroup to the Platyhelminthes, were included in the Compara family analysis. Four families survived filtering and were hypothesized to be PROFs. These families included 284 genes (238 with at least 5 and less than 8 transmembrane domains) with representatives from each of the 25 flatworms included in the analysis, making the PROF one of the largest clades of flatworm Class A GPCRs (unpublished data from 50HGI). The two largest PROFs contained at least one representative from each species included and are two of the top five largest phylum-specific families, further emphasizing the likely importance of these receptors to the biology of flatworms. The other two PROFs included representatives from each of the Trematoda and assorted others from Cestoda, Monogenea, and Turbellaria.

Species of the class Trematoda had on average more PROF genes per organism than the other flatworm classes (Table 3.1), and blood flukes (*Schistosoma* and *Trichobilharzia*) had substantially more than the liver or intestinal flukes (*Clonorchis, Fasciola, and Echinostoma*).

While both Monogenea (*Protopolystoma xenopodis*) and Turbellaria (*S. mediterranea*) only had one representative species, making statistically significant inference difficult, they each had fewer PROFs than the average Trematoda, but more than the average Cestoda. These findings corroborate the discovery of widespread GPCR family loss in Cestoda after splitting from Trematoda [214].

Table 3.1: Statistics of putative PROF subfamilies. PROF1 is the largest subfamily due to an expansion in trematodes. PROF3b was likely lost in cestodes. In general, trematodes have more PROF genes than cestodes, mimicking the pattern for the GPCR superfamily as a whole.

Species	Phylum	family 272753	family 302163	family 804503	family 1051524	Total
		PROF1	PROF2	PROF3a	PROF3b	
Diphyllobothrium latum	Cestoda	2	4	1	0	7
Echinococcus granulosus	Cestoda	1	3	1	0	5
Echinococcus multilocularis	Cestoda	1	4	1	0	6
Hydatigera taeniaeformis	Cestoda	1	4	1	0	6
Hymenolepis diminuta	Cestoda	1	4	1	0	6
Hymenolepis microstoma	Cestoda	1	2	1	0	4
Hymenolepis nana	Cestoda	1	4	1	0	6
Mesocestoides corti	Cestoda	1	3	1	0	5
Schistocephalus solidus	Cestoda	2	4	1	0	7
Spirometra erinaceieuropaei	Cestoda	1	6	3	0	10
Taenia asiatica	Cestoda	1	4	1	0	6
Taenia solium	Cestoda	1	4	1	0	6
Protopolystoma xenopodis	Monogenea	1	3	2	1	7
Clonorchis sinensis	Trematoda	3	5	2	1	11
Echinostoma caproni	Trematoda	3	6	2	1	12
Fasciola hepatica	Trematoda	1	5	1	1	8
Schistosoma curassoni	Trematoda	11	6	2	1	20
Schistosoma haematobium	Trematoda	10	6	2	1	19
Schistosoma japonicum	Trematoda	9	5	2	1	17
Schistosoma mansoni	Trematoda	10	4	2	1	17
Schistosoma margrebowiei	Trematoda	12	6	2	1	21
Schistosoma mattheei	Trematoda	10	6	2	1	19
Schistosoma rodhaini	Trematoda	9	6	2	1	18
Trichobilharzia regenti	Trematoda	22	5	2	1	30
Schmidtea mediterranea	Turbellaria	5	5	0	1	11
TOTAL		120	114	37	13	284

Upon the initial discovery of the PROF, 66 genes were identified in *S. mediterranea* and *S. mansoni*, but only 21 of these were linked to predicted gene models; the remainder were

manually found in the genomic sequence contigs [236]. Our analysis here was able to add 1 and 10 more putative PROF gene models to the *S. mediterranea* and *S. mansoni* complements, respectively, substantiating this approach, its power in identifying PROFs in other flatworms, and the value of using comparative approaches to review previous annotations of older genomes.

3.3.2 Phylogenetics

We used molecular phylogenetics to place these genes within the greater Class A of Rhodopsin-like GPCRs (Figure 3.1). As a model, we chose the receptors of *S. mansoni*, as they are the most completely annotated and contain the most nearly complete gene models. As expected, the 4 putative PROF subfamilies were found to form one monophyletic clade, paraphyletic with other Class A peptide receptors. We note that it appears that these orphan receptors are most closely related to neuropeptide receptors.

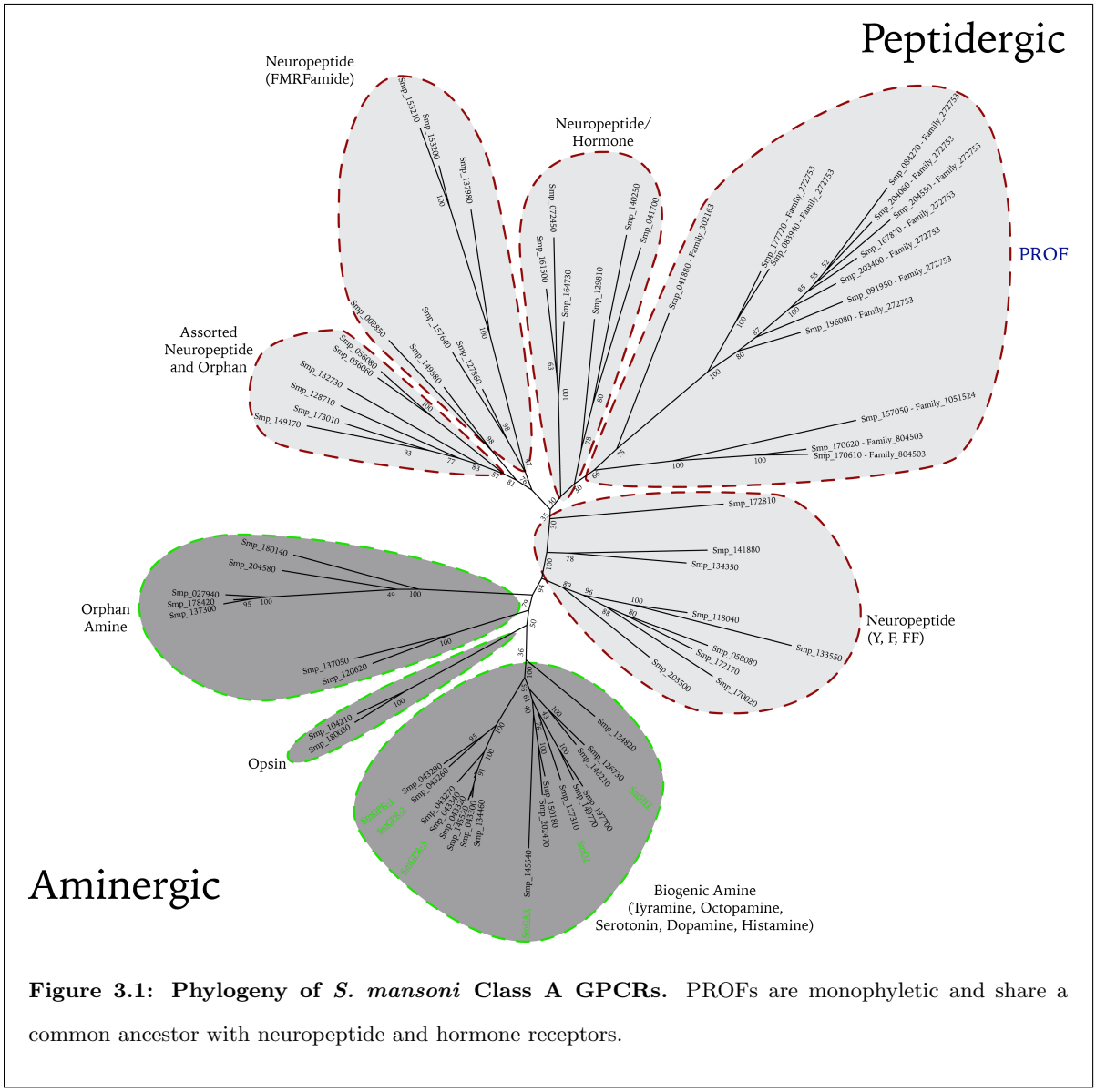

Figure 3.1: Phylogeny of *S. mansoni* Class A GPCRs. PROFs are monophyletic and share a common ancestor with neuropeptide and hormone receptors.

Next, the PROF phylogeny was constructed (Figure 3.2). Two separate runs, each with 4 chains, satisfied all relevant measures of convergence after 20,000,000 generations. The eight reversible-jump Metropolis-coupled Markov chains Monte Carlo calculated poster probabilities from the sampling frequency to compare each of 9 fixed rate protein substitution matrices to find the most suitable model. Interestingly, both runs predominantly sampled from the Cprev model, which was created from plastid genomes [4], as it routinely provided the highest likelihood score. Wary of this result, we reran the analysis with a model created from the

transmembrane domains of Class A GPCRs [174] and also attempted to estimate from the data a variable rate general time reversible model. However, neither of these performed better than the Cprev model, raising questions about the suitability of the most widespread models, which typically use orthologous sequences from model organisms and tend to be biased with vertebrate sequences, for flatworm data. However, we found that the reversible-jump algorithm had higher sampling rates for suitable and intuitive models when dealing with full-length globular proteins from flatworms (data not shown), suggesting that our result with transmembrane sequence may be specific to our dataset.

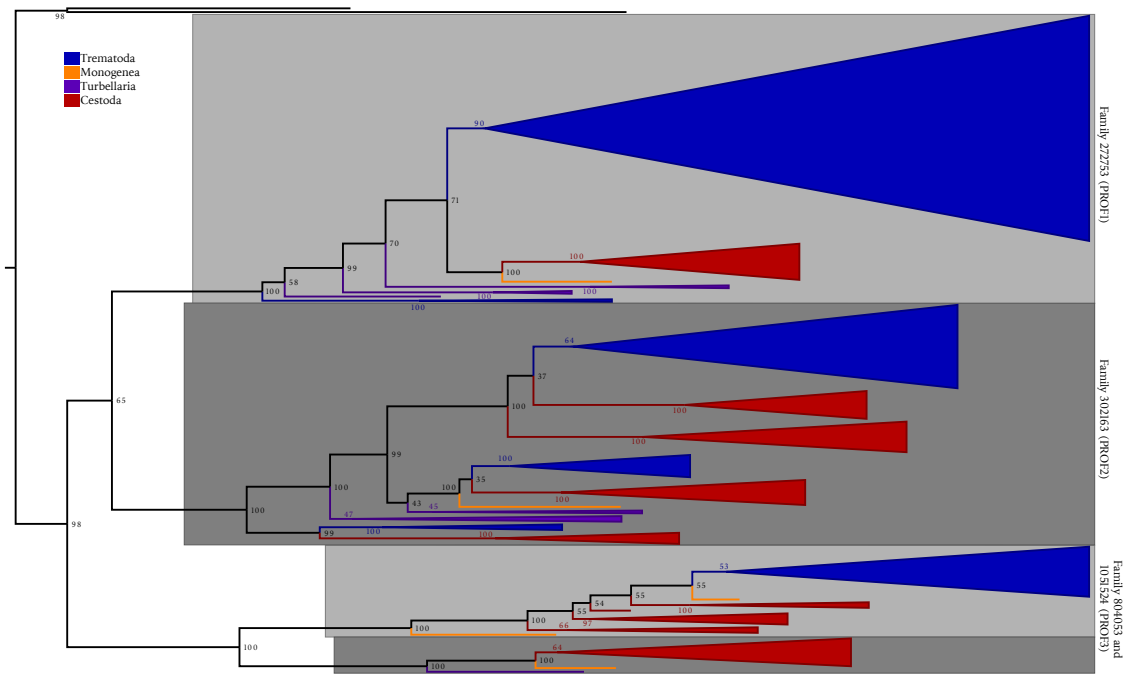

Figure 3.2: Complete phylogeny PROFs. Four Compara families (gray boxes) are classified into 3 PROF subfamilies. Cartoon-ized clades contain only genes from species of that particular flatworm class. In general, Turbellaria genes are basal to parasitic genes, following the species tree. Ectoparasitic Monogenea tend to be basal to the endoparasitic Trematoda and Cestoda, but sometimes cluster with Cestoda. These inconsistencies parallel the difficulty of confidently inferring the phylogeny of flatworms at the class level.

Regardless, the two runs were able to recapitulate each of the 4 families that were clustered by the Compara pipeline (Figure 3.2). Most of the nodes of this tree are highly supported

by the posterior probability (mean = 92) Notably, the least supported are those nodes that join different Compara families (bounded by gray boxes) and those joining genes from differing flatworm classes (red, blue, purple, and orange leaves). An exception, though, is the node that joins the two smallest families. This node has a posterior probability of 100%, indicating that these two families are probably a part of a larger individual gene family. Interestingly, when combined, this larger family contains one representative from each of the flatworms, mimicking the pattern of the other Compara families.

The convergence of the runs, along with the high support of the tree at-large, especially nodes within Compara-deduced families, compels us to hypothesize that these PROFs are actually three separate subfamilies, related by their common grouping into Class A (Rhodopsin-like) GPCRs. While it is tempting to infer that each family is activated by a separate ligand, only functional deorphanization and pharmacologic profiling can give such information. This conclusion is consistent with previous study of PROF phylogeny within *S. mansoni* and *S. mediterranea* [236]. Thus, we dub Compara Family 272753 as PROF1, Family 302163 as PROF2, and Families 804503 and 1051524 as PROF3a and PROF3b, respectively.

3.3.3 Expression Analyses

S. mediterranea was used as a model to analyze tissue expression of one PROF2 representative. Figure 3.3 shows robust mRNA expression in a 'tuning-fork' pattern along the cerebral ganglion and longitudinal nerve cords of the planarian nervous system. This pattern establishes PROF2 as an endoGPCR and is not activated by an exogenous chemical like chemosensory GPCRs. This is a significant discovery, as two recent bioinformatic GPCR surveys, using nematode chemoreceptors as bait, have annotated PROF receptors as chemosensory, calling into question the reliability of these annotations. [109, 184].

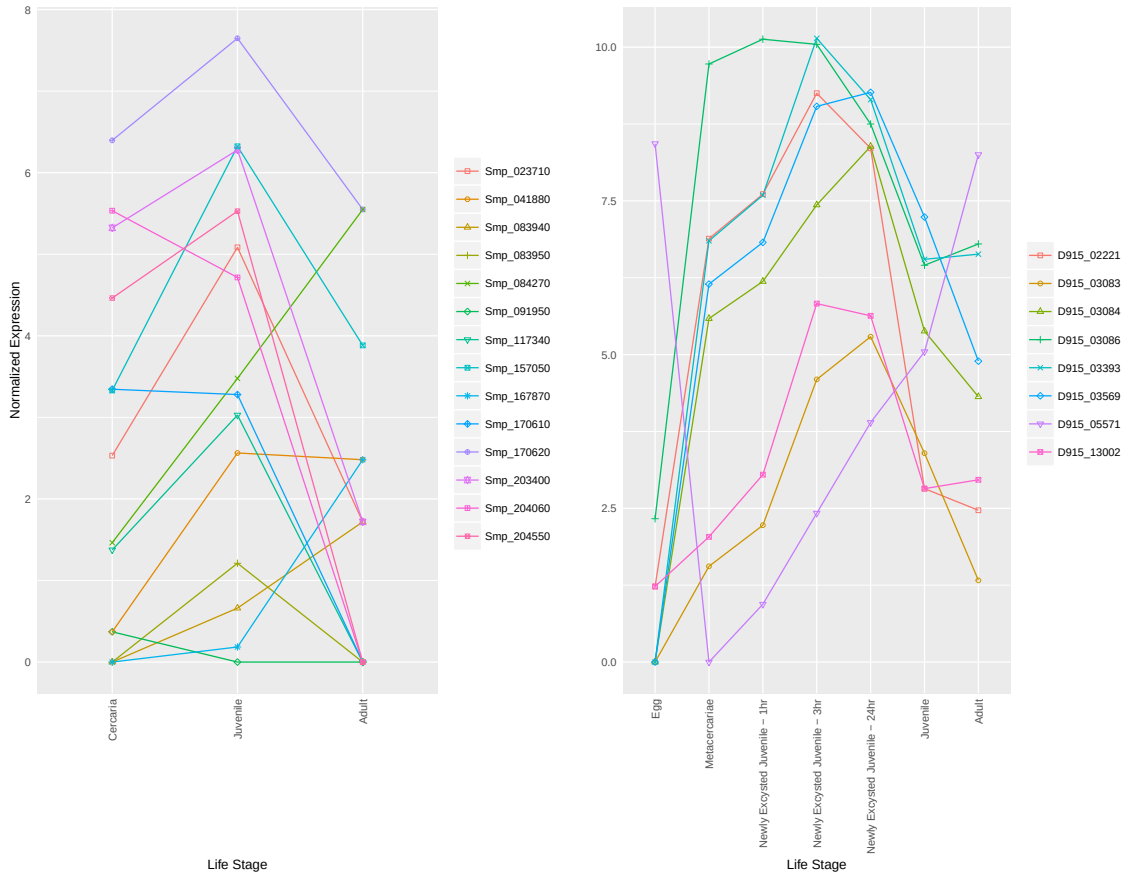

Figure 3.3: Relative expression of PROF complements in two parasitic flatworms. *S. mansoni* (left) and *F. hepatica* (right) demonstrate the general pattern of increased expression during juvenile life stages. NOTE: relative expressions should not be compared between organisms.

We also used available life stage-specific expression data from the trematodes *S. mansoni* and *F. hepatica* to create expression profiles for each of the PROF genes. While the plots (Figure 3.4) show some variability among the receptors, the general pattern of increased expression in the juvenile stages is immediately conspicuous. DESeq leveraged biological replicates to estimate the effective library size for each life stage, allowing stage-to-stage comparison and ensuring that this pattern is not simply due to differences in read depth or stage transcriptional activity. This pattern is notable, as one of the few flaws of praziquantel is its lack of efficacy against the juvenile stage (schistosomula) of schistosomes. Thus, it is desired that the next-generation of antischistosomals are effective in treating both juvenile and adult worm

stages.

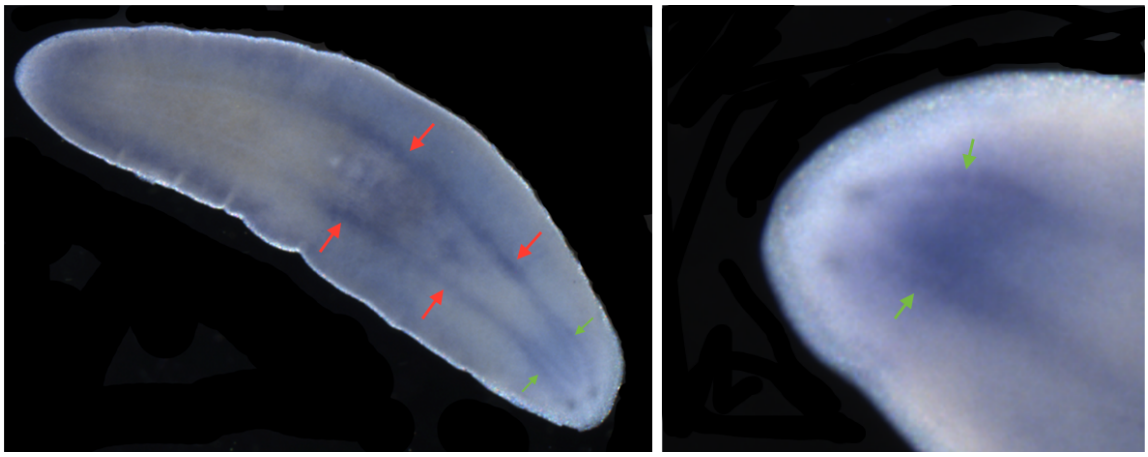

Figure 3.4: *In situ* hybridization of one PROF2 representative. (Left) The localization of a single PROF2 transcript is shown along the longitudinal nerve chords (red) and the cerebral ganglion (green) of *Schmidtea mediterranea*. (Right) The planarian head region reveals PROF2 expression in the cerebral ganglion (green).

3.4 Conclusion

We present here a comprehensive assemblage of all available data for the Platyhelminth-specific Rhodopsin-like Orphan Family. We have annotated a total of 284 protein coding genes in 25 flatworm organisms. We find that these genes are ubiquitous in all four Platyhelminth classes analyzed, suggesting importance to the phylum as a whole, and we note its expansion in Trematoda. We have established a high-confidence phylogeny for these genes and show that each of the three subfamilies is broadly conserved across the phylum and appeared early in flatworm speciation. Within *S. mansoni*, the PROF subfamilies are monophyletic within Class A (Rhodopsin-like) GPCRs, and they are most closely related to neuropeptide receptors. In situ hybridization in a free-living representative shows tissue-specific neurogenic expression. In two parasitic flukes, we show a pattern of increased expression in juvenile stages and adult stages, life stages that correspond to parasitism of humans.

Recently, two studies have used clustering, phylogenetics and a BLAST strategy to functionally annotate flatworm GPCRs [109, 184]. Both of these studies annotate PROF receptors as chemosensory receptors because of their apparent relatedness to nematode chemoreceptors. While the data presented here provides evidence that the PROFs may be neuropeptide receptors, and the tissue expression of one PROF2 representative implicates it as an endoGPCR, we are careful to show restraint in ascribing annotation without any supporting functional data. It may be that other PROF receptors are expressed in the tissue periphery and used in chemoreception. Flatworms certainly engage in complex chemosensation and, in effect, behavioral adaptation in accordance with sensitive and precise environmental cues. However, no chemosensory GPCR has yet been deorphanized in this phylum, and those presumptive chemoreceptors that flatworms do express are unlikely to be orthologous to the known nematode chemoreceptors.

Indeed, it is believed that diverse clades of 7 transmembrane-containing chemosensory receptors from vertebrates (olfactory, trace amine-associated, vomeronasal type 1 and 2, taste type 1 and 2), insects (olfactory, gustatory), and nematodes (19 separate families, including srw) arose independently after their respective ancestors evolved into land-dwelling organisms [15, 141], though insects and nematodes may still share one class of gustatory receptors [181]. Given that flatworms share no common land-dwelling ancestor with vertebrates, arthropods, and nematodes, it is unlikely that a flatworm chemoreceptor would be truly orthologous to chemoreceptors from these popular models. It is thus inappropriate to use chemoreceptor sequences from these diverse models as bait to search for or annotate flatworm chemoreceptors, and it is misleading to force flatworm sequences into a chemoreceptor phylogeny that does not include equal representation from all taxa involved in the analysis. Finally, the PROFs are scattered throughout the genome (Table 3.2) and do not mimic the genomic architecture of *C. elegans* chemoreceptors, which are all found clustered together on chromosome 5 [180]. Evolutionary information and life-history data such as these must be taken into account when attempting to annotate newly discovered (or newly sequenced) genes and genomes.

It has been hypothesized and supported in other taxonomies that orphan genes can be correlated with neofunctionalization [230], often in morphological or biochemical aspects, so

it is possible that the PROF receptors are involved in something particularly novel to the flatworm lineage. Furthermore, orphan GPCRs are the target of intense study in humans for their possible therapeutic exploitation [61, 202, 206]. The restriction of these receptors to flatworms and their expression in parasitic stages compels further interest in elucidation of their endogenous ligand(s), their biological function, and their study as drug targets.

3.5 Methods

3.5.1 Sequence Data

All predicted proteomes were acquired from Wormbase ParaSite. SRA files from the most recent *Fasciola hepatica* and *Schistosoma mansoni* transcriptomes (PRJNA179522 and PRJEA36577 [165]) were downloaded from the NCBI. Gene families inferred by the Compara pipeline were obtained from Matthew Berriman and Nancy Holyrod of the Wellcome Trust Sanger Institute.

3.5.2 GPCR Identification

Lists of seed sequences were populated by assembling known or predicted GPCR sequences from free-living and parasitic flatworms and nematodes. For the nematodes specifically, there was heavy reliance upon GO annotation and prediction. Sequences from *Schmidtea mediterranea, S. mansoni, Caenorhabditis elegans, Onchocerca volvulus*, and *Brugia malayi* were utilized for this initial list. While the analysis here deals specifically with platyhelminths, nematode sequences were included so that the full complement of helminth GPCRs could be identified from the data provided by the Sanger Institute. These seed accessions were then used to parse the Compara gene families provided, pulling out each family (and the accessions therein) that included at least one seed. Compara families contained a range of platyhelminth, nematode, and outgroup genes. An in-house Python script utilized these accessions to scroll through each helminths genome file and create FASTA files for each individual family.

Because most of the seed genes were uncurated, or at least had not yet been experimentally validated, and the putative GPCRs initially numbered over 10,000, we undertook a family-

based approach to supporting these genes as bona fide GPCRs. Family FASTA files were first aligned with MAFFT [99] (`mafft --auto family_number.fasta > family_number.aln`). Next, alignments were automatically trimmed to their most conserved sequence domains - likely the transmembrane regions [137] - with trimAL [35] (`trimal -automated1 -in family_number.aln -out family_number.trim.aln`). HH-suite 2.0 was used to build a Hidden Markov Model (HMM) for each family (`hhmake -i family_number.trim.aln -M -50`) and search it (`hhsearch -i family_number.hmm -d ~/hhsuite-2.0.16/database`) against clustered data-bases of HMMs from Pfam [62], SCOP [10], PDB [17], and UniProt. These databases were built using differing methodologies (see the HH-suite documentation), so use of each one ensured that GPCR identification was not overly dependent upon one database, especially one that prioritized mammalian sequences. The most significant hit from each of these databases were gathered, and each family was assigned a value (1-4) for GPCR support. Families that had GPCR hits for at least two of the databases were kept. Additionally, GPCRDB [89] class designation (Classes A, B, C and F) was assigned and validated for each family by randomly BLASTing 2-5 representative sequences against the NCBIs non-redundant translated nucleotide sequence database.

3.5.3 PROF Designation

Of the confirmed GPCR families, PROFs were nominated by (1) containing only platyhelminth sequences, (2) being Class A (Rhodopsin-like), and (3) being orphan receptors. Orphan status was confirmed by BLASTing 2-5 representative sequences against the NCBIs non-redundant translated nucleotide sequence database; these were most similar to either previously curated PROF genes or contained no consensus annotation within the significant hits (e-value 1e03).

3.5.4 Phylogenetics

We maintained the tradition of only using transmembrane sequences as phylogenetically informative [218]. Each GPCR or putative PROF sequence was trimmed to its transmembrane domains by identifying the domains with HMMTOP [215] and trimming it with an in-house Perl script. These trimmed sequences were aligned with MAFFT and subjected to Bayesian phy-

logenetic inference with MrBayes 3.2 [182]. Two reversible-jump Metropolis-coupled Markov Chain Monte Carlo (MCMC) runs, each with 4 chains sampling every 1000 generations, were run in parallel on the CyStorm high-performance computing system for 20,000,000 generations. The MCMC also sampled among all possible fixed-rate amino acid evolutionary models and used a gamma rate distribution with some invariable states (`prset aamodelpr=mixed; lset rates=invgamma`).

3.5.5 Expression Profiles

RNA-Seq reads were re-mapped to the most recent S. mansoni and F. hepatica genome and annotations available on WormBase ParaSite (BioProjects: PRJEA36577 and PRJNA179522; Assemblies: ASM23792v2 and F_hepatica_1.0.allpaths.pg; Genebuild version: 2014-05-WormBase). TopHat [213] was used to map RNA-Seq reads to these genomes, and featureCounts [124] was used to count the raw reads mapped to each gene. The R package of DESeq [8] was used to normalize count data and ggplot2 [229] was used to create line graphs for these data.

3.5.6 *In situ* Hybridization

Receptor gene fragments were PCR amplified with the minimal T7 polymerase promoter sequence appended to the 5 anti-sense primer. Digoxigenin-labeled antisense riboprobes were synthesized using these PCR products (Roche). Whole-mount *in situ* hybridization (WISH) was performed at 55C in hybridization solution (50% formamide, 5XSSC, 100 ug/ml yeast tRNA, 100 ug/ml heparin sodium salt, 0.1% Tween-20, 10 mM DTT, 10% dextran sulfate sodium salt). DIG-labeled riboprobe (40 ng/ml) was denatured at 72C for 15 min immediately prior to hybridization. BCIP/NBT was used for chromogenic color development, followed by paraformaldehyde fixation and imaging.

3.6 Additional Materials

Table 3.2: Genomic features of *S. mansoni* PROFs. Like most *S. mansoni* genes, the majority of the PROFs have a single exon. Unlike *C. elegans* chemoreceptors, PROFs are scattered throughout the genome; they are found on chromosomes 1, 3, 4, ZW and several unplaced scaffolds. The gene model for Smp_196080 is likely a fusion between a PROF receptor and a myosin protein.

Gene	Family	Locus	Exons
Smp_157050	family_1051524	Smp.SC_0313:39018-40310	2
Smp_084270	family_272753	Smp.SC_0677:2067-3176	1
Smp_091950	family_272753	Smp.Chr_3:21985382-21986482	1
Smp_167870	family_272753	Smp.SC_0111:252547-253623	1
Smp_177720	family_272753	Smp.Chr_3.unplaced.SC_0044:1913109-1914230	1
Smp_203400	family_272753	Smp.Chr_ZW:8222368-8223480	1
Smp_204060	family_272753	Smp.Chr_ZW:8409798-8410985	1
Smp_204550	family_272753	Smp.SC_0111:258383-259477	1
Smp_083950	family_272753	Smp.Chr_3.unplaced.SC_0044:1964528-1965808	2
Smp_083940	family_272753	Smp.Chr_3.unplaced.SC_0044:1934639-1963395	3
Smp_196080	family_272753	Smp.Chr_3:21988262-22007247	8
Smp_041880	family_302163	Smp.Chr_4:21927065-21928093	1
Smp_117340	family_302163	Smp.Chr_3:21718117-21719394	1
Smp_001070	family_302163	Smp.Chr_1:64165239-64230896	2
Smp_023710	family_302163	Smp.Chr_3:21772171-21773439	2
Smp_170620	family_804503	Smp.Chr_1:58945336-58948044	1
Smp_170610	family_804503	Smp.Chr_1:58871918-58873276	2

CHAPTER 4. A CASE FOR THE CAREFUL FUNCTIONAL ANNOTATION OF GPCRS IN PARASITIC WORMS

A paper to be submitted

Nicolas J Wheeler[1,2,3] Michael J Kimber[1], Mostafa Zamanian[4], Tim A Day[1,2,5]

4.1 Introduction

Parasitic worms of the phyla Nematoda and Platyhelminthes infect over one billion humans and devastate agricultural plants and livestock. The arrival of the genomic age and the subsequent sequencing of dozens of parasitic worm genomes have reinvigorated hope that target-based drug discovery may initiate a wave of next-generation anthelmintic development.

The rapid expansion of parasitic worm sequence data has largely outpaced our ability to make sense of the data. Reliable functional annotation of this abundance of newly discovered genes presents some significant challenges [73], and while misannotation has long been an issue [198, 25], it is exacerbated when new sequences are from the less studied phyla and branches of the tree of life. In these cases, experimental biologists are coerced to rely on progressively less informative data to make decisions and declarations about functional annotations of genes. This is hardly catastrophic; for the most part many genes can be accurately annotated with a reasonable degree of surety based on sequence homology, signature motifs, and the conservation of key functional domains.

Several tools make preliminary predictions of gene function by using similarity-driven al-

[1]Department of Biomedical Sciences, Iowa State University, Ames, IA USA
[2]Interdepartmental Genetics and Genomics, Iowa State University, Ames, IA USA
[3]Primary author
[4]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI USA
[5]Author for correspondence

gorithms that compare new sequences to accessions in curated databases. These approaches have limitations [96], especially for parasitic worms. A need for caution is compelled by the vast evolutionary distance between the phyla housing parasitic worms and the large bulk of well-characterized proteins at disposal for comparisons. While there are a significant number of functionally characterized proteins from organisms nearly related to parasitic nematodes, there are far fewer sources for platyhelminth parasites. In both cases, though, parasitologists are often obliged to shoehorn invertebrate data into mammalian or vertebrate annotation schema.

Concerns about the accuracy of sequence-based preliminary annotations are more pronounced with G protein-coupled receptors (GPCRs). Sequence-based comparisons do not reliably predict the ligands that will activate GPCRs, or the downstream signaling systems that they will activate, or the host of other important functional characteristics that are often the primary points of inquiry. These kinds of functional classifications are very difficult to parse in vivo, especially in many of the least tractable parasites, so preliminary predictions based solely on primary sequence data are often accepted. Here, we describe the confounding nature of these biologically important, potentially vulnerable parasite proteins, and we suggest a cautious, more holistic approach to thinking about their functional annotation.

4.2 Parasitic Worm GPCRs

Parasitic worm GPCR complements cohere reasonably well with either of the major classification systems that have emerged in attempts to make sense of the GPCR superfamily.

- Attwood and Findlay [14] and Kolakowsk [105] used transmembrane sequence patterns to classify or fingerprint GPCRs into groups labelled A-F. This system is incorporated by GPCRDB [90] and has been used to describe GPCRs in a parasitic worm [33].

- Schith and Fredriksson inferred the phylogeny of human GPCRs and showed that they segregate into five families that they labelled GRAFS. They also demonstrated the suitability of this system for other animals, vertebrates and invertebrates alike [63]. We and others have also used this system to classify the complements of parasitic worm GPCRs [236, 214].

Classification of parasitic worm GPCRs into the large A-F or GRAFS families is relatively simple, and automated pipelines do so accurately. However, these two systems depart from each other significantly when it comes to sub-family categorization, and automated approaches can become unreliable at this level. These sub-family categories attempt to align with functional characteristics, such as the specific ligands that activate the GPCRs in a certain sub-family, so it is notable that they are confounding for parasitic worm GPCR contingents. For example, the Rhodopsin-like group (the largest class of parasitic worm GPCRs) includes receptors that respond to a wide range of ligands, ranging from photons to peptides, making ligand prediction an arduous task. In contrast, Glutamate-like and Frizzled-like receptors respond to a much more limited spectrum of ligands [239, 88, 205]. Thus, for the majority of parasitic worm GPCRs, classification is only broadly useful and does not aid in precise functional annotation that is desired.

4.3 The Challenge

The enterprise of confidently annotating GPCRs *in silico* is complicated by three distinct qualities. First, outside the 7 canonical transmembrane (TM) regions, GPCRs contain very few definitive functional domains. Those that are known to be present are found in a subset of receptors in the non-Rhodopsin groups, which comprise a small fraction of worm GPCRs [236, 33]. Second, the Rhodopsin-like GPCRs have diverse and diffuse ligand-binding domains. For small ligands in particular, most of their cognate receptors binding residues are in the core of the transmembrane regions and dispersed throughout [111]. Even if ligand binding residues of a receptor are known, it is challenging to extrapolate these interactions to homologous receptors [133]. Third, while phylogenetics can place a novel GPCR upon an evolutionary backdrop, a complete phylogeny does not differentiate between gene duplications that have lead to neo- or sub-functionalization, and it cannot capture the idiosyncratic differences that may distinguish between receptors of similar sequence but divergent functional traits. The TM domains and the ligand-binding residues therein experience unique selection pressures [200], which should be accounted for when making evolutionary assumptions of functional relatedness. This likely contributes to the observation that divergent receptors showing little sequence similarity can

respond to the same ligand and, contrastingly, homologous receptors with high similarity can respond to different ligands [125].

The convoluted and opaque evolutionary history of the GPCR superfamily further impedes functional annotation via phylogenetics. Phylogenetic techniques assume that the sequences included are homologous. Whether or not the main families of GPCRs are descended from a single common ancestor is still a question, though recent evidence suggests that the Rhodopsin-like, Frizzled-like, Glutamate-like, and Adhesion-like families descended from a 7 TM gene in a eukaryotic common ancestor [108]. Nevertheless, convergent evolution of GPCRs is also widespread [54, 82, 156, 116, 122]. These complications - the ancient genesis of the superfamily, the lack of sequence-similarity between GPCR families, and reports of GPCR convergent evolution - can cause GPCR phylogenies to be misinferred and can impede functional annotation from them.

In an attempt to circumvent these issues, one might capitalize upon newer, innovative analyses that use similarity-driven clustering in multidimensional space [160, 92, 185]. These approaches are not as constrained by models of molecular evolution and can be preferable when dealing with large protein families, sets of gene models that have differing levels of confidence, or datasets with non-homologous genes. Parasitic worm GPCRs have these characteristics, but because of the confounding sequence traits described above, more granular inferences - those focused on protein function - can be spurious. This is particularly important when dealing with highly diverged GPCRs or datasets that include a diverse range of species.

Both phylogenetic and clustering approaches have the potential to produce inconsistent inferences about receptor function. For example, a large grouping of flatworm receptors (Platyhelminth-specific Rhodopsin-like Orphan Family; PROF) was initially discovered and annotated, using a phylogenetic approach, as a Rhodopsin family outgroup [236] with the hypothesis that they may be neuropeptide-responsive. Since then, gene models for *Schistosoma mansoni* have been improved with RNA-Seq [165], and a more sensitive Bayesian inference supports the original topology (Figure 1). In contrast, recent reports have annotated members of this receptor grouping as chemoreceptors based on sequence similarity to the nematode srw family of chemoreceptors [185, 109].

In the absence of experimental data, neither *in silico* approach is sufficient for confidently annotating the function of these receptors. The evolution of volatile compound chemosensation occurred quickly after the migration of water dwelling animals to the land. This functional adaptation happened independently in ancestors of nematodes and arthropods [15], after the split between Lophotrochozoa (flatworms) and Ecdysozoa (nematodes). Thus, even if the PROF and srw clades have an ancient relation, it is unsuitable, based on the phylogeny or sequence-similarity alone, to transfer organismal function from one to the other. Whether these receptors have true chemosensory or neuropeptidergic (or other) function remains to be seen, but the PROF case study demonstrates how diverse methodology can lead to incompatible inferences, and, critically, how molecular and pharmacologic experiments are necessary for true functional annotation of parasitic worm GPCRs.

4.4 Difficult and Important Cases

Most of what is known about GPCRs comes from experimentation in mammalian systems. When a parasitic worm GPCR is closely related to a mammalian receptor, similarity methods may be able to inform functional annotation. However, for many receptors from these less-studied taxa, reliance upon mammalian comparisons is less dependable. We will consider two types of GPCRs as examples to illustrate this weakness.

4.4.1 Taxonomically-Restricted GPCRs

As more nematode and flatworm genomes are sequenced, researchers are able to use comparative genomics to identify phylum, clade, class, or life-style specificity that may confer import traits involved in host-parasite interactions [112]. This is new frontier for parasitologists, as previous work has only been able postulate as to the specificity of gene families of interest. We are now able to confidently ascribe taxonomic-restriction, and we predict that taxonomically-restricted gene families will be and should be of great interest in the future, especially with GPCRs.

A recent comparative analysis of parasitic worm GPCRs found that over 60 of the 200 GPCR families identified are constrained to worms and were not found in any outgroups including

humans, zebrafish, oysters, flies, and others. Furthermore, at time of writing, the current release of WormBase ParaSite [85] (Version: WBPS7) contains 11,315 protein-coding genes with predicted GPCR activity (GO:0004930) in worms without an orthologue in flies, human, mouse, budding yeast, or zebrafish. These taxonomically-restricted receptors are of great interest and ought to receive increasing attention, but similarity-based analyses provide no clear direction to inform functional annotation, especially for the platyhelminth receptors.

4.4.2 Orphan GPCRs

Many of these taxonomically-restricted GPCRs are orphans. Like taxonomically-restricted GPCRs, functional annotation of orphans is also troublesome, and they are too often broadly and conclusively annotated in literature or databases without any molecular, biochemical, or pharmacologic evidence. As described, primary sequence can enable classification of these orphan receptors, but it is mostly impotent when seeking to draw more decisive and sensitive conclusions. In these cases, one is left directionless.

Few worm receptors have been definitively deorphanized *in vivo* – that is, few have been demonstrably paired with an endogenous ligand. In general, the first GPCRs in parasitic worms to be pharmacologically deorphanized are those for which measures of sequence similarity provide high-likelihood hypotheses that point to a particular ligand. As these conserved GPCRs continue to be cloned and profiled, the divergent, orphan GPCRs will remain, for which sequence similarity provides little direction. In *C. elegans*, there is evidence that the preliminary predictions by sequence similarity fail and mislead ligand identification [135], and this is likely to be true for more divergent orphans in parasitic worms as well. Indeed, the sequence databases allude toward this result. For instance, genes from 26 flatworms in ParaSite WormBase are annotated as dro/myosuppressin receptors. Myosuppresin, a peptide ligand, has never been described or isolated in either nematodes or flatworms, and this annotation comes from the extrapolation of work performed in fruit flies. *Ancylostoma duodenale* and *Romanomermis culicivorax*, two parasitic nematodes, also have Putative dro/myosuppressin receptors, and these annotations are inferred by orthology to a *S. mansoni* protein. This annotation had made vast evolutionary leaps from Arthropoda to Platyhelminthes, and from Platyhelminthes to Ne-

matoda. Such examples are ubiquitous across the functional annotations of parasitic worm genomes. Of course, most know to be wary of unsupervised automated annotating pipelines, but these examples manifest the impropriety of using the underlying logic - sequence similarity - to annotate parasitic worm GPCRs. Preliminary annotations can be useful indeed, but caution has to be taken when evaluating the predicted ligands of orphans.

4.4.3 Related Difficulties

These examples are not to suggest that *in silico* studies are worthless. Rather, studies of GPCR similarity, relatedness, homology, and putative function are all important, but the conclusions of each are not always correlated, and we need to avoid drawing bold arrows from one to the other. That is, homologous receptors are not always obviously similar, and receptors that are obviously similar do not always share a common function [197]. When *in silico* studies are undertaken, it is important that the bounds and limitations of each approach are made clear and that restraint is shown in their conclusions.

Computational identification and preliminary annotation should be supplemented with experimentation that probes the expanse of different GPCR characteristics. This work is difficult but necessary in order to fully appreciate the purpose and function of parasitic worm GPCRs and their possible exploitation in chemotherapy. Below, we briefly describe five main attributes and associated caveats to be considered when seeking to a newly identified parasitic worm GPCR.

4.4.3.1 Identification of ligand(s)

A GPCRs ligand is arguably the most important piece of functional data that can be determined, and much of the GPCR literature deals with receptor deorphanization, the process of pairing receptors with ligands. There are several caveats that should be considered in this endeavor. For example, there are examples of ligand-independent and constitutive GPCR signaling [167] that potentially extend to parasitic worms [130]. GPCRs can also form homo- and heteromeric complexes [196, 203], making identification and delineation of the endogenous ligands non-trivial. For peptidergic receptors, receptor promiscuity is likely widespread in

parasitic worms [138], and *in vitro* sensitivity to a molecule may not correlate with ligand responsivity in the native environment [150]. Despite these warnings, the vast majority of GPCRs have at least one ligand of biological relevance, and receptor deorphanization should be one of the first aims when annotating a newly identified GPCR.

4.4.3.2 Spatial expression

Knowledge of the spatial distribution of receptor expression can provide clues about receptor signaling and physiology. Animals dictate GPCR expression patterns with cellular [23, 78] and subcellular specificity [42], and receptors are likely to be expressed in the same tissues or in proximity to the signals they receive. For instance, GPCRs in chemosensory organs or cell types are likely to be chemoreceptors [23, 51] and GPCRs at the neural-muscle interface may receive one of the many known chemical signals that are involved in muscle control [159]. Learning the cells, tissues, and organs in which GPCRs are expressed can aid interpretation of native and *in vitro* GPCR function, as well as the appraisal of specific GPCRs as parasite drug targets.

4.4.3.3 Temporal expression

It is also essential to know the parasite life stages in which GPCRs are expressed and accessible to potential ligands. In terms of therapeutic potential, it is necessary that the targeted receptor be expressed in the intra-host life stages amenable to pharmacologic intervention [134]. Expression data can include both transcript and protein abundances, but they are not necessarily linked. Measuring transcript levels at certain points in time does not indicate that the receptor is being translated or trafficked to the membrane. Receptor internalization, turnover, and recycling must be considerations when drawing inferences from these expression data.

4.4.3.4 Downstream signaling pathways

Canonical GPCR signaling progresses through activation of the Gα subunit, but it is now clear that there are a variety of signaling cascades. Typically, the C-terminus of the receptor directly interacts with one of a number of different Gαs, leading to a specific downstream result. Alternative signaling through G$\beta\gamma$ was demonstrated early on[128], and other non-

G protein mediated routes have been discovered, particularly through β-arrestin [119]. For parasitic worms, less is known about the repertoire of G proteins and their function, although their structure largely resembles that of mammalian G proteins.

4.4.3.5 Phenotype

Understanding the biological functions mediated by these GPCRs is usually the ultimate goal. Those biological functions are generally probed by the association of particular phenotypes with GPCR loss of function. However, in parasites there are limitations in the available genetic tools, and it is difficult to relate the few, limited phenotypes available in the laboratory to the breadth of biological functions that are important to parasites throughout their complex life-cycle. For example, one needs to be especially wary of RNAi phenotypes, as some believe that many behavioral phenotypes, especially a decrease in general motility of whole organisms, could be due to general sickness rather than acute inhibition of a single transcript [68, 103]. Availability of a only a few, limited phenotypes can induce us to ascribe unwarranted importance to any detected function, often to the exclusion of other possible functions for which we simply have no phenotypic probe. Nonetheless, when properly employed, RNAi can be a powerful annotation tool and can reveal relevant biological functions. As the experimental toolkit advances, it will be helpful to be able to create knockdowns, or even knockouts, in specific tissues or cell-types, and progress would be greatly facilitated by the development of more relevant phenotypic assays in parasitic worms [235].

4.5 Conclusion

There are a variety of approaches to gather this information, but these annotations have only been studied in a limited number of GPCRs in helminths (Table 4.1). Classification of receptors of the non-Rhodopsin groups (Glutamate, Adhesion, Secretin, and Frizzled) will also inform some of the other canonical GPCR attributes. However, for the Rhodopsin-like family of parasitic worm GPCRs, classification, sequence-similarity, and homology can give clues toward some functional attributes, but alone these threads are insufficient and will provide an incomplete, and even misleading, picture. Unfortunately, there is momentum in the literature

to turn many of these clues into hard and fast functional annotations, and, as demonstrated, these annotations have waterfall effects when provided to pipelines that continually annotate new genomes.

Table 4.1: Summary of all experimental results studying GPCRs of parasitic worms.

		Deorphanization						
Name	Species	Ligand	Cell Type	*In Vitro* G protein coupling	Reporter of GPCR Activity	Localization	Stage of Confirmed Expression	RNAi Phenotype
SmGPR-1 [77]	*Schistosoma mansoni*	Histamine	HEK293, COS7	Native G proteins	Ca^{2+} accumulation		Cercaria, higher in young juveniles, adults	
SmGPR-2 [57]	*S. mansoni*	Histamine	*Saccharomyces cerevisiae* strain YEX108	Chimeric Gαq	Growth	Nervous system; subtegumental neuronal plexus of adult and larvae	Cercaria, higher in young juveniles, adults	
SmGPR-3 [58]	*S. mansoni*	Dopamine	*S. cerevisiae* strain YEX108	Chimeric Gαq	Growth	Nervous system; main nerve cords and the peripheral innervation of the body wall muscles		
SmGAR-1 [130]	*S. mansoni*	Acetylcholine	*S. cerevisiae* strain Cy13393	Chimeric Gαi	Growth			Hypoactive larvae
Sm5HTR [159]	*S. mansoni*	Serotonin	HEK293	Native G proteins	cAMP stimulation	Nervous system, including the cerebral ganglia and main nerve cords and the peripheral innervation of the body wall muscles and tegument		Hypoactive juveniles and adults
SmGluR [205]	*S. mansoni*	Glutamate	HEK293	Native G proteins	cAMP stimulation	Longitudinal nerve cords and cerebral commissures; peripheral nerve fibers and plexuses innervating the acetabulum and the somatic musculature; length of the female reproductive system	Cercaria, higher in young juveniles, adults	
SmD2 [204]	*S. mansoni*	Dopamine	HEK293, *S. cerevisiae* strain Cy13393	Native G proteins, Chimeric Gαs	cAMP stimulation	Subtegumental somatic musculature and acetabulum of all larval stages tested; enriched in the somatic muscles of adults and the muscular lining of the caecum		
AsGAR-1 [102]	*Ascaris suum*	Acetylcholine	*S. cerevisiae* strain YEX108	Chimeric Gαq	Growth	Head and tail		
As5HTR2 [87]	*A. suum*	Serotonin, LSD	COS7, HEK293	Native G proteins	Radioligand binding; PI turnover	Pharynx and body wall muscle		
GpFLP-32 [13]	*Globodera pallida*							Increase in migration rate
BmNPR-4 [9]	*Brugia malayi*	FLP-18 family	HEK293	Native Gαi proteins	cAMP attenuation			
Bm4 [194]	*B. malayi*	Tyramine, LSD	HEK293	Native G proteins	Radioligand binding			
Hc110-R [186, 139]	*Haemonchus contortus*	α-latrotoxin (LTX), AF1, AF10, PF2; PF1022A and emodepside (antagonists)	HEK293	Native G proteins	Ca^{2+} accumulation			
Hc5HTR1 [195]	*H. contortus*	Serotonin	Sf9, AV12-Ga_{15}	Promiscuous $G\alpha_{15}$	Radioligand binding; Ca^{2+} accumulation		L1/L2/L3, adult	

The preponderance of the tools leveraged to functionally annotate parasitic worm GPCRs rely upon sequence similarity to known receptors. Unfortunately, it is becoming clear that the resulting inferences are not as extensible as once was hoped. Outside of similarity-driven analyses, other comparative *in silico* approaches that make use of genomic synteny will be more useful for annotation as parasitic worm genomes progress closer to chromosome-level assemblies [7]. Until then, given the current difficulties, we suggest that parasitologists commit to a more precise experimental toolset for studying and annotating GPCRs. This may seem prosaic, but it will be worthwhile to take time to sensitively functionally annotate individual receptors rather than add to the already overwhelming dataset of GPCRs of unknown function. Collaboration and teamwork is key, and a commitment by the parasitology community to caution and rigor will have important outcomes on the continuing functional annotation of diverse parasite genomes.

CHAPTER 5. SUMMARY AND DISCUSSION

There are four conceptual trajectories toward anthelmintic discovery and development (Table 1.1). The high cost of screening infected animals (Trajectory #1) limits its applicability, which has demanded the innovation of new techniques that will also produce improved understanding of flatworm biology.

Here, we generated a *de novo* transcriptome for a species of planaria, *Dugesia tigrina*. This transcriptome will prove to be a valuable resource for the basic study of planarian and flatworm biology as well as more translational approaches that are oriented toward anthelmintic discovery. As a proof of this claim, the transcriptome was used in a comparative analysis that innovated a new pipeline to prioritize putative drug targets that are shared within the flatworm phylum. By comparing the *D. tigrina* transcripts to genes from a blood fluke *Schistosoma mansoni* and a tapeworm *Echinococcus multilocularis*, filters were constructed to remove unshared genes and genes that are unlikely to be worthy of further study as targets. Putative targets that survived these filters were then compared to human genes, with the working hypothesis that toxicity in humans could be avoided by choosing targets not conserved in the host. Of course, there are some drugs for infectious diseases that do not support this hypothesis - the benizimadizoles being a prime example (see chapter 1). Nevertheless, comparison to host genes remains a popular filter in other subtractive and comparative genomic approaches [55, 207].

After a set of putative targets conserved amongst flatworms were pooled, several experiments were performed to demonstrate that phenotypes caused by pharmacologic inhibition or transcript knock-down in planaria were reproducible in schistosomes. Pharmacologic inhibition of the actin binding complex (ARP2/3) resulted in reduced motility in both planaria and schistosomes (Figure 2.5A), and gene specific knock-down in planaria impaired regenerative capacity and resulted in eventual death (Figure 2.6B). The chemical used to inhibit ARP2/3,

CK-666, is commercially available and unlikely to be a good lead because its inhibiting action is efficacious in all sorts of organisms [147, 175]. Instead, these experiments serve to demonstrate how debilitating phenotypes in planaria are often reproduced in schistosomes. The contrast was shown to be true as well for the shared protein NADH dehydrogenase. The NADH dehydrogenase inhibitor rotenone showed no effect on either planaria or schistosomes (Figure 2.5C), and knock-down of NADH dehydrogenase also showed no phenotype (Figure 2.6A).

Thus, chapter 2 contributed to two separate anthelmintic discovery trajectories (Figure 5.1). It demonstrated the utility of *D. tigrina* as a platyhelminth model for *in vivo* screens, and it demonstrated a method for putative drug target prioritization using genetic comparisons. This work also showed how both trajectories can provide valuable biological conclusions. As stated, the *D. tigrina* transcriptome is a valuable resource in and of itself, not only for potential discovery of novel anthelmintics; its comparison to other flatworm or metazoan sequences could also be useful in evolution and ecology, toxicology, regeneration, and other sorts of studies. In addition, the RNAi and pharmacologic inhibition experiments for ARPC2 and SDH1 (Figure 2.5B) provided insight into the biological functions of these two proteins.

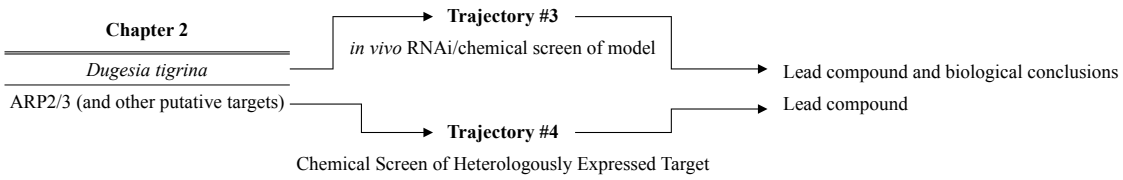

Figure 5.1: Objectives of Chapter 2. Chapter 2 demonstrates how focus on anthelmintic development can lead to translational lead compounds as well as biological conclusions.

Systematic RNAi screens are already underway in the planaria *Schmidtea mediterranea*, with particular focus given to regeneration [179, 221]. *D. tigrina* could be an added resource in this pursuit, one that could possibly enable a higher throughput because of their more reliable hardiness in lab cultures. Diversifying the species of planaria used in screens could also provide hits that would have been otherwise missed when using only *S. mediterranea*. Finally, these screens ought not solely be performed with focus given to stem cells and regeneration, but

phenotypes that are classically associated with anthelmintic mechanisms should be considered as well, such as membrane depolarization, motility, death, or neural dysfunction [39, 37].

Reducing the size of container needed for screening, allowing for increased throughput, would be a helpful development for this platform. In the assays described here, planaria were individually observed in 35 mm Petri dishes, with a maximum of 6 dishes per experiment. This is a far too large container for screening with any sort of throughput. It is unlikely that the high-throughput phenotyping machinery used for screening *C. elegans* will ever be matched for planaria, but work could be performed to adapt planaria culture techniques that are amenable for 96 well plates. Immmortalizing a planarian cell line has not yet been accomplished, but an individual worm can be cut into hundreds of pieces that could be sorted into plates [142]. New robotic cell imaging and computer vision approaches could track tissue regeneration and survival, looking for similar effects to that seen in ARPC2 RNAi experiments (Figure 2.6C).

In an alternative to the unbiased filtering and pooling of potentially druggable proteins, one can identify potential proteins for trajectory 4 by hand-selecting superfamilies that are known to be druggable. We emphasize G protein-coupled receptors in this approach because of their proven druggability in other systems [152]. As plasma membrane receptors, GPCRs are easily accessed by small molecule chemicals, and GPCRs have many critical biological functions in platyhelminths. Indeed, emodepside, a drug used to treat gastrointestinal nematode parasites, was originally shown to have latrophilin, a Class B GPCR, as its target [186, 79]. However, further experiments implicated a potassium channel, SLO-1, as the more important target [75]. The differing pathways and their cross-talk are still being worked out, but it is clear that attenuation of latrophilin signaling is an important part of emodepside's activity. Outside of this one anecdote, no other GPCR is targeted by marketed anthelmintics.

It is often misconceived that GPCRs act solely as modulators in platyhelminths, with actions too slow to be a good drug target. This is often contrasted with fast acting ion channels, and it is argued that this is the reason that ion channels are the targets of the vast majority of marketed anthelmintics. However, GPCRs have important neuromuscular roles, and their activation can have immediate and strong effect on these systems [159, 53, 52]. Even so, it is unclear why GPCRs have yet to be confirmed as an anthelmintic target when they seem

so suitable for other types of chemotherapeutics. Because each of the anthelmintics to target ion channels (pyrantel, derquantel, levamisole, ivermectin, monepantel, and emodepside, for example) were discovered through screens of infected animals, it is possible that these screens are not equipped to identify other types of targets. If so, new approaches may validate GPCRs and other proteins as true anthelmintic targets.

In a collaboration with the parasite genomics group at the Wellcome Trust Sanger Institute, a census was taken for the GPCRs of 81 different flatworms and nematodes, including many parasites (Appendix A), and the GPCR complement of each organism was compared (Figure A.1). The constructed heatmap allows for comparisons between clades, classes, species, and phyla, and it reveals a number of interesting patterns. The Platyhelminth-Specific Rhodopsin-Like Orphan Family (PROF, see chapter 3) was striking in this analysis. When first identified in *Schmidtea mediterranea* and *Schistosoma mansoni*, it was hypothesized to be a putative drug target that deserved further analysis on the basis that it 1) was a GPCR, 2) was a large (and therefore important) clade of GPCRs within these two flatworms, 3) phylogenetically clustered with putative neuropeptide receptors, and 3) showed evidence of being taxonomically-restricted to the flatworms. However, at the time, it was inconclusive if this lineage of GPCRs was expanded in all flatworms, if it indeed had neuropeptidergic capacity, and if it was actually phylum-specific.

The sequencing of 25 flatworm genomes in the 50HGI allowed for a broader analysis of the PROF (chapter 3). Through this analysis, it was shown that the PROF is actually the largest subfamily of GPCRs found in platyhelminths. In fact, the PROF is the largest phylum-specific gene family overall. It was also confirmed that the PROF is indeed expanded throughout the phylum (Table 3.1) and that one of its members is expressed in neural tissue (Figure 3.4). Moreover, publicly available stage-specific RNA-seq datasets show that many PROF members are highly expressed in the juvenile stages of *Fasciola hepatica* and *S. mansoni*, and some are expressed throughout the life cycle (Figure 3.3).

After all this, can we now confirm that PROF receptors are good candidates for heterologous expression and screening? Unfortunately, the fact remains that the ligands of any of the PROF receptors are still unknown. Because of this, there would be no positive control for PROF

expression in heterologous systems, which would make it impossible to find small molecule inhibitors for a PROF. This brings up an important note. Although screening of heterologously expressed targets will not provide new information regarding the biological function of targets within the worm (Table 1.2), the only targets that ought to be moved through this trajectory are those for which a good deal of biology is already known. For a receptor, knowledge of the endogenous ligand is necessary to validate it as a screening candidate. Though comparative genomics and filtering pipelines can highlight proteins that may be druggable, these targets are not good candidates for heterologous expression and screening unless their function is confirmed. In this tangential way, trajectory 4 can also further our understanding of basic flatworm biology (Figure 5.2). For the majority of flatworm proteins that do not have well studied orthologs, precise experimental functional annotation must be performed before embarking on mechanistic-based screening protocols. This is of immeasurable importance to the success of the screen, but it also provides important, novel understanding of flatworm biology.

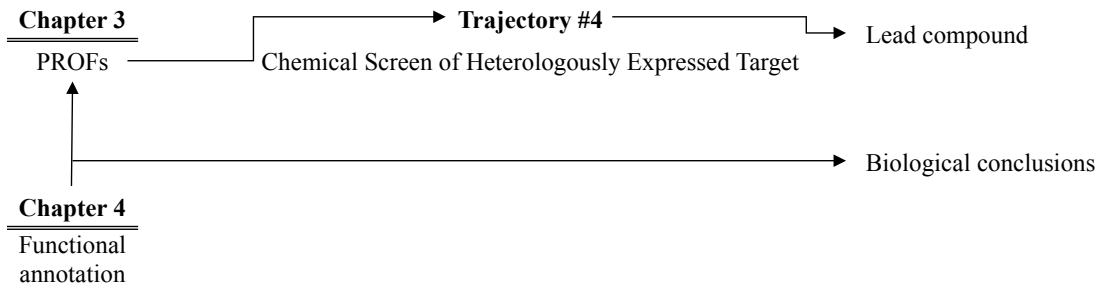

Figure 5.2: Objectives of Chapters 3 and 4. Chapters 3 and 4 demonstrate how focus on anthelmintic development can lead to translational lead compounds as well as biological conclusions.

Unfortunately, functional annotation of new platyhelminth genes tends to be imprecise, especially for GPCRs (chapter 4). For the most confounding (and most interesting) GPCRs, those that are taxonomically-restricted, automatic annotations performed *in silico* are essentially empty. From both basic and applied research perspectives, it is absolutely critical to be confident in the functional annotations that are provided in literature and online databases. For this reason, chapter 4 called for a recommitment to careful functional annotation of GPCRs,

especially those that are good drug targets. Care and caution in functional annotation will be crucial not only for the development of new anthelmintics, but for the wholistic understanding of the biology of flatworms. This same care needs to be given to the PROF receptors before they are screened for anthelmintics.

A recommitment to a more nuanced approach to the functional annotation of GPCRs should hearken back to the thrust of chapter 2. Parasite GPCRs are difficult to study; parasites themselves are intractable to transgenesis, difficult and costly to culture, and they have few phenotypic assays by which to probe protein function. These difficulties are exasperated for GPCRs that are taxonomically-restricted to the flatworms, like the PROFs. The obvious answer to this conundrum is to use the planarian model to study these receptors.

RNAi in planaria is straightforward and cheap, and knock-down experiments have short turnaround times. As mentioned, there are many more phenotypic assays available for planaria than parasitic flatworms, making it more likely that a knock-down experiment would give positive results. A comprehensive RNAi screen of planaria GPCRs would be a reasonable next step for this work.

For aminergic receptors, this approach would undoubtedly provide a wealth of information regarding their function. There are still multiple monoamines that are synthesized in flatworms that have yet to be linked to their cognate receptor; tyramine and octopamine are the most prominent of these orphan ligands. Deorphanization of these could be accomplished by establishing phenotypes for exogenous tyramine or octopamine treatment, and then attempting to ablate that phenotype by iterative RNAi knock-down of the hypothesized receptors. This approach has been used for a planarian serotonin receptor, but it has yet to be leveraged for any other flatworm GPCR [234]. After confirmation in planaria, these experiments could then be repeated in parasitic flatworms, using the sequence of the confirmed planaria receptor to point to the likely homolog in the parasite.

However, for PROF receptors and the other putative neuropeptide receptors, it is possible that redundancy could mask any RNAi phenotypes. Some neuropeptides are believed to have pleiotropic effects in these systems [5], and the expansion and high sequence conservation among the PROF receptors hints toward this redundancy. This could be potentially confounding for

RNAi screens, and it could cause false negatives for the previously described deorphanization approach by "rescuing" an ablation phenotype through functional redundancy. In this case, degenerate dsRNA constructs could act to knock-down several PROF receptors at once, but a high dose of dsRNA and a buildup of digested nucleotides may result in toxicity that would give a falsely positive result. Thus, knock-down approaches are unlikely to be informative for the PROF and other putative neuropeptide receptors. In cases such as these, the planarian model may fall short.

Heterologous expression of GPCRs and screening with potential endogenous ligands (instead of potential anthelmintics) is the best way to match ligands and receptors. This is difficult in cell line expression systems like yeast, insect, or mammalian cells, because flatworm receptors often run into problems with proper folding, membrane trafficking, or signal transduction. For these experiments, there is no positive control, and it is time consuming and expensive to confirm membrane expression by raising antibodies against the receptor. Expression vectors could be designed with tags, but it is possible that a GFP or His tag, for instance, would interfere with wild type ligand recognition or downstream signaling. If a first-pass screen fails to provide any hits, the logical next step is unclear. In a departure from traditional heterologous expression in cell lines, perhaps there are other more suitable systems. Mammalian receptors have been expressed in a tissue-specific manner in *C. elegans* [209, 188], and parasitic nematode GPCRs have been screened in this way [117]. It is possible that this system would be amenable for flatworm receptors as well. Further, with the advent of precise genome editing tools like CRISPR-Cas9 [47], potentially masking homologous endogenous receptors could be removed and replaced with the flatworm receptor of interest. If co-transfected with a genetically-encoded calcium or cAMP reporter (see [225] or [157] for reviews), this would be a robust platform for the deorphanization of flatworm GPCRs.

5.0.1 Conclusion

Together, these chapters form a cohesive unit that develops and comments upon modern techniques for anthelmintic development, but it also serves as a reminder that traditional biological experimentation remains the foundation of this field. The genomic age enables novel

trajectories toward drug discovery while simultaneously reasserting a demand to more attentive biological engagement with these animals. To advance understanding of these important parasites, there must be the sort of multidisciplinary engagement that is demonstrated here.

APPENDIX A. THE 50 HELMINTHS GENOME INITIATIVE (50HGI)

A paper to be submitted

Matthew Berriman[1] et al., (Nicolas J Wheeler[2,3], Tim A Day[2,3])

This appendix includes selected sections for which I had primary responsibility in both the analysis and textual summary.

Summary

GPCR Analysis

A comparative approach to identifying helminth GPCRs uncovered a number of fascinating data (Figure A.1). First, the massive radiation of chemosensory GPCRs that occurred in *C. elegans* is unmatched by any other nematode species. Indeed, over 87% of the *C. elegans* GPCR complement is comprised of chemosensory GPCRs. In contrast, the next largest chemosensory fraction occurs in the hookworm *A. ceylanicum*, comprising 30% of the total GPCR set. All examined parasitic nematodes possess chemosensory receptors, with the greatest numerical representation apparent in Clade V parasites. Although fewer chemoreceptors are found in Clade I, III, and IV parasites, those present represent receptor subtypes that are near completely conserved across all parasites in the phylum. These include homologs of *C. elegans* daf-37, which is known to mediate ascaroside signaling, consistent with the importance of this pheromone signaling pathway in the likely evolutionary transition from free-living dauer larvae to infective larvae, the so-called dauer hypothesis [49, 145].

[1]Parasite Genomics, Wellcome Trust Sanger Institute, Hinxton, England
[2]Department of Biomedical Sciences, Iowa State University, Ames, IA USA
[3]Interdepartmental Genetics and Genomics, Iowa State University, Ames, IA USA

There are distinctive differences between the GPCR complements of filarial nematodes and non-filarial nematodes. Filarial nematodes (family Filarioidea) are phylogenetically diverged from their counterparts and paraphyletic to the rest of Clade III. *O. volvulus, L. loa, B. malayi*, and *L. sigmodontis* show an overall decrease in GPCR families and, on average, possess about 40% fewer GPCRs than their clade-mates.

One interesting contrast can be found in examination of the GPCR complements of Clade I nematodes. It is seen that *Trichinella spiralis* and *Trichuris muris* have very similarly reduced GPCR complements when compared to the non-filarial nematodes. In contrast *Romanomermis culcivorax* appears to contain a more standard GPCR complement. In the *R. culcivorax* genome paper [189], the authors performed a similar clustering analysis to the one performed here and also noticed distinct differences within the proteomes of Clade I species. However, while they surmised that *T. spiralis* is not representative of the rest of the clade, the GPCR analysis here seems to suggest that it is actually *R. culicivorax* that stands out.

In regards to platyhelminths, it is seen that these worms express far fewer GPCRs than nematodes overall, though many families are conserved in both phyla. There are, however, a number of platyhelminth-specific groups, highlighted by the PROFs (Platyhelminth-specific Rhodopsin-like Orphan Families) [236]. These receptors are Class A receptors and hypothesized to be peptide responsive, but they do not show any significant homology to any annotated or deorphanized proteins. Additionally, there are several families that are specific to flukes, showing a cladistic GPCR expansion.

The two liver flukes *C. sinensis* and *F. hepatica* seem to lack any homolog to the classical secretin receptor (Class B). Class B receptors are predominantly peptide hormone receptors (glucagon receptor and parathyroid hormone receptor, for instance). Class B peptide hormone receptors were found in more detailed searches of the *S. mansoni* and *S. mediterranea* genomes [236], and these were recapitulated in this analysis, so it is interesting to discover that this class of GPCRs has completely disappeared in the non-schistosome Digenea.

Interestingly, no chemosensory GPCRs are readily identifiable in platyhelminth species. Given that flatworms, like nematodes, must navigate diverse environmental cues, it is likely that these species do in fact engage in sophisticated forms of chemosensation. We hypothesize

that either some subset of identified phylum-specific orphan receptors are involved in such processes, or that chemosensation in flatworm parasites is not primarily driven by GPCRs.

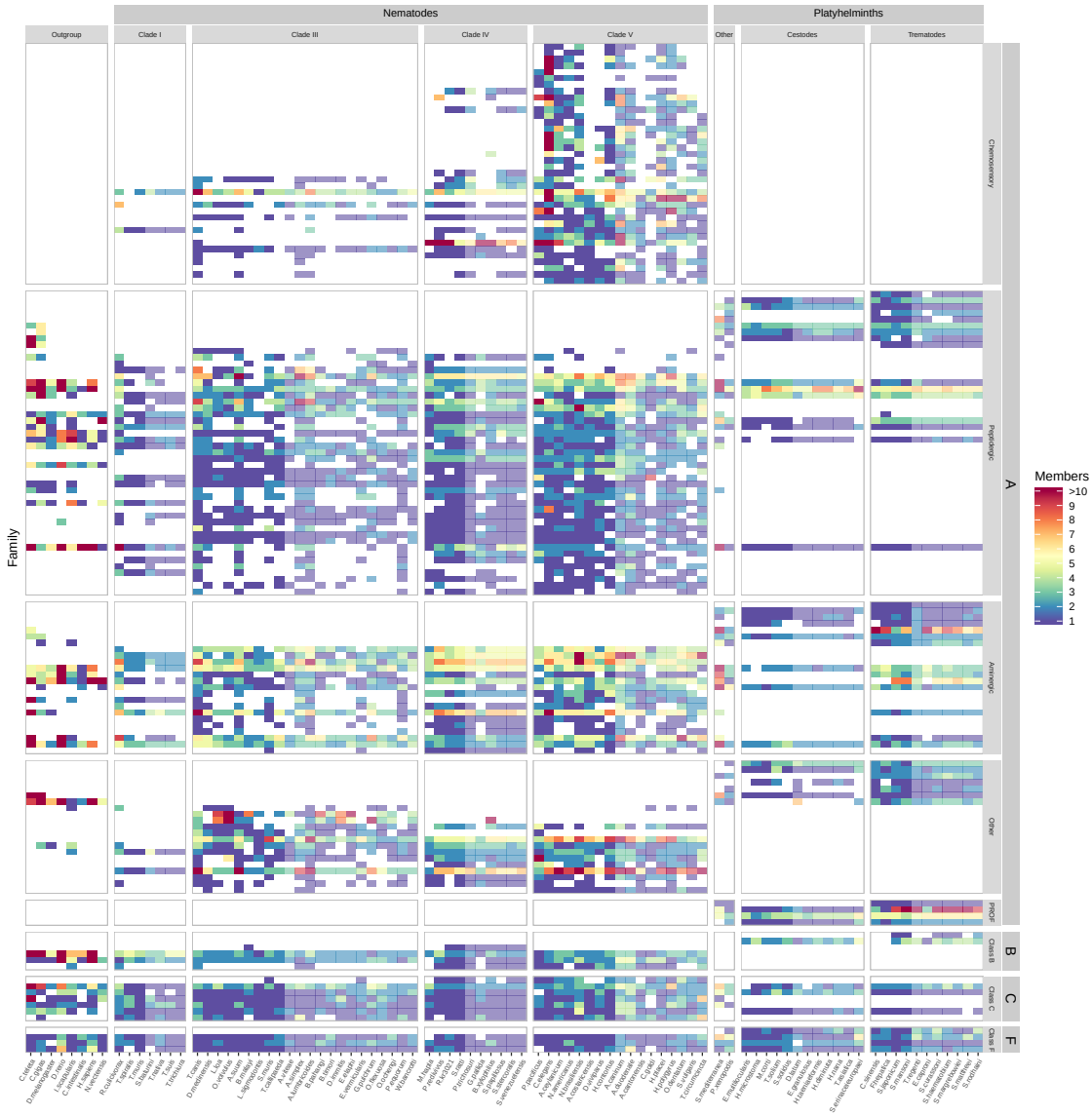

Figure A.1: Comparison of GPCR families for 81 nematode and platyhelminth species. Nematodes (left) are organized by clade and platyhelminths (right) are organized by class. Each row signifies an individual Compara family from the Ensembl comparative genomics pipeline.

Cys-Loop Ligand-Gated Ion Channel Analysis

Several intriguing features can be seen in the comparative analysis of cys-loop receptors in nematodes and platyhelminths (Figure A.2). All nematodes analyzed contain multiple subunits from the four previously described subfamilies of nicotinic acetylcholine receptor channels (nAChRs; DEG-3, ACR-16, UNC-29, and ACR-8/UNC-38). In contrast, flatworms contain a more restricted distribution that varies by class UNC-29-like and ACR-8/UNC-38-like for cestodes and ACR-16-like for trematodes. These phyla also contain a previously described flatworm-enriched family of divergent nicotinic acetylcholine anion channels that is not homologous to nematode ACCs, and we have expanded this family to additional flatworms. We confirm that flatworms only contain one other family of anion channels - the GluCls of the AVR-14 family.

While much is known about the cys-loop superfamily in nematodes, this group of important proteins remains rich in unannotated and unstudied channels. This is particularly true of flatworms, and the identification of these receptors will pave the way for *in vivo* and *in vitro* studies to more fully appreciate their diverse and essential functions.

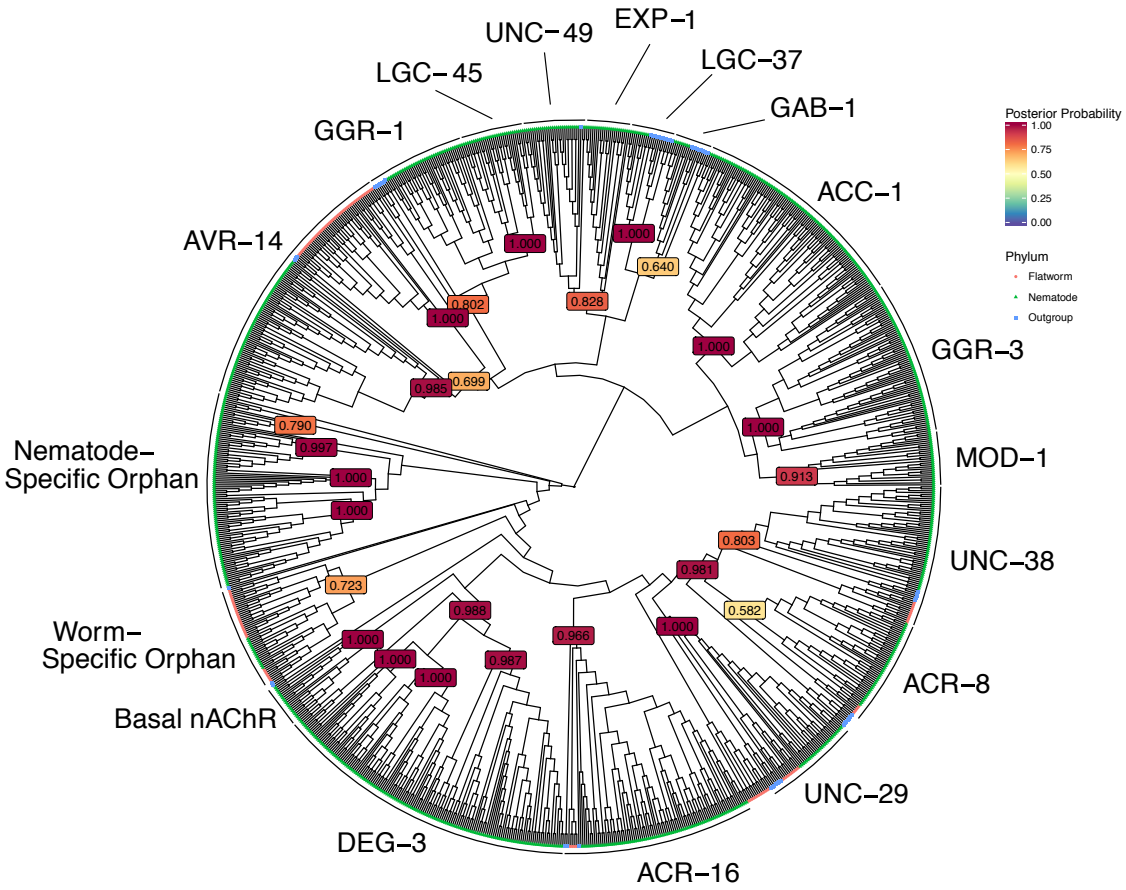

Figure A.2: The phylogeny of cys-loop receptors in nematodes and platyhelmiths. MrBayes was used to infer the phylogeny of the cys-loop superfamily in flatworms and nematodes. Posterior probabilities were calculated from 8 reversible jump MCMC chains. Tree is rooted between nicotinic acetylcholine receptors and non-nicotinic anion channels. Node posterior probability labels correspond to heatmap rows in Figure A.3. List of species included is on x-axis of Figure A.3.

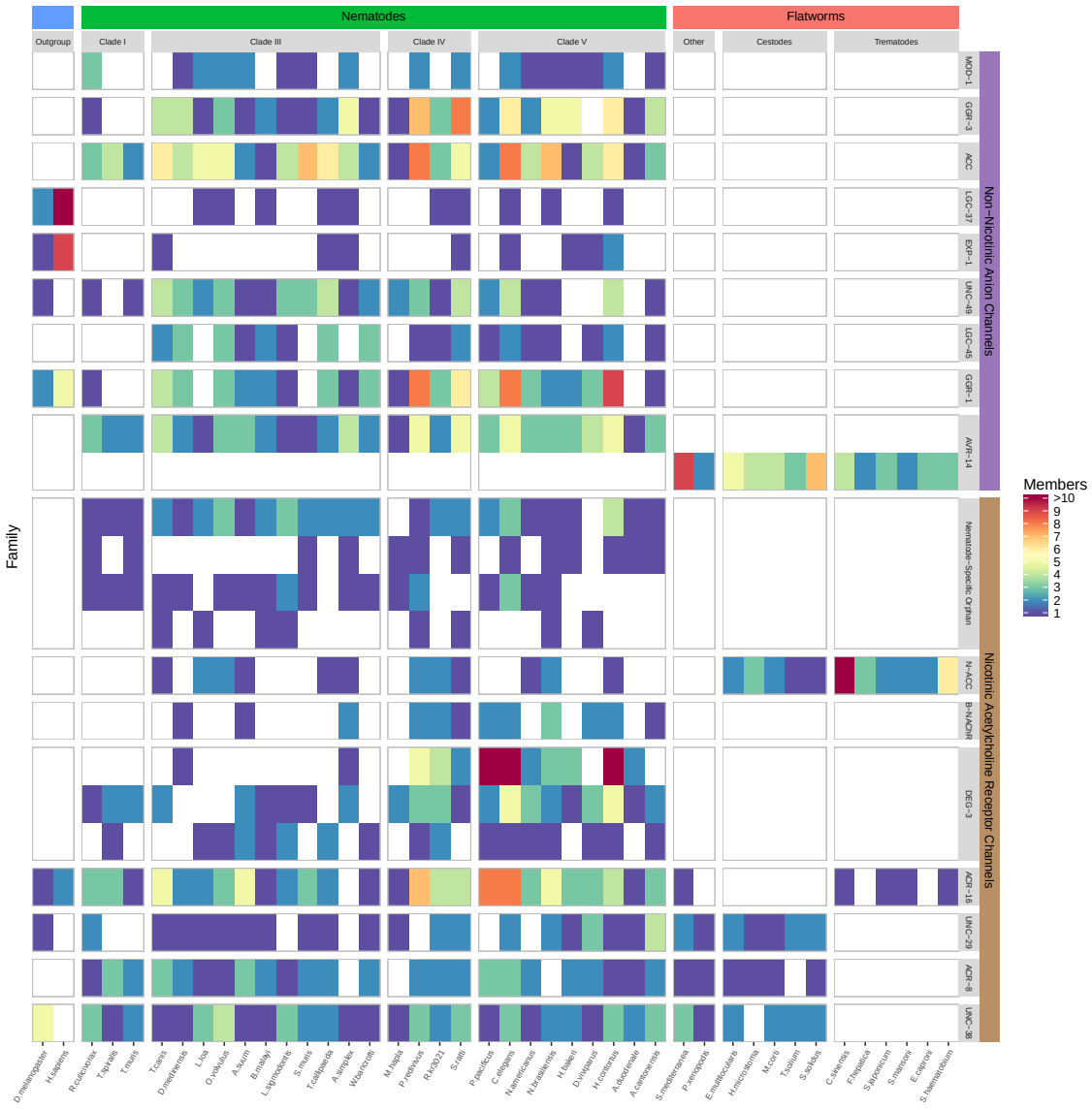

Figure A.3: Comparison of cys-loop families for 81 nematode and platyhelminth species. Families were extracted from the cys-loop phylogeny based upon posterior probability of clades. Taxa from each clade were extracted, tallied, and organized in a heatmap.

Methods

Known, annotated GPCRs from *C. elegans, B. malayi, O. volvulus, S. mansoni*, and *S. mediterranea* were identified from literature mining and previous GO annotations, and these

were used as seeds for extraction of GPCR families created by the Compara pipeline. This list of GPCR seeds included almost 2000 accessions: 1339 from *C. elegans*, 58 from *B. malayi*, 84 from *O. volvulus*, 150 from *S. mansoni*, and 343 from *S. mediterranea*. It was presumed that the vast majority of the *bona fide* GPCR families could be identified from this set of seeds, although species-specific, or possibly even clade-specific, duplications would have been missed. These seeds were used to parse the Compara families, searching for any family that contained at least one seed. A cursory examination of these putative GPCR families revealed a substantial amount of false-positives, likely due to imprecise GO annotations by which the seeds were curated.

A family-centric algorithm was then used to filter out false-positives and assign GPCR class information. The sequence for each family member was extracted and families were aligned with MAFFT [99] (`mafft --auto`), alignments were trimmed with trimal1.4 [35] (`trimal -automated1`) to remove uninformative sites. Finally, HHSuite [171] was used to build an HMM for each family and search (hhsearch) against databases of HMMs created by HHSuites clustering algorithm (Uniprot [1], SCOP [10], Pfam [62], and PDB [17]). This strategy focused on the conserved element of each family, likely the element that initially compelled Compara clustering. The approach output best-hit information for each family against each of the 4 databases. Families supported as a GPCR by at least 2 databases were retained for downstream analyses. To ensure proper designation and to assign classification, 2-5 random members from each putative GPCR family was used in blastp searches against the non-redundant protein database at the NCBI. All of the best-hit information was used to assign each family to one of the GPCRDB classes (Class A, B, C, and F).

A similar approach was used for the cys-loop receptors. Known cys-loop receptor accession IDs from *C. elegans* [94], *Brugia malayi* [120], *Haemonchus contortus* [27], *Oesophagostomum dentatum* [27], and *Schistosoma mansoni* [129] were gathered and used to parse Compara families. The resulting families were concatenated to a superfamily, aligned with MAFFT [99] (`mafft --thread 4 --maxiterate 1000 --localpair fasta > alignment`), and trimmed with trimAl [35] (`trimal -in alignment -out trimmed_alignment -gt 0.85 -cons 2`).

A combinatorial approach was used to the infer the phylogeny of the superfamily. First, a maximum-likelihood (ML) consensus tree was produced with RAxML (`/raxmlHPC-PTHREADS-SSE3 -T 8 -x 12345 -p 12345 -U -# 100 -m PROTCATAUTO -s alignment -n name`) by plotting bootstrap values from 100 replicates onto the ML tree (`/raxmlHPC-PTHREADS-SSE3 -f T -T 4 -p 12345 -m PROTCATAUTO -s alignment -n bestTree`). To filter out taxa that reduced the nodal support, RogueNaRok [2] was used to remove unstable rogue taxa. The subsequent reduced alignment was then used to infer the phylogeny with MrBayes3.2 [182]. Two simultaneous runs, each with 4 reversible-jump MCMC chains, were allowed to proceed for 20,000,000 generations. The chain was given a temperature of 0.05 and took samples every 1000 generations. The parameters below were provided:

- `prset shapepr=exponential(0.05);`
- `prset aamodelpr=mixed;`
- `lset rates=invgamma.`

A 50% majority rule consensus tree that included all compatible groups was created from the resulting 40,000 trees (`sumt contype=allcompat`), and this tree was visualized and annotated with ggtree [233] (see Appendix B for a sample of the script used to process the tree). Cys-loop subfamilies were redefined by selecting nodes with high posterior probability and using the annotations of the original cys-loop seeds. A census of the tips of each subfamily was taken and visualized in a heatmap with the ggplots2 [229] package within the R [83] statistical language.

APPENDIX B. SELECTED PYTHON AND R SCRIPTS

Python Parser

```
#!/usr/bin/python

with open("seeds.txt", "r") as data:
search_terms = data.read().splitlines()

missing_terms = set(search_terms)

with open("families.txt", "r") as db, open("output.txt", "w") as output:
    for line in db:
        for term in search_terms:
            if term in line:
                missing_terms.discard(term)
                next_line = db.next()
                output.write(">" + head + "\n" + next_line)
                print("Found {}".format(term))
                break

if missing_terms:
    diagnose_not_found(missing_terms)
```

Tree Plot

```
1  library("ggplot2")
2  library("ggtree")
3  library("ape")
4  library("phytools")
5  library("phylobase")
6  library("geiger")
7  library(dplyr)
8  library(reshape2)
9  library(tidyr)
10 library(RColorBrewer)
11 library(stringr)
12
13 palette <- colorRampPalette(rev(brewer.pal(11, "Spectral")), space = "Lab")
14
15 setwd("~/Box_Sync/Projects/50HGI/Wheeler/Tree")
16 bayes_file <- "cl-rj.con.tre"
17
18 bayes <- ape::read.nexus(bayes_file)
19 bayes <- bayes[[1]]
20
21 #load in reference file matching gene id <-> species <-> family
22 reference <- read.csv("reference.csv", header = FALSE)
23 colnames(reference) <- c("Gene_ID", "Species", "Family")
24
25 #load in reference file matching species <-> clade
26 reference2 <- read.csv("clade_species_phylum.csv", header = FALSE)
27 colnames(reference2) <- c("Species", "Clade", "Phylum")
28
29 #gene_id, species, family, clade
30 reference <- merge(reference,reference2,by="Species")
31
32 #add species name to tip label, format:
     Species."-".Gene_ID."-".Family."-".Clade."-".Phylum
33 tip_labels <- as.data.frame(bayes[]$tip.label)
```

```
34 colnames(tip_labels) <- c("tip_label")
35 tip_labels <- tip_labels %>%
36   tidyr::separate(tip_label, c("Gene_ID", "Family"), sep = "-", remove = TRUE)
     %>%
37   select(-Family)
38 reference <- reference %>%
39   filter(Gene_ID %in% tip_labels$Gene_ID) %>%
40   group_by(Gene_ID) %>% distinct(.keep_all=TRUE)
41 tip_labels$id  <- 1:nrow(tip_labels)
42 tip_labels <- merge(tip_labels,reference,by="Gene_ID") %>%
43   arrange(id) %>% select(-id)
44 tip_labels$final_label <-
    paste0(tip_labels$Species,"-",tip_labels$Gene_ID,"-",tip_labels$Family,"-",tip_labels$Clade,"
45
46 #replace the original tip labels
47 bayes[]$tip.label <- tip_labels$final_label
48
49 #basic tree with node labels
50 #ggtree(bayes, branch.length = "none", layout = "circular") +
    geom_text2(aes(subset=!isTip, label=node), hjust=-.3) + geom_tiplab2()
51 #reroot based on anion/cation / replot
52 bayes <- phytools::reroot(bayes, 1252)
53 #Create a df containing internal node numbers and probs, as well as terminal
    node numbers and NA (for labels)
54 prob <- dplyr::filter(as.data.frame(bayes), isTip  == FALSE)
55 prob <- dplyr::select(prob, node, label)
56 prob$label[prob$label == "Root"] <- NA
57 node <- c(1:1122)
58 label <- c(rep(NA, 1122))
59 prob2 <- data.frame(node, label)
60 prob <- dplyr::bind_rows(prob, prob2)
61 colnames(prob) <- c("node", "prob")
62
63
64
```

```
65 ##############################
66 #create entire ggtree object
67 ##############################
68
69 t <- ggtree(bayes,branch.length = "none", layout = "circular")
70 t <- ggtree::rotate(t, 1127) %>% ggtree::rotate(1129)
71
72 #add columns to t$data object to include all metadata
73 df <- tip_labels %>%
74   rename(label = final_label)
75 t_data <- t$data
76 t_data_replace <- merge(t_data,df,by="label",all=TRUE) %>%
77   arrange(node)
78 t$data <- t_data_replace
79
80 #add support values to nodes (for manual selection of nodes)
81 #tbs <- t2 %<+% prob + geom_label2(aes(subset=!is.na(prob), label = prob, fill =
     prob))
82 #selected nodes based on support values
83 node <- c(lgc45 = 1130, mod1 = 1504, ggr3 = 1429, acc = 1530, exp1a = 1409,
     lgc37 = 1388, unc49 = 1335, ggr1 = 1127, avr14 = 1206, avr14a = 1207, avr14b =
     1258, nsoa = 1643, nsob = 2222, nsoc = 2210, nsod = 2166, wso = 1654, bna =
     1992, deg3 = 2012, deg3a = 2067, deg3b = 2021, deg3c = 2087, acr16 = 1707,
     unc29 = 1938, au838 = 1803, acr8 = 1874, unc38 = 1805)
84
85 #plot entire tree
86 t_all <- t +
87   geom_cladelabel(node=1429, label="GGR-3", fontsize=8, offset.text=1) +
88   geom_cladelabel(node=1504, label="MOD-1", fontsize = 8, offset.text = 1) +
89   geom_cladelabel(node=1530, label="ACC-1", fontsize=8, offset.text=1) +
90   geom_cladelabel(node=1336, label="EXP-1", fontsize=8, offset.text=1) +
91   geom_cladelabel(node=1409, label="GAB-1", fontsize=8, offset.text=1) +
92   geom_cladelabel(node=1388, label="LGC-37", fontsize=8, offset.text=1) +
93   geom_cladelabel(node=1357, label="UNC-49", fontsize=8, offset.text=1) +
94   geom_cladelabel(node=1130, label="LGC-45", fontsize=8, offset.text=2) +
```

```
95    geom_strip(509, 343, label = "GGR-1", fontsize = 8, offset.text = 4) +
96    geom_strip(54, 351, label="AVR-14", fontsize=8, offset.text=7) +
97    geom_strip(1118, 1102, label = "Nematode-Specific Orphan", offset.text = 20,
        fontsize = 8) +
98    geom_strip(540, 1025, label = "Nicotinic ACC", offset.text = 13, fontsize = 8)
        +
99    geom_cladelabel(node=1992, label="Basal nAChR", fontsize=8, offset.text=5) +
100   geom_cladelabel(node=2012, label="DEG-3", fontsize=8, offset.text=3) +
101   geom_cladelabel(node=1707, label="ACR-16", fontsize=8, offset.text=1) +
102   geom_cladelabel(node=1932, label="UNC-29", fontsize=8, offset.text=1) +
103   geom_cladelabel(node=1805, label="UNC-38", fontsize=8, offset.text=1) +
104   geom_cladelabel(node=1874, label="ACR-8", fontsize=8, offset.text=1) +
105   geom_tippoint(aes(shape=Phylum, color=Phylum), size = 0.9)
106 #Manually annotate: "Nicotinic Acetylcholine Receptor Channels"
107
108 #select which probs to show (nodes picked above)
109 prob3 <- prob[prob$node %in% node,]
110
111 t_all <- t_all %<+% prob3 + geom_label2(aes(na.rm=TRUE, label = prob, fill =
      as.numeric(prob))) +
112   scale_fill_gradientn(colours = palette(11), name="Posterior Probability",
        limits=c(0,1)) +
113   theme(legend.position="right")
114 t_all
115
116 ############################
117 #create ggtree subtree object
118 ############################
119
120 t <- ggtree(bayes)
121 t <- ggtree::rotate(t, 1127) %>% ggtree::rotate(1129)
122
123 #add columns to t$data object to include all metadata
124 df <- tip_labels %>%
125   rename(label = final_label)
```

```
126 t_data <- t$data
127 t_data_replace <- merge(t_data, df, by="label", all=TRUE) %>%
128   arrange(node)
129 t$data <- t_data_replace
130
131 #add support values to nodes (for manual selection of nodes)
132 #tbs <- t2 %<+% prob + geom_label2(aes(subset=!is.na(prob), label = prob, fill =
      prob))
133 #selected nodes based on support values
134 node <- c(lgc45 = 1130, mod1 = 1504, ggr3 = 1429, acc = 1530, exp1a = 1409,
      lgc37 = 1388, unc49 = 1335, ggr1 = 1127, avr14 = 1206, avr14a = 1207, avr14b =
      1258, nsoa = 1643, nsob = 2222, nsoc = 2210, nsod = 2166, wso = 1654, bna =
      1992, deg3 = 2012, deg3a = 2067, deg3b = 2021, deg3c = 2087, acr16 = 1707,
      unc29 = 1938, au838 = 1803, acr8 = 1874, unc38 = 1805)
135
136 #plot subtree based on node
137 node_select <- 1206
138 t_clade <- t %>%
139   viewClade(node=node_select) +
140   geom_tippoint(aes(shape=Phylum, color=Clade), size = 2) +
141   geom_text2(aes(subset=!isTip, label = label), size = 1) +
142   #geom_text(aes(label=Species), hjust=-.1, size=3) +
143   geom_text(aes(label=label), hjust=-.1, size=3) +
144   theme(legend.position="right")
145 t_clade
146
147 #Pull out nodes (and color by Clade and label with species)
148 #haemonchus_contortus glc-1? HCOI00130400 family_1584651 Ve Nematoda
149 #we need avr-15
150 #avr-14 (glu-cls)
151 #expansion within nematodes after clade I
152
153 #get clades for heatmap
154 des <- list()
155 n <- 1
```

```
156 for(x in node) {
157   y <- geiger::tips(bayes, x)
158   des[n] <- list(y)
159   n <- n + 1
160 }
161 names(des) <- names(node)
162
163 len <- sapply(des, length)
164 n <- max(len)
165 len <- n - len
166 csv <- mapply(function(x,y) c( x , rep( NA , y ) ) , des , len )
167 csv.m <- melt(csv, na.rm = TRUE)
168
169 write.csv(csv.m, file="clades.csv")
```

REFERENCES

[1] The universal protein resource (UniProt). *Nucleic acids research*, 36(Database issue):D190–5, jan 2008.

[2] Andre J Aberer, Denis Krompass, and Alexandros Stamatakis. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic biology*, 62(1):162–6, jan 2013.

[3] Josep F Abril, Francesc Cebrià, Gustavo Rodríguez-Esteban, Thomas Horn, Susanna Fraguas, Beatriz Calvo, Kerstin Bartscherer, and Emili Saló. Smed454 dataset: unravelling the transcriptome of Schmidtea mediterranea. *BMC Genomics*, 11(1):731, dec 2010.

[4] J Adachi, P J Waddell, W Martin, and M Hasegawa. Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *Journal of molecular evolution*, 50(4):348–58, apr 2000.

[5] Mariana Laura Adami, Cristina Damborenea, and Jorge Rafael Ronderos. Expression of a neuropeptide similar to allatotropin in free living turbellaria (platyhelminthes). *Tissue and Cell*, 43(6):377–383, 2011.

[6] C. Adamidi, Y. Wang, D. Gruen, G. Mastrobuoni, X. You, D. Tolle, M. Dodt, S. D. Mackowiak, A. Gogol-Doering, P. Oenal, A. Rybak, E. Ross, A. S. Alvarado, S. Kempa, C. Dieterich, N. Rajewsky, and W. Chen. De novo assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics. *Genome Research*, 21(7):1193–1200, jul 2011.

[7] Et al Yun S. Prevertebrate Local Gene Duplication Facilitated Expan-

sion of the Neuropeptide {GPCR} Superfamily. - {PubMed} - {NCBI}. \url{https://www.ncbi.nlm.nih.gov/pubmed/26337547}, 2015.

[8] Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106, jan 2010.

[9] Ross C. Anderson, Claire L. Newton, Robert P. Millar, and Arieh A. Katz. The Brugia malayi neuropeptide receptor-4 is activated by FMRFamide-like peptides and signals via Gαi. 2014.

[10] Antonina Andreeva, Dave Howorth, John-Marc Chandonia, Steven E Brenner, Tim J P Hubbard, Cyrus Chothia, and Alexey G Murzin. Data growth and its impact on the SCOP database: new developments. *Nucleic acids research*, 36(Database issue):D419–25, jan 2008.

[11] P Andrews. A summary of the efficacy of praziquantel against schistosomes in animal experiments and notes on its mode of action. *Arzneimittel-Forschung*, 31(3a):538–41, 1981.

[12] P ANDREWS. Praziquantel: mechanisms of anti-schistosomal activity. *Pharmacology & Therapeutics*, 29(1):129–156, 1985.

[13] Louise E Atkinson, Michael Stevenson, Ciaran J McCoy, Nikki J Marks, Colin Fleming, Mostafa Zamanian, Tim A Day, Michael J Kimber, Aaron G Maule, and Angela Mousley. flp-32 Ligand/receptor silencing phenocopy faster plant pathogenic nematodes. *PLoS pathogens*, 9(2):e1003169, feb 2013.

[14] T K Attwood and J B Findlay. Fingerprinting G-protein-coupled receptors. *Protein engineering*, 7(2):195–203, feb 1994.

[15] Cornelia I Bargmann. Comparative chemosensation from receptors to ecology. *Nature*, 444(7117):295–301, nov 2006.

[16] Richard Benton, Silke Sachse, Stephen W Michnick, and Leslie B Vosshall. Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. *PLoS biology*, 4(2):e20, feb 2006.

[17] H. M. Berman. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, jan 2000.

[18] Matthew Berriman, Brian J Haas, Philip T LoVerde, R Alan Wilson, Gary P Dillon, Gustavo C Cerqueira, Susan T Mashiyama, Bissan Al-Lazikani, Luiza F Andrade, Peter D Ashton, Martin A Aslett, Daniella C Bartholomeu, Gaelle Blandin, Conor R Caffrey, Avril Coghlan, Richard Coulson, Tim A Day, Art Delcher, Ricardo DeMarco, Appolinaire Djikeng, Tina Eyre, John A Gamble, Elodie Ghedin, Yong Gu, Christiane Hertz-Fowler, Hirohisha Hirai, Yuriko Hirai, Robin Houston, Alasdair Ivens, David A Johnston, Daniela Lacerda, Camila D Macedo, Paul McVeigh, Zemin Ning, Guilherme Oliveira, John P Overington, Julian Parkhill, Mihaela Pertea, Raymond J Pierce, Anna V Protasio, Michael A Quail, Marie-Adèle Rajandream, Jane Rogers, Mohammed Sajid, Steven L Salzberg, Mario Stanke, Adrian R Tivey, Owen White, David L Williams, Jennifer Wortman, Wenjie Wu, Mostafa Zamanian, Adhemar Zerlotini, Claire M Fraser-Liggett, Barclay G Barrell, and Najib M El-Sayed. The genome of the blood fluke Schistosoma mansoni. *Nature*, 460(7253):352–8, jul 2009.

[19] Matthew Berriman and Midori Harris. Annotation of Parasite Genomes. In *Parasite Genomics Protocols*, pages 017–044. Humana Press, New Jersey, 2004.

[20] Martin J. Blythe, Damian Kao, Sunir Malla, Joanna Rowsell, Ray Wilson, Deborah Evans, Jamie Jowett, Amy Hall, Virginie Lemay, Sabrina Lam, and A. Aziz Aboobaker. A Dual Platform Approach to Transcript Discovery for the Planarian Schmidtea Mediterranea to Establish RNAseq for Stem Cell and Regeneration Biology. *PLoS ONE*, 5(12):e15617, dec 2010.

[21] Jérôme Boissier, Hélène Moné, Guillaume Mitta, M Dolores Bargues, David Molyneux,

and Santiago Mas-Coma. Schistosomiasis reaches Europe. *The Lancet. Infectious diseases*, 15(7):757–8, jul 2015.

[22] Anthony M Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, pages 1–7, apr 2014.

[23] Andrea G Brear, Jason Yoon, Martin Wojtyniak, and Piali Sengupta. Diverse cell type-specific mechanisms localize G protein-coupled receptors to Caenorhabditis elegans sensory cilia. *Genetics*, 197(2):667–84, jun 2014.

[24] K. BREHM. Echinococcus multilocularis as an experimental model in stem cell research and molecular host-parasite interaction. *Parasitology*, 137(03):537, mar 2010.

[25] S E Brenner. Errors in genome annotation. *Trends in genetics : TIG*, 15(4):132–3, apr 1999.

[26] Andrew R. Burns, Genna M. Luciani, Gabriel Musso, Rachel Bagg, May Yeo, Yuqian Zhang, Luckshika Rajendran, John Glavin, Robert Hunter, Elizabeth Redman, Susan Stasiuk, Michael Schertzberg, G. Angus McQuibban, Conor R. Caffrey, Sean R. Cutler, Mike Tyers, Guri Giaever, Corey Nislow, Andy G. Fraser, Calum A. MacRae, John Gilleard, Peter J. Roy, J. Keiser, J. Utzinger, B. Besier, V. L. Fuller, C. J. Lilley, P. E. Urwin, I. A. Sutherland, D. M. Leathwick, R. Kaminsky, R. Kaminsky, D. Mosimann, H. Sager, P. Stein, B. Hosking, I. Scott, T. C. Kwok, G. A. Lemieux, M. Petrascheck, X. Ye, L. B. Buck, C. K. Leung, A. K. Jones, S. D. Buckingham, D. B. Sattelle, K. G. Simpkin, G. C. Coles, T. G. Geary, D. P. Thompson, J. S. Gilleard, B. H. Lee, J. S. Gilleard, H. Hungenberg, H. Fursch, H. Rieck, E. Hellwege, A. R. Burns, H. Ichimiya, F. Sun, H. Sierotzki, G. Scalliet, C. Sakai, E. Tomitsuka, H. Esumi, S. Harada, K. Kita, H. Ishii, T. Miyamoto, S. Ushio, M. Kakishima, A. Osanai, S. Harada, D. K. Inaoka, J. Ohmori, K. Kita, A. R. Burns, J. Wasmuth, R. Schmid, A. Hedley, M. Blaxter, M. Guest, B. Lehner, J. Tischler, A. G. Fraser, J. Demeler, U. Kuttler, G. von Samson-Himmelstjerna, R. Laing, E. Redman, D. J. Bartley, M. S. Cline, A. M. Bolger, M. Lohse, B. Usadel, H. Li, R. Durbin, M. A. DePristo, K. Wang, M. Li, H. Hakonarson, and J. M. Walker. Caenorhabditis

elegans is a useful model for anthelmintic discovery. *Nature Communications*, 6:7485, jun 2015.

[27] Samuel K Buxton, Claude L Charvet, Cedric Neveu, Jacques Cabaret, Jacques Cortet, Nicolas Peineau, Melanie Abongwa, Elise Courtot, Alan P Robertson, and Richard J Martin. Investigation of acetylcholine receptor diversity in a nematode parasite leads to characterization of tribendimidine- and derquantel-sensitive nAChRs. *PLoS pathogens*, 10(1):e1003870, jan 2014.

[28] C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science (New York, N.Y.)*, 282(5396):2012–8, dec 1998.

[29] Miguel M. Cabada, Martha Lopez, Maria Cruz, Jennifer R. Delgado, Virginia Hill, A. Clinton White, L Valdez, J Huerta-Mercado, DO Freedman, JM Vinetz, E Gotuzzo, and XN Zhou. Treatment Failure after Multiple Courses of Triclabendazole among Patients with Fascioliasis in Cusco, Peru: A Case Series. *PLOS Neglected Tropical Diseases*, 10(1):e0004361, jan 2016.

[30] Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. BLAST+: architecture and applications. *BMC bioinformatics*, 10:421, jan 2009.

[31] Cyril Caminade, Jan van Dijk, Matthew Baylis, and Diana Williams. Modelling recent and future climatic suitability for fasciolosis in Europe. *Geospatial health*, 9(2):301–8, jan 2015.

[32] William C. Campbell, John Janovy, and Gerald W. Esch. Chemotherapy of helminth infections: A centennial reflection. In *A Century of Parasitology*, pages 202–214. John Wiley & Sons, Ltd, Chichester, UK, feb 2016.

[33] Tulio D L Campos, Neil D Young, Pasi K Korhonen, Ross S Hall, Stefano Mangiola, Andrew Lonie, and Robin B Gasser. Identification of G protein-coupled receptors in

Schistosoma haematobium and S. mansoni by comparative genomics. *Parasites & vectors*, 7:242, 2014.

[34] Brandi L Cantarel, Ian Korf, Sofia M C Robb, Genis Parra, Eric Ross, Barry Moore, Carson Holt, Alejandro Sánchez Alvarado, and Mark Yandell. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research*, 18(1):188–96, jan 2008.

[35] Salvador Capella-Gutiérrez, José M Silla-Martínez, and Toni Gabaldón. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, 25(15):1972–3, aug 2009.

[36] John D. Chan, Prince N. Agbedanu, Thomas Grab, Mostafa Zamanian, Peter I. Dosa, Timothy A. Day, and Jonathan S. Marchant. Ergot Alkaloids (Re)generate New Leads as Antiparasitics. *PLoS Neglected Tropical Diseases*, 9(9):1–18, 2015.

[37] John D Chan, Prince N Agbedanu, Mostafa Zamanian, Sarah M Gruba, Christy L Haynes, Timothy A Day, and Jonathan S Marchant. 'Death and axes': unexpected Ca2+ entry phenologs predict new anti-schistosomal agents. *PLoS pathogens*, 10(2):e1003942, feb 2014.

[38] John D. Chan, Thomas Grab, and Jonathan S. Marchant. Kinetic profiling an abundantly expressed planarian serotonergic GPCR identifies bromocriptine as a perdurant antagonist. *International Journal for Parasitology: Drugs and Drug Resistance*, 2016.

[39] John D Chan, Magdalena Zarowiecki, and Jonathan S Marchant. Ca channels and praziquantel: a view from the free world. *Parasitology international*, 62(6):619–28, dec 2013.

[40] John D. Chan, Dan Zhang, Xiaolong Liu, Magdalena Zarowiecki, Matthew Berriman, and Jonathan S. Marchant. Utilizing the planarian voltage-gated ion channel transcriptome to resolve a role for a Ca2+ channel in neuromuscular function and regeneration. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, oct 2016.

[41] Bing Chen and Jian-Fan Wen. The adaptive evolution divergence of triosephosphate isomerases between parasitic and free-living flatworms and the discovery of a potential universal target against flatworm parasites. *Parasitology research*, 109(2):283–289, aug 2011.

[42] Changchun Chen, Eisuke Itakura, Katherine P Weber, Ramanujan S Hegde, and Mario de Bono. An ER complex of ODR-4 and ODR-8/Ufm1 specific protease 2 promotes GPCR maturation by a Ufm1-independent mechanism. *PLoS genetics*, 10(3):e1004082, mar 2014.

[43] James J. Collins, Xiaowen Hou, Elena V. Romanova, Bramwell G. Lambrus, Claire M. Miller, Amir Saberi, Jonathan V. Sweedler, and Phillip A. Newmark. Genome-wide analyses reveal a role for peptide hormones in planarian germline development. *PLoS Biology*, 8(10), 2010.

[44] James J Collins and Phillip A Newmark. It's no fluke: the planarian as a model for understanding schistosomes. *PLoS pathogens*, 9(7):e1003396, jan 2013.

[45] James J. Collins III, Bo Wang, Bramwell G. Lambrus, Marla E. Tharp, Harini Iyer, and Phillip A. Newmark. Adult somatic stem cells in the human parasite Schistosoma mansoni. *Nature*, 494(7438):476–479, feb 2013.

[46] Ana Conesa, Stefan Götz, Juan Miguel Garc\'\ia-Gómez, Javier Terol, Manuel Talón, and Montserrat Robles. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)*, 21(18):3674–3676, sep 2005.

[47] Le Cong, F Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D Hsu, Xuebing Wu, Wenyan Jiang, Luciano A Marraffini, and Feng Zhang. Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, 339(6121):819–23, feb 2013.

[48] Thomas Crellen, Martin Walker, Poppy H. L. Lamberton, Narcis B. Kabatereine, Edridah M. Tukahebwa, James A. Cotton, and Joanne P. Webster. Reduced Efficacy of

Praziquantel Against Schistosoma mansoni Is Associated With Multiple Rounds of Mass Drug Administration. *Clinical Infectious Diseases*, 6(9):ciw506, jul 2016.

[49] Matt Crook. The dauer hypothesis and the evolution of parasitism: 20 years on and still going strong. *International journal for parasitology*, 44(1):1–8, jan 2014.

[50] Gregory J Crowther, Dhanasekaran Shanmugam, Santiago J Carmona, Maria A Doyle, Christiane Hertz-Fowler, Matthew Berriman, Solomon Nwaka, Stuart A Ralph, David S Roos, Wesley C Van Voorhis, and Fernán Agüero. Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. *PLoS neglected tropical diseases*, 4(8):e804, 2010.

[51] Scott F Cummins, Dirk Erpenbeck, Zhihua Zou, Charles Claudianos, Leonid L Moroz, Gregg T Nagle, Bernard M Degnan, I Antonov, I Antonova, ER Kandel, RD Hawkins, R Balakrishnan, JR Nambu, RH Scheller, SD Painter, SF Cummins, AE Nichols, DB Akalal, CH Schein, W Braun, JS Smith, AJ Susswein, M Levy, PA de Boer, A ter Maat, MW Miller, C Scanlan, RM Milberg, JV Sweedler, GT Nagle, L Buck, R Axel, SD Liberles, LB Buck, C Dulac, R Axel, NJ Ryba, R Tirindelli, X Li, L Staszewski, H Xu, K Durick, M Zoller, E Adler, H Matsunami, JP Montmayeur, LB Buck, K Touhara, F Zufall, T Leinders-Zufall, Y Cao, BC Oh, L Stryer, X Zhang, S Firestein, PJ Clyne, CG Warr, MR Freeman, D Lessing, J Kim, JR Carlson, LB Vosshall, H Amrein, PS Morozov, A Rzhetsky, R Axel, HM Robertson, KW Wanner, R Benton, S Sachse, SW Michnick, LB Vosshall, K Sato, M Pellegrino, T Nakagawa, T Nakagawa, LB Vosshall, K Touhara, F Raible, K Tessmar-Raible, E Arboleda, T Kaller, P Bork, D Arendt, MI Arnone, AS Mobley, G Mahendra, MT Lucero, SF Cummins, CH Schein, Y Xu, W Braun, GT Nagle, SF Cummins, AE Nichols, A Amare, AB Hummon, JV Sweedler, GT Nagle, SF Cummins, AE Nichols, CJ Warso, GT Nagle, M Levy, M Blumberg, AJ Susswein, K Gobbeler, A Klussmann-Kolb, DG Emery, TE Audesirk, A Wertz, W Rossler, M Obermayer, U Bickmeyer, SF Cummins, MR De Vries, KS Hill, D Boehning, GT Nagle, LL Moroz, KT Clifford, L Gross, K Johnson, KJ Martin, N Shaheen, MA Harrington, LL Moroz, JR Edwards, SV Puthanveettil, AB Kohn, T Ha, A Heyland, B Knud-

sen, A Sahni, F Yu, L Liu, S Jezzini, P Lovell, W Lannucculli, M Chen, T Nguyen, H Sheng, R Shaw, S Kalachikov, YV Panchin, W Farmerie, JJ Russo, J Ju, ER Kandel, NJ Mulder, R Apweiler, TK Attwood, A Bairoch, A Bateman, D Binns, P Bork, V Buillard, L Cerutti, R Copley, E Courcelle, U Das, L Daugherty, M Dibley, R Finn, W Fleischmann, J Gough, D Haft, N Hulo, S Hunter, D Kahn, A Kanapin, A Kejariwal, A Labarga, PS Langendijk-Genevaux, D Lonsdale, R Lopez, I Letunic, M Madera, J Maslen, J Freitag, G Ludwig, I Andreini, P Rossler, H Breer, S Katada, M Tanaka, K Touhara, K Scott, R Brady, A Cravchik, P Morozov, A Rzhetsky, C Zuker, R Axel, MM Rosenkilde, TN Kledal, TW Schwartz, S Acharya, SS Karnik, J Ballesteros, S Kitanovic, F Guarnieri, P Davies, BJ Fromme, K Konvicka, L Chi, RP Millar, JS Davidson, H Weinstein, SC Sealfon, A Pathe-Neuschäfer-Rube, F Neuschäfer-Rube, GP Püschel, LM Wodicka, DE Morse, H Shinohara, T Asano, K Kato, C Jia, G Goldman, M Halpern, A Berghard, L Buck, R Vassar, SK Chao, R Sitcheran, JM Nunez, LB Vosshall, R Axel, N Ferrand, M Pessah, S Frayon, J Marais, JM Garel, M Parmentier, F Libert, S Schurmans, S Schiffmann, A Lefort, D Eggerickx, C Ledent, C Mollereau, C Gerard, J Perret, A Grootegoed, G Vassart, N Fukuda, K Yomogida, M Okabe, K Touhara, A Branscomb, J Seger, RL White, J Krieger, O Klink, C Mohl, K Raming, H Breer, CA Hill, AN Fox, RJ Pitts, LB Kent, PL Tan, MA Chrystal, A Cravchik, FH Collins, HM Robertson, LJ Zwiebel, Y Go, Y Niimura, J Bohbot, RJ Pitts, HW Kwon, M Rutzler, HM Robertson, LJ Zwiebel, ER Troemel, JH Chou, ND Dwyer, HA Colbert, CI Bargmann, SY Xie, P Feinstein, P Mombaerts, N Nozawa, N Masatochi, MK Stewart, NL Clark, G Merrihew, EM Galloway, JH Thomas, C McBride, HM Robertson, CG Warr, JR Carlson, ER Liman, H Innan, C Notredame, DG Higgins, J Heringa, A Stamatakis, F Ronquist, JP Huelsenbeck, S Tavare, S Whelan, and N Goldman. Candidate chemoreceptor subfamilies differentially expressed in the chemosensory organs of the mollusc Aplysia. *BMC Biology*, 7(1):28, 2009.

[52] T A Day, A G Maule, C Shaw, D W Halton, S Moore, J L Bennett, and R A Pax. Platyhelminth FMRFamide-related peptides (FaRPs) contract Schistosoma mansoni (Trema-

toda: Digenea) muscle fibres in vitro. *Parasitology*, 109 (Pt 4:455–459, nov 1994.

[53] Tim A. Day, Aaron G. Maule, Chris Shaw, and Ralph A. Pax. Structure-activity relationships of FMRFamide-related peptides contracting Schistosoma mansoni muscle. *Peptides*, 18(7):917–921, 1997.

[54] D Donnelly, J B Findlay, and T L Blundell. The evolution and structure of aminergic G protein-coupled receptors. *Receptors & channels*, 2(1):61–78, 1994.

[55] Maria A Doyle, Robin B Gasser, Ben J Woodcroft, Ross S Hall, Stuart A Ralph, P Trouiller, P Olliaro, E Torreele, J Orbinski, R Laing, N Ford, S Nwaka, A Hudson, SY Gerdes, MD Scholle, JW Campbell, G Balazsi, E Ravasz, MD Daugherty, AL Somera, NC Kyrpides, I Anderson, MS Gelfand, G Giaever, AM Chu, L Ni, C Connelly, L Riles, S Veronneau, S Dow, A Lucau-Danila, K Anderson, B Andre, EA Winzeler, DD Shoemaker, A Astromoff, H Liang, K Anderson, B Andre, R Bangham, R Benito, JD Boeke, H Bussey, Y Kang, T Durfee, JD Glasner, Y Qiu, D Frisch, KM Winterberg, FR Blattner, CM Sassetti, DH Boyd, EJ Rubin, CM Sassetti, DH Boyd, EJ Rubin, J Bethony, S Brooker, M Albonico, SM Geiger, A Loukas, D Diemert, PJ Hotez, MS Chan, AJ Wolstenholme, I Fairweather, R Prichard, G von Samson-Himmelstjerna, NC Sangster, WE Pomroy, B Besier, JS Gilleard, RN Beech, M Nguyen, A Alfonso, CD Johnson, JB Rand, JA Lewis, CH Wu, JH Levine, H Berg, BE Campbell, SH Nagaraj, M Hu, W Zhong, PW Sternberg, EK Ong, A Loukas, S Ranganathan, I Beveridge, RL McInnes, R Hoekstra, A Visser, M Otsen, J Tibben, JA Lenstra, MH Roos, JP McCarter, MD Mitreva, J Martin, M Dante, T Wylie, U Rao, D Pape, Y Bowers, B Theising, CV Murphy, M Mitreva, DP Jasmer, J Appleton, J Martin, M Dante, T Wylie, SW Clifton, RH Waterston, JP McCarter, AJ Nisbet, RB Gasser, FJ Thompson, M Mitreva, GL Barker, J Martin, RH Waterston, JP McCarter, ME Viney, Y Yin, J Martin, JP McCarter, SW Clifton, RK Wilson, M Mitreva, E Ghedin, S Wang, D Spiro, E Caler, Q Zhao, J Crabtree, JE Allen, AL Delcher, DB Guiliano, D Miranda-Saavedra, M Mitreva, DS Zarlenga, JP McCarter, DP Jasmer, JS Gilleard, DJ Woods, JAT Dow, P Geldhof, L Murray, A Couthier, JS Gilleard, G McLauchlan, DP Knox,

C Britton, P Geldhof, A Visser, D Clark, G Saunders, C Britton, J Gilleard, M Berriman, D Knox, DP Knox, P Geldhof, A Visser, C Britton, RS Kamath, J Ahringer, RS Kamath, AG Fraser, Y Dong, G Poulin, R Durbin, M Gotta, A Kanapin, N Le Bot, S Moreno, M Sohrmann, JL Mummery-Widmer, M Yamazaki, T Stoeger, M Novatchkova, S Bhalerao, D Chen, G Dietzl, BJ Dickson, JA Knoblich, M Blaxter, JS Gilleard, W Zhong, PW Sternberg, I Lee, B Lehner, C Crombie, W Wong, AG Fraser, EM Marcotte, K Zhang, PK Rathod, CJ Bacchi, HC Nathan, SH Hutner, PP McCann, A Sjoerdsma, JM Kraus, CL Verlinde, M Karimi, GI Lepesheva, MH Gelb, FS Buckner, LF Silvian, J Wang, TA Steitz, GC Conant, A Wagner, X Gu, B-Y Liao, J Zhang, AR Mushegian, EV Koonin, A Amsterdam, RM Nissen, Z Sun, EC Swindell, S Farrington, N Hopkins, KR Sakharkar, MK Sakharkar, VTK Chow, F Schreiber, K Pick, D Erpenbeck, G Worheide, B Morgenstern, K Lundstrom, YC Hwang, CC Lin, JY Chang, H Mori, HF Juan, HC Huang, MS Kwa, JG Veenstra, MH Roos, DF Cully, DK Vassilatis, KK Liu, PS Paress, LH Van der Ploeg, JM Schaeffer, JP Arena, RK Prichard, JA Dent, MM Smith, DK Vassilatis, L Avery, RJ Martin, JA Thanassi, SL Hartman-Neumann, TJ Dougherty, BA Dougherty, MJ Pucci, T Roemer, B Jiang, J Davison, T Ketela, K Veillette, A Breton, F Tandia, A Linteau, S Sillaots, C Marta, W Hu, S Sillaots, S Lemieux, J Davison, S Kauffman, A Breton, A Linteau, C Xin, J Bowman, J Becker, F Aguero, B Al-Lazikani, M Aslett, M Berriman, FS Buckner, RK Campbell, S Carmona, IM Carruthers, AW Chan, F Chen, AR Joyce, BO Palsson, G Lamichhane, M Zignol, NJ Blades, DE Geiman, A Dougherty, J Grosset, KW Broman, WR Bishai, M Seringhaus, A Paccanaro, A Borneman, M Snyder, M Gerstein, C Volker, JR Brown, G Butland, JM Peregrin-Alvarez, J Li, W Yang, X Yang, V Canadien, A Starostine, D Richards, B Beattie, N Krogan, NJ Krogan, G Cagney, H Yu, G Zhong, X Guo, A Ignatchenko, J Li, S Pu, N Datta, AP Tikuisis, JP Overington, B Al-Lazikani, AL Hopkins, EL Sonnhammer, EV Koonin, L Li, CJ Stoeckert, and DS Roos. Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. *BMC Genomics*, 11(1):222, 2010.

[56] Joseph El-On. Benzimidazole treatment of cystic echinococcosis. *Acta Tropica*, 85(2):243–252, feb 2003.

[57] Fouad El-Shehabi and Paula Ribeiro. Histamine signalling in Schistosoma mansoni: Immunolocalisation and characterisation of a new histamine-responsive receptor (SmGPR-2). *International Journal for Parasitology*, 40(12):1395–1406, oct 2010.

[58] Fouad El-Shehabi, Amira Taman, Lorena S. Moali, Nelly El-Sakkary, and Paula Ribeiro. A novel g protein-coupled receptor of schistosoma mansoni (smgpr-3) is activated by dopamine and is widely expressed in the nervous system. *PLoS Neglected Tropical Diseases*, 6(2), 2012.

[59] Deborah J Evans, Suthira Owlarn, Belen Tejada Romero, Chen Chen, and a Aziz Aboobaker. Combining classical and molecular approaches elaborates on the complexity of mechanisms underpinning anterior regeneration. *PloS one*, 6(11):e27927, jan 2011.

[60] P G Fallon and M J Doenhoff. Drug-resistant schistosomiasis: resistance to praziquantel and oxamniquine induced in Schistosoma mansoni in mice is drug specific. *The American journal of tropical medicine and hygiene*, 51(1):83–8, jul 1994.

[61] Ye Fang, Terry Kenakin, and Changlu Liu. Editorial: Orphan GPCRs As Emerging Drug Targets. *Frontiers in pharmacology*, 6:295, jan 2015.

[62] Robert D Finn, Penelope Coggill, Ruth Y Eberhardt, Sean R Eddy, Jaina Mistry, Alex L Mitchell, Simon C Potter, Marco Punta, Matloob Qureshi, Amaia Sangrador-Vegas, Gustavo A Salazar, John Tate, and Alex Bateman. The Pfam protein families database: towards a more sustainable future. *Nucleic acids research*, 44(D1):D279–D285, dec 2015.

[63] Robert Fredriksson and Helgi B Schiöth. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol. Pharmacol.*, 67(5):1414–1425, may 2005.

[64] Mireille Galloni. Global irradiation effects, stem cell genes and rare transcripts in the planarian transcriptome. *The International journal of developmental biology*, 56(1-3):103–116, jan 2012.

[65] Jordi Garcia-Fernandez, Jose Ram, Gemma Mar-any, and Ana Maria Mun. High Copy Number of Highly Similar mariner-like Transposons in Planarian (Platyhelminthe) : Evidence for a Trans-Phyla Horizontal Transfer. *Molecular biology and evolultion*, 12(3), 1995.

[66] Timothy G. Geary, Judy A. Sakanari, and Conor R. Caffrey. Anthelmintic Drug Discovery: Into the Future. *Journal of Parasitology*, 101(2):125–133, apr 2015.

[67] Timothy G Geary and David P Thompson. Development of antiparasitic drugs in the 21st century. *Veterinary Parasitology*, 115(2):167–184, jul 2003.

[68] P Geldhof, A Visser, D Clark, G Saunders, C Britton, J Gilleard, M Berriman, and D Knox. RNA interference in parasitic helminths: current situation, potential pitfalls and future prospects. *Parasitology*, 134(Pt 5):609–19, may 2007.

[69] Ian H Gilbert. Drug discovery for neglected diseases: molecular target-based and phenotypic approaches. *Journal of medicinal chemistry*, 56(20):7719–7726, oct 2013.

[70] J S Gilleard. The use of Caenorhabditis elegans in parasitic nematode research. *Parasitology*, 128 Suppl:S49—-70, jan 2004.

[71] Danielle Gordon, Ruth Zadoks, Philip Skuce, and Neil Sargison. Confirmation of triclabendazole resistance in liver fluke in the UK. *The Veterinary record*, 171(6):159–60, aug 2012.

[72] M G Grabherr, B J Haas, and M Yassour. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature*, 29(7):644–652, 2011.

[73] Marek Grabowski, Andrzej Joachimiak, Zbyszek Otwinowski, and Wladek Minor. Structural genomics: keeping up with expanding knowledge of the protein universe. *Current opinion in structural biology*, 17(3):347–53, jun 2007.

[74] Christoph G Grevelding, L.C Leite, E Dias-Neto, C.F Menck, R.A Wilson, G.M Williams, and D.P McManus. Schistosoma. *Current biology : CB*, 14(14):R545, jul 2004.

[75] Marcus Guest, Kathryn Bull, Robert J. Walker, Kiran Amliwala, Vincent O'Connor, Achim Harder, Lindy Holden-Dye, and Neil A. Hopper. The calcium-activated potassium channel, SLO-1, is required for the action of the novel cyclo-octadepsipeptide anthelmintic, emodepside, in Caenorhabditis elegans. *International Journal for Parasitology*, 37(14):1577–1588, dec 2007.

[76] Danielle Hagstrom, Olivier Cochet-Escartin, Siqi Zhang, Cindy Khuu, and Eva-Maria S. Collins. Freshwater Planarians as an Alternative Animal Model for Neurotoxicology. *Toxicological Sciences*, 147(1):270–285, sep 2015.

[77] Fadi F. Hamdan, Mark Abramovitz, Aisha Mousa, Jinling Xie, Yves Durocher, and Paula Ribeiro. A novel Schistosoma mansoni G protein-coupled receptor is responsive to histamine. *Molecular and Biochemical Parasitology*, 119(1):75–86, 2002.

[78] Yun Hao and Nicholas P Tatonetti. Predicting G Protein-Coupled Receptor Downstream Signaling by Tissue Expression. *Bioinformatics (Oxford, England)*, aug 2016.

[79] A. Harder, L. HoldenDye, R. Walker, and F. Wunderlich. Mechanisms of action of emodepside. *Parasitology Research*, 97(S01):S1–S10, oct 2005.

[80] Javier Herrero, Matthieu Muffato, Kathryn Beal, Stephen Fitzgerald, Leo Gordon, Miguel Pignatelli, Albert J. Vilella, Stephen M. J. Searle, Ridwan Amode, Simon Brent, William Spooner, Eugene Kulesha, Andrew Yates, and Paul Flicek. Ensembl comparative genomics resources. *Database*, 2016:bav096, feb 2016.

[81] Byron Hetrick, Min Suk Han, Luke a Helgeson, and Brad J Nolen. Small molecules CK-666 and CK-869 inhibit actin-related protein 2/3 complex by blocking an activating conformational change. *Chemistry & biology*, 20(5):701–712, may 2013.

[82] Florence Horn, Eleonora M. van der Wenden, Laerte Oliveira, Adriaan P. IJzerman, and Gerrit Vriend. Receptors coupling to G proteins: Is there a signal behind the sequence? *Proteins: Structure, Function, and Genetics*, 41(4):448–459, dec 2000.

[83] Kurt Hornik. The R FAQ, 2014.

[84] Peter J Hotez and Aruna Kamath. Neglected tropical diseases in sub-saharan Africa: review of their prevalence, distribution, and disease burden. *PLoS neglected tropical diseases*, 3(8):e412, aug 2009.

[85] Kevin L Howe, Bruce J Bolt, Scott Cain, Juancarlos Chan, Wen J Chen, Paul Davis, James Done, Thomas Down, Sibyl Gao, Christian Grove, Todd W Harris, Ranjana Kishore, Raymond Lee, Jane Lomax, Yuling Li, Hans-Michael Muller, Cecilia Nakamura, Paulo Nuin, Michael Paulini, Daniela Raciti, Gary Schindelman, Eleanor Stanley, Mary Ann Tuli, Kimberly Van Auken, Daniel Wang, Xiaodong Wang, Gary Williams, Adam Wright, Karen Yook, Matthew Berriman, Paul Kersey, Tim Schedl, Lincoln Stein, and Paul W Sternberg. WormBase 2016: expanding to enable helminth genomic research. *Nucleic acids research*, 44(D1):D774–80, jan 2016.

[86] Kevin L Howe, Bruce J Bolt, Myriam Shafie, Paul Kersey, and Matthew Berriman. WormBase ParaSite - a comprehensive resource for helminth genomics. *Molecular and biochemical parasitology*, nov 2016.

[87] X Huang, E Duran, F Diaz, H Xiao, W S Messer, and R Komuniecki. Alternative-splicing of serotonin receptor isoforms in the pharynx and muscle of the parasitic nematode, Ascaris suum. *Molecular and biochemical parasitology*, 101(1-2):95–106, jun 1999.

[88] Philip W. Ingham, Yoshiro Nakano, and Claudia Seger. Mechanisms and functions of Hedgehog signalling across the metazoa. *Nature Reviews Genetics*, 12(6):393–406, jun 2011.

[89] Vignir Isberg, Stefan Mordalski, Christian Munk, Krzysztof Rataj, Kasper Harpsøe, Alexander S Hauser, Bas Vroling, Andrzej J Bojarski, Gert Vriend, and David E Gloriam. GPCRdb: an information system for G protein-coupled receptors. *Nucleic acids research*, 44(D1):D356–64, jan 2016.

[90] Vignir Isberg, Isberg Vignir, Vroling Bas, Rob van der Kant, Li Kang, Vriend Gert, and Gloriam David. {GPCRDB}: an information system for {G} protein-coupled receptors. *Nucleic Acids Res.*, 42(D1):D422—-D425, 2013.

[91] M Ismail, S Botros, A Metwally, S William, A Farghally, L F Tao, T A Day, and J L Bennett. Resistance to praziquantel: direct evidence from Schistosoma mansoni isolated from Egyptian villagers. *The American journal of tropical medicine and hygiene*, 60(6):932–5, jun 1999.

[92] G. Jekely. Global view of the evolution and diversity of metazoan neuropeptide signaling. *Proceedings of the National Academy of Sciences*, 110(21):8702–8707, may 2013.

[93] S.S. Johnson, E.M. Thomas, and T.G. Geary. Intestinal Worm Infections. In Oto Zak and Merle A. Sande, editors, *Handbook of Animal Models of Infection: Experimental Models in Antimicrobial Chemotherapy*, chapter 106. Academic Press, 1 edition, 1999.

[94] Andrew K. Jones, Paul Davis, Jonathan Hodgkin, and David B. Sattelle. The nicotinic acetylcholine receptor gene family of the nematode Caenorhabditis elegans: an update on nomenclature. *Invertebrate Neuroscience*, 7(2):129–131, may 2007.

[95] Philip Jones, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, John Maslen, Alex Mitchell, Gift Nuka, Sebastien Pesseat, Antony F Quinn, Amaia Sangrador-Vegas, Maxim Scheremetjew, Siew-Yit Yong, Rodrigo Lopez, and Sarah Hunter. InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, 30(9):1236–1240, may 2014.

[96] Trupti Joshi and Dong Xu. Quantitative assessment of relationship between sequence similarity and function similarity. *BMC Genomics*, 8:222, 2007.

[97] Minoru Kanehisa, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research*, 42(Database issue):D199—-205, jan 2014.

[98] Damian Kao, Daniel Felix, and Aziz Aboobaker. The planarian regeneration transcriptome reveals a shared but temporally shifted regulatory program between opposing head and tail scenarios. *BMC genomics*, 14:797, jan 2013.

[99] Kazutaka Katoh and Daron M Standley. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–80, apr 2013.

[100] Jennifer Keiser, Kigbafori D Silué, Lukas K Adiossan, Nicaise A N'Guessan, N'Chou Monsan, Jürg Utzinger, and Eliézer K N'Goran. Praziquantel, mefloquine-praziquantel, and mefloquine-artesunate-praziquantel against Schistosoma haematobium: a randomized, exploratory, open-label trial. *PLoS neglected tropical diseases*, 8(7):e2975, jul 2014.

[101] Jennifer Keiser and Jürg Utzinger. Food-borne trematodiases. *Clinical microbiology reviews*, 22(3):466–83, jul 2009.

[102] Michael J. Kimber, Laura Sayegh, Fouad El-Shehabi, Chuanzhe Song, Mostafa Zamanian, Debra J. Woods, Tim A. Day, and Paula Ribeiro. Identification of an Ascaris G protein-coupled acetylcholine receptor with atypical muscarinic pharmacology. *International Journal for Parasitology*, 39(11):1215–1222, sep 2009.

[103] David P Knox, Peter Geldhof, Aline Visser, and Collette Britton. {RNA} interference in parasitic nematodes of animals: a reality check? *Trends Parasitol.*, 23(3):105–107, mar 2007.

[104] A. B. Kohn, P. A. V. Anderson, J. M. Roberts-Misterly, and R. M. Greenberg. Schistosome Calcium Channel Subunits: UNUSUAL MODULATORY EFFECTS AND POTENTIAL ROLE IN THE ACTION OF THE ANTISCHISTOSOMAL DRUG PRAZIQUANTEL. *Journal of Biological Chemistry*, 276(40):36873–36876, oct 2001.

[105] L F Kolakowski Jr. {GCRDb}: a G-protein-coupled receptor database. *Receptors Channels*, 2(1):1–7, 1994.

[106] Triinu Koressaar and Maido Remm. Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23(10):1289–1291, may 2007.

[107] Natalia D Kreshchenko. Functions of flatworm neuropeptides NPF, GYIRF and FMRF

in course of pharyngeal regeneration of anterior body fragments of planarian, Girardia tigrina. *Acta biologica Hungarica*, 59 Suppl:199–207, jan 2008.

[108] Arunkumar Krishnan, Markus Sällman Almén, Robert Fredriksson, and Helgi B Schiöth. The origin of {GPCRs}: identification of mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled {GPCRs} in fungi. *PLoS One*, 7(1):e29817, 2012.

[109] Arunkumar Krishnan, Markus Sällman Almén, Robert Fredriksson, and Helgi B Schiöth. Insights into the origin of nematode chemosensory GPCRs: putative orthologs of the Srw family are found across several phyla of protostomes. *PloS one*, 9(3):e93048, jan 2014.

[110] Martin Krzywinski, Jacqueline Schein, Inanç Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J Jones, and Marco a Marra. Circos: an information aesthetic for comparative genomics. *Genome research*, 19(9):1639–1645, sep 2009.

[111] Malin C. Lagerström and Helgi B. Schiöth. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature Reviews Drug Discovery*, 7(4):339–357, apr 2008.

[112] Roz Laing, Taisei Kikuchi, Axel Martinelli, Isheng J Tsai, Robin N Beech, Elizabeth Redman, Nancy Holroyd, David J Bartley, Helen Beasley, Collette Britton, David Curran, Eileen Devaney, Aude Gilabert, Martin Hunt, Frank Jackson, Stephanie L Johnston, Ivan Kryukov, Keyu Li, Alison A Morrison, Adam J Reid, Neil Sargison, Gary I Saunders, James D Wasmuth, Adrian Wolstenholme, Matthew Berriman, John S Gilleard, and James A Cotton. The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery. *Genome biology*, 14(8):R88, 2013.

[113] Mahima Lall, Ajay Kumar Sahni, and A K Rajput. Pleuropulmonary paragonimiasis: mimicker of tuberculosis. *Pathogens and global health*, 107(1):40–2, jan 2013.

[114] Ben Langmead, Cole Trapnell, Mihai Pop, and Steven L Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3):R25, jan 2009.

[115] SylvainW. Lapan and PeterW. Reddien. Transcriptome Analysis of the Planarian Eye Identifies ovo as a Specific Regulator of Eye Regeneration. *Cell Reports*, 2(2):294–307, aug 2012.

[116] Nicholas D Larusso, Brian E Ruttenberg, Ambuj K Singh, and Todd H Oakley. Type II opsins: evolutionary origin by internal domain duplication? *Journal of molecular evolution*, 66(5):417–23, may 2008.

[117] Wenjing Law, Leah M. Wuescher, Amanda Ortega, Vera M. Hapiak, Patricia R. Komuniecki, and Richard Komuniecki. Heterologous Expression in Remodeled C. elegans: A Platform for Monoaminergic Agonist Identification and Anthelmintic Screening. *PLoS Pathogens*, 11(4):1–17, 2015.

[118] Marcus Lechner, Sven Findeiss, Lydia Steiner, Manja Marz, Peter F Stadler, and Sonja J Prohaska. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC bioinformatics*, 12(1):124, jan 2011.

[119] R. J. Lefkowitz and Sudha K Shenoy. Transduction of Receptor Signals by -Arrestins. *Science*, 308(5721):512–517, apr 2005.

[120] Ben-Wen Li, Amy C Rush, and Gary J Weil. Expression of five acetylcholine receptor subunit genes in Brugia malayi adult worms. *International journal for parasitology. Drugs and drug resistance*, 5(3):100–9, dec 2015.

[121] Bo Li and Colin N Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12:323, jan 2011.

[122] Qian Li, Yaw Tachie-Baffour, Zhikai Liu, Maude W Baldwin, Andrew C Kruse, Stephen D Liberles, MW. Baldwin, Y. Toda, T. Nakagita, MJ. O'Connell, KC. Klasing, T. Misaka, SV. Edwards, SD. Liberles, M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, TG. Cassarino, M. Bertoni, L. Bordoli, T. Schwede, B. Borowsky, N. Adham, KA. Jones, R. Raddatz, R. Artymyshyn, KL. Ogozalek, MM. Durkin, PP. Lakhlani, JA. Bonini, S. Pathirana, N. Boyle, X. Pu, E. Kouranova, H. Lichtblau, FY.

Ochoa, TA. Branchek, C. Gerald, JR. Bunzow, MS. Sonders, S. Arttamangkul, LM. Harrison, G. Zhang, DI. Quigley, T. Darland, KL. Suchland, S. Pasumamula, JL. Kennedy, SB. Olson, RE. Magenis, SG. Amara, DK. Grandy, A. Dewan, R. Pacifico, R. Zhan, D. Rinberg, T. Bozza, DM. Ferrero, JK. Lemon, D. Fluegge, SL. Pashkovski, WJ. Korzan, SR. Datta, M. Spehr, M. Fendt, SD. Liberles, DM. Ferrero, D. Wacker, MA. Roque, MW. Baldwin, RC. Stevens, SD. Liberles, RA. Friesner, RB. Murphy, MP. Repasky, LL. Frye, JR. Greenwood, TA. Halgren, PC. Sanschagrin, DT. Mainz, TJ. Hara, LF. Horowitz, LR. Saraiva, D. Kuang, KH. Yoon, LB. Buck, A. Hussain, LR. Saraiva, DM. Ferrero, G. Ahuja, VS. Krishna, SD. Liberles, SI. Korsching, A. Hussain, LR. Saraiva, SI. Korsching, P. Jiang, J. Josue, X. Li, D. Glaser, W. Li, JG. Brand, RF. Margolskee, DR. Reed, GK. Beauchamp, K. Katoh, K. Misawa, K. Kuma, T. Miyata, V. Katritch, V. Cherezov, RC. Stevens, F. Kermen, LM. Franco, C. Wyatt, E. Yaksi, D. Krautwurst, KW. Yau, RR. Reed, Q. Li, WJ. Korzan, DM. Ferrero, RB. Chang, DS. Roy, M. Buchi, JK. Lemon, AW. Kaur, L. Stowers, M. Fendt, SD. Liberles, Q. Li, SD. Liberles, SD. Liberles, LB. Buck, SD. Liberles, LF. Horowitz, D. Kuang, JJ. Contos, KL. Wilson, J. Siltberg-Liberles, DA. Liberles, LB. Buck, SD. Liberles, SD. Liberles, ER. Liman, L. Lindemann, MC. Hoener, JD. Morrow, GR. Margolies, J. Rowland, LJ. Roberts, M. Nei, Y. Niimura, M. Nozawa, EA. Reese, Y. Norimatsu, MS. Grandy, KL. Suchland, JR. Bunzow, DK. Grandy, FG. Revel, J-L. Moreau, RR. Gainetdinov, A. Bradaia, TD. Sotnikova, R. Mory, S. Durkin, KG. Zbinden, R. Norcross, CA. Meyer, V. Metzler, S. Chaboz, L. Ozmen, G. Trube, B. Pouzet, B. Bettler, MG. Caron, JG. Wettstein, MC. Hoener, AM. Ring, A. Manglik, AC. Kruse, MD. Enos, WI. Weis, KC. Garcia, BK. Kobilka, S. Rivière, L. Challet, D. Fluegge, M. Spehr, I. Rodriguez, SH. Rolen, PW. Sorensen, D. Mattson, J. Caprio, H. Saito, M. Kubota, RW. Roberts, Q. Chi, H. Matsunami, LLC. Schrodinger, LLC. Schrodinger, R. Seifert, A. Strasser, EH. Schneider, D. Neumann, S. Dove, A. Buschauer, AR. Shalaby, L. Shi, JA. Javitch, P. Shi, J. Zhang, A. Stamatakis, P. Hoover, J. Rougemont, BC. Stöver, KF. Müller, J-S. Surgand, J. Rodrigo, E. Kellenberger, D. Rognan, JA. Tessarolo, MJ. Tabesh, M. Nesbitt, WS. Davidson, A. Wisman, I. Shrira, J. Zhang, R. Pacifico, D. Cawley, P. Feinstein, T. Bozza, J. Zhang, and DM. Webb. Non-classical

amine recognition evolved in a large clade of olfactory receptors. *eLife*, 4:e10441, 2015.

[123] Weizhong Li and Adam Godzik. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659, jul 2006.

[124] Yang Liao, Gordon K Smyth, and Wei Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, 30(7):923–30, apr 2014.

[125] Henry Lin, Maria F Sassano, Bryan L Roth, and Brian K Shoichet. A pharmacological organization of G proteincoupled receptors. *Nature Methods*, 10(2):140–146, jan 2013.

[126] Nathan C Lo, Isaac I Bogoch, Brian G Blackburn, Giovanna Raso, Eliézer K N'Goran, Jean T Coulibaly, Sören L Becker, Howard B Abrams, Jürg Utzinger, and Jason R Andrews. Comparison of community-wide, integrated mass drug administration strategies for schistosomiasis and soil-transmitted helminthiasis: a cost-effectiveness modelling study. *The Lancet Global Health*, 3(10):e629–e638, oct 2015.

[127] Flora J Logan-Klumpler, Nishadi De Silva, Ulrike Boehme, Matthew B Rogers, Giles Velarde, Jacqueline a McQuillan, Tim Carver, Martin Aslett, Christian Olsen, Sandhya Subramanian, Isabelle Phan, Carol Farris, Siddhartha Mitra, Gowthaman Ramasamy, Haiming Wang, Adrian Tivey, Andrew Jackson, Robin Houston, Julian Parkhill, Matthew Holden, Omar S Harb, Brian P Brunk, Peter J Myler, David Roos, Mark Carrington, Deborah F Smith, Christiane Hertz-Fowler, and Matthew Berriman. GeneDB–an annotation database for pathogens. *Nucleic acids research*, 40(Database issue):D98—-108, jan 2012.

[128] Diomedes E. Logothetis, Yoshihisa Kurachi, Jonas Galper, Eva J. Neer, and David E. Clapham. The $\beta\gamma$ subunits of GTP-binding proteins activate the muscarinic K+ channel in heart. *Nature*, 325(6102):321–326, jan 1987.

[129] Kevin MacDonald, Samuel Buxton, Michael J Kimber, Tim A Day, Alan P Robertson, and Paula Ribeiro. Functional characterization of a novel family of acetylcholine-gated chloride channels in Schistosoma mansoni. *PLoS pathogens*, 10(6):e1004181, jun 2014.

[130] Kevin Macdonald, Michael J Kimber, Tim A Day, and Paula Ribeiro. A constitutively active G protein-coupled acetylcholine receptor regulates motility of larval Schistosoma mansoni. *Molecular & Biochemical Parasitology*, 202:29–37, 2015.

[131] María P Magariños, Santiago J Carmona, Gregory J Crowther, Stuart A Ralph, David S Roos, Dhanasekaran Shanmugam, Wesley C Van Voorhis, and Fernán Agüero. TDR Targets: a chemogenomics resource for neglected diseases. *Nucleic acids research*, 40(Database issue):D1118–27, jan 2012.

[132] Tanja Magoč and Steven L Salzberg. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21):2957–2963, nov 2011.

[133] Orna Man, Yoav Gilad, and Doron Lancet. Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. *Protein Sci.*, 13(1):240–254, jan 2004.

[134] Paul McVeigh, Louise Atkinson, Nikki J Marks, Angela Mousley, Johnathan J Dalzell, Ann Sluder, Lance Hammerland, and Aaron G Maule. Parasite neuropeptide biology: Seeding rational drug target selection? *Int. J. Parasitol. Drugs Drug Resist.*, 2:76–91, dec 2012.

[135] Inge Mertens, Isabelle Clinckspoor, Tom Janssen, Ronald Nachman, and Liliane Schoofs. FMRFamide related peptide ligands activate the Caenorhabditis elegans orphan GPCR Y59H11AL.1. *Peptides*, 27(6):1291–1296, 2006.

[136] John N Milligan and Emmitt R Jolly. Cercarial transformation and in vitro cultivation of Schistosoma mansoni schistosomules. *Journal of visualized experiments : JoVE*, (54):4–7, jan 2011.

[137] Tara Mirzadegan, Gil Benkö, Sławomir Filipek, and Krzysztof Palczewski. Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry*, 42(10):2759–67, mar 2003.

[138] Angela Mousley, Nikki J Marks, David W Halton, Timothy G Geary, David P Thompson, and Aaron G Maule. Arthropod FMRFamide-related peptides modulate muscle activity in helminths. *International journal for parasitology*, 34(6):755–68, may 2004.

[139] Stefanie Mühlfeld, Hans-Peter Schmitt-Wrede, Achim Harder, and Frank Wunderlich. FMRFamide-like neuropeptides as putative ligands of the latrophilin-like HC110-R from Haemonchus contortus. 2009.

[140] Christopher J L Murray, Theo Vos, Rafael Lozano, Mohsen Naghavi, Abraham D Flaxman, Catherine Michaud, Majid Ezzati, Kenji Shibuya, Joshua a Salomon, Safa Abdalla, Victor Aboyans, Jerry Abraham, Ilana Ackerman, Rakesh Aggarwal, Stephanie Y Ahn, Mohammed K Ali, Miriam Alvarado, H Ross Anderson, Laurie M Anderson, Kathryn G Andrews, Charles Atkinson, Larry M Baddour, Adil N Bahalim, Suzanne Barker-Collo, Lope H Barrero, David H Bartels, Maria-Gloria Basáñez, Amanda Baxter, Michelle L Bell, Emelia J Benjamin, Derrick Bennett, Eduardo Bernabé, Kavi Bhalla, Bishal Bhandari, Boris Bikbov, Aref Bin Abdulhak, Gretchen Birbeck, James a Black, Hannah Blencowe, Jed D Blore, Fiona Blyth, Ian Bolliger, Audrey Bonaventure, Soufiane Boufous, Rupert Bourne, Michel Boussinesq, Tasanee Braithwaite, Carol Brayne, Lisa Bridgett, Simon Brooker, Peter Brooks, Traolach S Brugha, Claire Bryan-Hancock, Chiara Bucello, Rachelle Buchbinder, Geoffrey Buckle, Christine M Budke, Michael Burch, Peter Burney, Roy Burstein, Bianca Calabria, Benjamin Campbell, Charles E Canter, Hélène Carabin, Jonathan Carapetis, Loreto Carmona, Claudia Cella, Fiona Charlson, Honglei Chen, Andrew Tai-Ann Cheng, David Chou, Sumeet S Chugh, Luc E Coffeng, Steven D Colan, Samantha Colquhoun, K Ellicott Colson, John Condon, Myles D Connor, Leslie T Cooper, Matthew Corriere, Monica Cortinovis, Karen Courville de Vaccaro, William Couser, Benjamin C Cowie, Michael H Criqui, Marita Cross, Kaustubh C Dabhadkar, Manu Dahiya, Nabila Dahodwala, James Damsere-Derry, Goodarz Danaei, Adrian Davis, Diego De Leo, Louisa Degenhardt, Robert Dellavalle, Allyne Delossantos, Julie Denenberg, Sarah Derrett, Don C Des Jarlais, Samath D Dharmaratne, Mukesh Dherani, Cesar Diaz-Torne, Helen Dolk, E Ray Dorsey, Tim Driscoll, Herbert Duber, Beth Ebel, Karen

Edmond, Alexis Elbaz, Suad Eltahir Ali, Holly Erskine, Patricia J Erwin, Patricia Espindola, Stalin E Ewoigbokhan, Farshad Farzadfar, Valery Feigin, David T Felson, Alize Ferrari, Cleusa P Ferri, Eric M Fèvre, Mariel M Finucane, Seth Flaxman, Louise Flood, Kyle Foreman, Mohammad H Forouzanfar, Francis Gerry R Fowkes, Marlene Fransen, Michael K Freeman, Belinda J Gabbe, Sherine E Gabriel, Emmanuela Gakidou, Hammad a Ganatra, Bianca Garcia, Flavio Gaspari, Richard F Gillum, Gerhard Gmel, Diego Gonzalez-Medina, Richard Gosselin, Rebecca Grainger, Bridget Grant, Justina Groeger, Francis Guillemin, David Gunnell, Ramyani Gupta, Juanita Haagsma, Holly Hagan, Yara a Halasa, Wayne Hall, Diana Haring, Josep Maria Haro, James E Harrison, Rasmus Havmoeller, Roderick J Hay, Hideki Higashi, Catherine Hill, Bruno Hoen, Howard Hoffman, Peter J Hotez, Damian Hoy, John J Huang, Sydney E Ibeanusi, Kathryn H Jacobsen, Spencer L James, Deborah Jarvis, Rashmi Jasrasaria, Sudha Jayaraman, Nicole Johns, Jost B Jonas, Ganesan Karthikeyan, Nicholas Kassebaum, Norito Kawakami, Andre Keren, Jon-Paul Khoo, Charles H King, Lisa Marie Knowlton, Olive Kobusingye, Adofo Koranteng, Rita Krishnamurthi, Francine Laden, Ratilal Lalloo, Laura L Laslett, Tim Lathlean, Janet L Leasher, Yong Yi Lee, James Leigh, Daphna Levinson, Stephen S Lim, Elizabeth Limb, John Kent Lin, Michael Lipnick, Steven E Lipshultz, Wei Liu, Maria Loane, Summer Lockett Ohno, Ronan Lyons, Jacqueline Mabweijano, Michael F MacIntyre, Reza Malekzadeh, Leslie Mallinger, Sivabalan Manivannan, Wagner Marcenes, Lyn March, David J Margolis, Guy B Marks, Robin Marks, Akira Matsumori, Richard Matzopoulos, Bongani M Mayosi, John H McAnulty, Mary M McDermott, Neil McGill, John McGrath, Maria Elena Medina-Mora, Michele Meltzer, George a Mensah, Tony R Merriman, Ana-Claire Meyer, Valeria Miglioli, Matthew Miller, Ted R Miller, Philip B Mitchell, Charles Mock, Ana Olga Mocumbi, Terrie E Moffitt, Ali a Mokdad, Lorenzo Monasta, Marcella Montico, Maziar Moradi-Lakeh, Andrew Moran, Lidia Morawska, Rintaro Mori, Michele E Murdoch, Michael K Mwaniki, Kovin Naidoo, M Nathan Nair, Luigi Naldi, K M Venkat Narayan, Paul K Nelson, Robert G Nelson, Michael C Nevitt, Charles R Newton, Sandra Nolte, Paul Norman, Rosana Norman, Martin O'Donnell, Simon O'Hanlon, Casey Olives, Saad B Omer, Katrina Ortblad, Richard Osborne, Doruk Ozgediz, Andrew

Page, Bishnu Pahari, Jeyaraj Durai Pandian, Andrea Panozo Rivero, Scott B Patten, Neil Pearce, Rogelio Perez Padilla, Fernando Perez-Ruiz, Norberto Perico, Konrad Pesudovs, David Phillips, Michael R Phillips, Kelsey Pierce, Sébastien Pion, Guilherme V Polanczyk, Suzanne Polinder, C Arden Pope, Svetlana Popova, Esteban Porrini, Farshad Pourmalek, Martin Prince, Rachel L Pullan, Kapa D Ramaiah, Dharani Ranganathan, Homie Razavi, Mathilda Regan, Jürgen T Rehm, David B Rein, Guiseppe Remuzzi, Kathryn Richardson, Frederick P Rivara, Thomas Roberts, Carolyn Robinson, Felipe Rodriguez De Leòn, Luca Ronfani, Robin Room, Lisa C Rosenfeld, Lesley Rushton, Ralph L Sacco, Sukanta Saha, Uchechukwu Sampson, Lidia Sanchez-Riera, Ella Sanman, David C Schwebel, James Graham Scott, Maria Segui-Gomez, Saeid Shahraz, Donald S Shepard, Hwashin Shin, Rupak Shivakoti, David Singh, Gitanjali M Singh, Jasvinder a Singh, Jessica Singleton, David a Sleet, Karen Sliwa, Emma Smith, Jennifer L Smith, Nicolas J C Stapelberg, Andrew Steer, Timothy Steiner, Wilma a Stolk, Lars Jacob Stovner, Christopher Sudfeld, Sana Syed, Giorgio Tamburlini, Mohammad Tavakkoli, Hugh R Taylor, Jennifer a Taylor, William J Taylor, Bernadette Thomas, W Murray Thomson, George D Thurston, Imad M Tleyjeh, Marcello Tonelli, Jeffrey a Towbin, Thomas Truelsen, Miltiadis K Tsilimbaris, Clotilde Ubeda, Eduardo a Undurraga, Marieke J van der Werf, Jim van Os, Monica S Vavilala, N Venketasubramanian, Mengru Wang, Wenzhi Wang, Kerrianne Watt, David J Weatherall, Martin a Weinstock, Robert Weintraub, Marc G Weisskopf, Myrna M Weissman, Richard a White, Harvey Whiteford, Natasha Wiebe, Steven T Wiersma, James D Wilkinson, Hywel C Williams, Sean R M Williams, Emma Witt, Frederick Wolfe, Anthony D Woolf, Sarah Wulf, Pon-Hsiu Yeh, Anita K M Zaidi, Zhi-Jie Zheng, David Zonies, Alan D Lopez, Mohammad a AlMazroa, and Ziad a Memish. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380(9859):2197–2223, dec 2012.

[141] Masatoshi Nei, Yoshihito Niimura, and Masafumi Nozawa. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature reviews.*

Genetics, 9(12):951–63, dec 2008.

[142] Philip A Newmark and Alejandro Sánchez Alvarado. Not your father's planarian: a classic model enters the era of functional genomics. *Nature reviews. Genetics*, 3(3):210–9, mar 2002.

[143] Osamu Nishimura, Yukako Hirao, Hiroshi Tarui, and Kiyokazu Agata. Comparative transcriptome analysis between planarian Dugesia japonica and other platyhelminth species. *BMC genomics*, 13(1):289, jan 2012.

[144] Taisaku Nogi, Dan Zhang, John D Chan, and Jonathan S Marchant. A novel biological activity of praziquantel requiring voltage-operated Ca2+ channel beta subunits: subversion of flatworm regenerative polarity. *PLoS neglected tropical diseases*, 3(6):e464, 2009.

[145] Jaime H Noguez, Elizabeth S Conner, Yue Zhou, Todd A Ciche, Justin R Ragains, and Rebecca A Butcher. A novel ascaroside controls the parasitic life cycle of the entomopathogenic nematode Heterorhabditis bacteriophora. *ACS chemical biology*, 7(6):961–6, jun 2012.

[146] L P Noldus, a J Spink, and R a Tegelenbosch. EthoVision: a versatile video tracking system for automation of behavioral experiments. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, 33(3):398–414, aug 2001.

[147] B. J. Nolen, N. Tomasevic, A. Russell, D. W. Pierce, Z. Jia, C. D. McCormick, J. Hartman, R. Sakowicz, and T. D. Pollard. Characterization of two classes of small molecule inhibitors of Arp2/3 complex. *Nature*, 460(7258):1031–1034, aug 2009.

[148] Alan Fenwick OBE. Praziquantel:do we need another antischistosoma treatment? *Future Medicinal Chemistry*, 7(6):677–680, apr 2015.

[149] F. Olaechea, V. Lovera, M. Larroza, F. Raffo, and R. Cabrera. Resistance of Fasciola hepatica against triclabendazole in cattle in Patagonia (Argentina). *Veterinary Parasitology*, 178(3-4):364–366, jun 2011.

[150] Hanan H Omar, Judith E Humphries, Martha J Larsen, Teresa M Kubiak, Timothy G Geary, Aaron G Maule, Michael J Kimber, and Tim A Day. Identification of a platyhelminth neuropeptide receptor. *Int. J. Parasitol.*, 37(7):725–733, jun 2007.

[151] E A Ottesen, M M Ismail, J Horton, E.A. Ottesen, M.M. Ismail, and J. Horton. The role of albendazole in programmes to eliminate lymphatic filariasis. *Parasitology today (Personal ed.)*, 15(9):382–6, sep 1999.

[152] John P. Overington, Bissan Al-Lazikani, and Andrew L. Hopkins. How many drug targets are there? *Nature Reviews Drug Discovery*, 5(12):993–996, dec 2006.

[153] Joanna Owens. Determining druggability. *Nature Reviews Drug Discovery*, 6(3):187–187, mar 2007.

[154] O R Pagán, S Deats, D Baker, E Montgomery, G Wilk, M Tenaglia, and J Semon. Planarians require an intact brain to behaviorally react to cocaine, but not to react to nicotine. *Neuroscience*, 246:265–270, aug 2013.

[155] Qiuxiang Pang, Lili Gao, Wenjing Hu, Yang An, Hongkuan Deng, Yichao Zhang, Xiaowen Sun, Guangzhong Zhu, Baohua Liu, and Bosheng Zhao. De Novo Transcriptome Analysis Provides Insights into Immune Related Genes and the RIG-I-Like Receptor Signaling Pathway in the Freshwater Planarian (Dugesia japonica). *PLOS ONE*, 11(3):e0151597, mar 2016.

[156] P K Papasaikas, P G Bagos, Z I Litou, and S J Hamodrakas. A novel method for {GPCR} recognition and family classification from sequence alone using signatures derived from profile hidden Markov models. *SAR QSAR Environ. Res.*, 14(5-6):413–420, oct 2003.

[157] Valeriy M Paramonov, Veronika Mamaeva, Cecilia Sahlgren, and Adolfo Rivero-Müller. Genetically-encoded tools for cAMP probing and modulation in living systems. *Frontiers in pharmacology*, 6:196, 2015.

[158] Nicholas Patocka and Paula Ribeiro. The functional role of a serotonin transporter

in Schistosoma mansoni elucidated through immunolocalization and RNA interference (RNAi). *Molecular and biochemical parasitology*, 187(1):32–42, jan 2013.

[159] Nicholas Patocka, Nidhi Sharma, Mohammed Rashid, and Paula Ribeiro. Serotonin signaling in Schistosoma mansoni: a serotonin-activated G protein-coupled receptor controls parasite movement. *PLoS pathogens*, 10(1):e1003878, jan 2014.

[160] Julien Pelé, Pelé Julien, Abdi Hervé, Moreau Matthieu, Thybert David, and Chabbert Marie. Multidimensional Scaling Reveals the Main Evolutionary Pathways of Class A {G-Protein-Coupled} Receptors. *PLoS One*, 6(4):e19094, 2011.

[161] A. M. PHIRI, D. I. DE POMERAI, D. J. BUTTLE, and J. M. BEHNKE. A novel assay for the detection of anthelmintic activity mediated by cuticular damage to nematodes: validation on Caenorhabditis elegans exposed to cysteine proteinases. *Parasitology*, pages 1–11, jan 2017.

[162] L Pica-Mattoccia and D Cioli. Studies on the mode of action of oxamniquine and related schistosomicidal drugs. *The American journal of tropical medicine and hygiene*, 34(1):112–8, jan 1985.

[163] Livia Pica-Mattoccia, Daniele Carlini, Alessandra Guidi, Velasco Cimica, Fabio Vigorosi, and Donato Cioli. The schistosome enzyme that activates oxamniquine has the characteristics of a sulfotransferase. *Memórias do Instituto Oswaldo Cruz*, 101:307–312, oct 2006.

[164] Roger K. Prichard, María-Gloria Basáñez, Boakye A. Boatin, James S. McCarthy, Héctor H. García, Guo-Jing Yang, Banchob Sripa, and Sara Lustigman. A Research Agenda for Helminth Diseases of Humans: Intervention for Control and Elimination. *PLoS Neglected Tropical Diseases*, 6(4):e1549, apr 2012.

[165] Anna V. Protasio, Isheng J. Tsai, Anne Babbage, Sarah Nichol, Martin Hunt, Martin A. Aslett, Nishadi de Silva, Giles S. Velarde, Tim J C Anderson, Richard C. Clark, Claire Davidson, Gary P. Dillon, Nancy E. Holroyd, Philip T. LoVerde, Christine Lloyd,

Jacquelline McQuillan, Guilherme Oliveira, Thomas D. Otto, Sophia J. Parker-Manuel, Michael A. Quail, R. Alan Wilson, Adhemar Zerlotini, David W. Dunne, and Matthew Berriman. A systematically improved high quality genome and transcriptome of the human blood fluke Schistosoma mansoni. *PLoS Neglected Tropical Diseases*, 6(1), 2012.

[166] Kim D Pruitt, Garth R Brown, Susan M Hiatt, Françoise Thibaud-Nissen, Alexander Astashyn, Olga Ermolaeva, Catherine M Farrell, Jennifer Hart, Melissa J Landrum, Kelly M McGarvey, Michael R Murphy, Nuala a O'Leary, Shashikant Pujar, Bhanu Rajput, Sanjida H Rangwala, Lillian D Riddick, Andrei Shkeda, Hanzhen Sun, Pamela Tamez, Raymond E Tully, Craig Wallin, David Webb, Janet Weber, Wendy Wu, Michael DiCuccio, Paul Kitts, Donna R Maglott, Terence D Murphy, and James M Ostell. RefSeq: an update on mammalian reference sequences. *Nucleic acids research*, 42(Database issue):D756—-63, jan 2014.

[167] Kriti Rakesh, ByungSu Yoo, Il-Man Kim, Natasha Salazar, Ki-Seok Kim, and Howard A Rockman. beta-Arrestin-biased agonism of the angiotensin receptor induced by mechanical stress. *Science signaling*, 3(125):ra46, 2010.

[168] Latha Ramakrishnan, Christina Amatya, Cassie J DeSaer, Zachary Dalhoff, and Michael R Eggerichs. Galantamine reverses scopolamine-induced behavioral alterations in Dugesia tigrina. *Invertebrate neuroscience : IN*, 14(2):91–101, sep 2014.

[169] Peter W Reddien, Adam L Bermange, Kenneth J Murfitt, Joya R Jennings, and Alejandro Sánchez Alvarado. Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Developmental cell*, 8(5):635–49, may 2005.

[170] Michael R. Reich, Ramesh Govindaraj, Karin Dumbaugh, Bong-min Yang, Agnes Brinkmann, and Sameh El-Saherty. International strategies for tropical disease treatments: Experiences with praziquantel. Technical report, World Health Organization, 1998.

[171] Michael Remmert, Andreas Biegert, Andreas Hauser, and Johannes Söding. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*, 9(2):173–175, dec 2011.

[172] Alissa M. Resch, Dasaradhi Palakodeti, Yi-Chien Lu, Michael Horowitz, and Brenton R. Graveley. Transcriptome Analysis Reveals Strain-Specific and Conserved Stemness Genes in Schmidtea mediterranea. *PLoS ONE*, 7(4):e34447, apr 2012.

[173] H. C. RICHARDS and R. FOSTER. A New Series of 2-Aminomethyltetrahydroquinoline Derivatives displaying Schistosomicidal Activity in Rodents and Primates. *Nature*, 222(5193):581–582, may 1969.

[174] Santiago Rios, Marta F Fernandez, Gianluigi Caltabiano, Mercedes Campillo, Leonardo Pardo, and Angel Gonzalez. GPCRtm: An amino acid substitution matrix for the transmembrane region of class A G Protein-Coupled Receptors. *BMC bioinformatics*, 16:206, jan 2015.

[175] Syed A. Rizvi, Erin M. Neidt, Jiayue Cui, Zach Feiger, Colleen T. Skau, Margaret L. Gardel, Sergey A. Kozmin, and David R. Kovar. Identification and Characterization of a Small Molecule Inhibitor of Formin-Mediated Actin Assembly. *Chemistry & Biology*, 16(11):1158–1168, nov 2009.

[176] S. M.C. Robb, E. Ross, and A. S. Alvarado. SmedGD: the Schmidtea mediterranea genome database. *Nucleic Acids Research*, 36(Database):D599–D606, dec 2007.

[177] Sofia M C Robb, Eric Ross, and Alejandro Sa. SmedGD : the Schmidtea mediterranea genome database. *Nucleic acids research*, 36:599–606, 2008.

[178] Sofia M.C. Robb, Kirsten Gotting, Eric Ross, and Alejandro Sánchez Alvarado. SmedGD 2.0: The Schmidtea mediterranea genome database. *genesis*, 53(8):535–546, aug 2015.

[179] Rachel H Roberts-Galbraith, John L Brubacher, and Phillip A Newmark. A functional

genomics screen in planarians reveals regulators of whole-brain regeneration. *eLife*, 5, sep 2016.

[180] Hugh M Robertson and James H Thomas. The putative chemoreceptor families of C. elegans *.

[181] Hugh M Robertson, Coral G Warr, and John R Carlson. Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, 100 Suppl:14537–42, nov 2003.

[182] Fredrik Ronquist, Maxim Teslenko, Paul van der Mark, Daniel L Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc A Suchard, and John P Huelsenbeck. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3):539–42, may 2012.

[183] Labib Rouhana, Jennifer A Weiss, David J Forsthoefel, Hayoung Lee, Ryan S King, Takeshi Inoue, Norito Shibata, Kiyokazu Agata, and Phillip A Newmark. RNA interference by feeding in vitro-synthesized double-stranded RNA to planarians: methodology and dynamics. *Developmental dynamics : an official publication of the American Association of Anatomists*, 242(6):718–30, jun 2013.

[184] Amir Saberi, Ayana Jamal, Isabel Beets, Liliane Schoofs, and Phillip A. Newmark. GPCRs Direct Germline Development and Somatic Gonad Function in Planarians. *PLOS Biology*, 14(5):e1002457, may 2016.

[185] Amir Saberi, Ayana Jamal, Isabel Beets, Liliane Schoofs, and Phillip A Newmark. {GPCRs} Direct Germline Development and Somatic Gonad Function in Planarians. *PLoS Biol.*, 14(5):e1002457, may 2016.

[186] B Saeger, H P Schmitt-Wrede, M Dehnhardt, W P Benten, J Krücken, A Harder, G Von Samson-Himmelstjerna, H Wiegand, and F Wunderlich. Latrophilin-like receptor from the parasitic nematode Haemonchus contortus as target for the anthelmintic depsipeptide PF1022A. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 15(7):1332–4, may 2001.

[187] M Sajid, J Rogers, and M Rajandream. The genome of the blood fluke Schistosoma mansoni. *Nature*, 460(7253):352–358, 2009.

[188] David Salom, Pengxiu Cao, Wenyu Sun, Kristopher Kramp, Beata Jastrzebska, Hui Jin, Zhaoyang Feng, and Krzysztof Palczewski. Heterologous expression of functional G-protein-coupled receptors in Caenorhabditis elegans. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 26(2):492–502, feb 2012.

[189] Philipp H Schiffer, Michael Kroiher, Christopher Kraus, Georgios D Koutsovoulos, Sujai Kumar, Julia I R Camps, Ndifon A Nsah, Dominik Stappert, Krystalynne Morris, Peter Heger, Janine Altmüller, Peter Frommolt, Peter Nürnberg, W Thomas, Mark L Blaxter, Einhard Schierenberg, M Liu, JP Mccarter, J Appleton, ER Mardis, RK Wilson, M Berriman, P Leonetti, M Magliano, TR Maier, GV Markov, P Mcveigh, G Pesole, J Poulain, and M Robinson-Rechavi. The genome of Romanomermis culicivorax: revealing fundamental changes in the core developmental genetic toolkit in Nematoda. *BMC Genomics*, 14(1):923, 2013.

[190] Caroline a Schneider, Wayne S Rasband, and Kevin W Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, jun 2012.

[191] Marcel H Schulz, Daniel R Zerbino, Martin Vingron, and Ewan Birney. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 28(8):1086–1092, apr 2012.

[192] M Lucila Scimone, Kellie M Kravarik, Sylvain W Lapan, and Peter W Reddien. Neoblast Specialization in Regeneration of the Planarian Schmidtea mediterranea. *Stem Cell Reports*, 3(2):339–352, aug 2014.

[193] W Evan Secor and Susan P Montgomery. Something old, something new: is praziquantel enough for schistosomiasis control? *Future medicinal chemistry*, 7(6):681–4, 2015.

[194] Katherine A Smith, Elizabeth B Rex, and Richard W Komuniecki. Are Caenorhabditis elegans receptors useful targets for drug discovery: pharmacological comparison of tyra-

mine receptors with high identity from C. elegans (TYRA-2) and Brugia malayi (Bm4). *Molecular and biochemical parasitology*, 154(1):52–61, jul 2007.

[195] Martin W Smith, Tracy L Borts, Renee Emkey, Carolyn A Cook, Christina J Wiggins, and Jesus A Gutierrez. Characterization of a novel G-protein coupled receptor from the parasitic nematode H. contortus with high affinity for serotonin. *Journal of neurochemistry*, 86(1):255–66, jul 2003.

[196] Nicola J. Smith and Graeme Milligan. Allostery at G Protein-Coupled Receptor Homo- and Heteromers: Uncharted Pharmacological Landscapes. *Pharmacological Reviews*, 62(4), 2010.

[197] Stephen A Smith and James B Pease. Heterogeneous molecular processes among the causes of how sequence similarity scores can fail to recapitulate phylogeny. *Briefings in bioinformatics*, page bbw034, apr 2016.

[198] Temple F. Smith and Xiaolin Zhang. The challenges of genome sequence annotation or The devil is in the details. *Nature Biotechnology*, 15(12):1222–1223, nov 1997.

[199] Jordi Solana, Damian Kao, Yuliana Mihaylova, Farah Jaber-Hijazi, Sunir Malla, Ray Wilson, and Aziz Aboobaker. Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNAseq, RNA interference and irradiation approach. *Genome biology*, 13(3):R19, jan 2012.

[200] Stephanie J Spielman and Claus O Wilke. Membrane environment imposes unique selection pressures on transmembrane domains of {G} protein-coupled receptors. *J. Mol. Evol.*, 76(3):172–182, mar 2013.

[201] Peter Steinmann, Jennifer Keiser, Robert Bos, Marcel Tanner, and Jürg Utzinger. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases*, 6(7):411–425, jul 2006.

[202] Jennifer A Stockert and Lakshmi A Devi. Advancements in therapeutically targeting orphan GPCRs. *Frontiers in pharmacology*, 6:100, jan 2015.

[203] Amira Taman, Taman Amira, and Ribeiro Paula. Characterization of a Truncated Metabotropic Glutamate Receptor in a Primitive Metazoan, the Parasitic Flatworm Schistosoma mansoni. *PLoS One*, 6(11):e27119, 2011.

[204] Amira Taman and Paula Ribeiro. Investigation of a dopamine receptor in Schistosoma mansoni: Functional studies and immunolocalization. *Molecular and Biochemical Parasitology*, 168(1):24–33, nov 2009.

[205] Amira Taman and Paula Ribeiro. Glutamate-mediated signaling in Schistosoma mansoni: A novel glutamate receptor is expressed in neurons and the female reproductive tract. *Molecular and Biochemical Parasitology*, 176(1):42–50, 2011.

[206] Xiao-long Tang, Ying Wang, Da-li Li, Jian Luo, and Ming-yao Liu. Orphan G protein-coupled receptors (GPCRs): biological functions and potential drug targets. *Acta Pharmacologica Sinica*, 33(3):363–371, 2012.

[207] Christina M. Taylor, Qi Wang, Bruce A. Rosa, Stanley Ching-Cheng Huang, Kerrie Powell, Tim Schedl, Edward J. Pearce, Sahar Abubucker, Makedonka Mitreva, E van den Enden, L Holden-Dye, RJ Walker, MY Osei-Atweneboana, JK Eng, DA Boakye, JO Gyapong, RK Prichard, S Geerts, B Gryseels, GJ Weil, W Kastens, M Susapu, SJ Laney, SA Williams, KR Barker, RS Hussey, LR Krusberg, GW Bird, RA Dunn, IA Zasada, JM Halbrendt, N Kokalis-Burelle, J Lamondia, MV McKenry, I Yeh, T Hanekamp, S Tsoka, PD Karp, RB Altman, MC Palumbo, A Colosimo, A Giuliani, L Farina, S Singh, BK Malik, DK Sharma, D Perumal, CS Lim, KR Sakharkar, MK Sakharkar, SA Rahman, D Schomburg, S Fatumo, K Plaimas, JP Mallm, G Schramm, E Adebiyi, TY Kim, HU Kim, SY Lee, PJ Brindley, M Mitreva, E Ghedin, S Lustigman, M Mitreva, DS Zarlenga, JP McCarter, DP Jasmer, ML Blaxter, P De Ley, JR Garey, LX Liu, P Scheldeman, M Kanehisa, S Goto, M Furumichi, M Tanabe, M Hirakawa, DS Wishart, C Knox, AC Guo, D Cheng, S Shrivastava, E Ghedin, S Wang, D Spiro, E Caler, Q Zhao, CH Opperman, DM Bird, VM Williamson, DS Rokhsar, M Burke, P Abad, J Gouzy, JM Aury, P Castagnone-Sereno, EG Danchin,

C Dieterich, SW Clifton, LN Schuster, A Chinwalla, K Delehaunty, M Mitreva, DP Jasmer, DS Zarlenga, Z Wang, S Abubucker, DJ Brownlee, L Holden-Dye, RJ Walker, DP Jasmer, C Yao, A Rehman, S Johnson, CL Charvet, AP Robertson, J Cabaret, RJ Martin, C Neveu, P Shannon, A Markiel, O Ozier, NS Baliga, JT Wang, CA Lipinski, F Lombardo, BW Dominy, PJ Feeney, SC Huang, TC Freitas, E Amiel, B Everts, EL Pearce, A Morgulis, EM Gertz, AA Schaffer, R Agarwala, C Trapnell, L Pachter, SL Salzberg, LW Hillier, V Reinke, P Green, M Hirst, MA Marra, MD Robinson, DJ McCarthy, GK Smyth, EM Zdobnov, R Apweiler, M Goujon, H McWilliam, W Li, F Valentin, S Squizzato, M Ashburner, CA Ball, JA Blake, D Botstein, H Butler, Y Benjamini, Y Hochberg, NM O'Boyle, M Banck, CA James, C Morley, T Vandermeersch, HB Schioth, R Fredriksson, GM Morris, R Huey, W Lindstrom, MF Sanner, RK Belew, YS Hsiao, G Jogl, V Esser, L Tong, J Martin, S Abubucker, E Heizer, CM Taylor, M Mitreva, JD McGarry, NF Brown, IO Zubovych, S Straud, MG Roth, RR Ramsay, VA Zammit, EA Park, GA Cook, GM Morris, R Huey, W Lindstrom, MF Sanner, RK Belew, J Noroes, G Dreyer, A Santos, VG Mendes, Z Medeiros, O Doumbo, JF Rossignol, E Pichard, HA Traore, TM Dembele, WC Campbell, MH Fisher, EO Stapley, G Albers-Schonberg, TA Jacob, RJ Martin, AP Robertson, H Bjorn, RK Prichard, RK Prichard, M Berriman, BJ Haas, PT LoVerde, RA Wilson, GP Dillon, RB Giordani, M Weizenmann, DB Rosemberg, GA De Carli, MR Bogo, FM Sansom, SC Robson, EL Hartland, K Gounaris, S Thomas, P Najarro, ME Selkirk, I Ghosh, N Raghavan, PC FitzGerald, AL Scott, MS Sajid, Z Iqbal, G Muhammad, MU Iqbal, TW Harris, I Antoshechkin, T Bieri, D Blasiar, J Chan, M Tanaka, J Gilbert, AJ Pappano, T Haruna, M Horie, M Takano, Y Kono, H Yoshida, JA Kennedy, SA Unger, JD Horowitz, BA Catto, M Boutros, J Ahringer, W Qu, C Ren, Y Li, J Shi, J Zhang, F Simmer, C Moorman, AM van der Linden, E Kuijk, PV van den Berghe, M Chan, DS Tan, TS Sim, R Zoraghi, RH See, P Axerio-Cilies, NS Kumar, H Gong, PD van Poelje, SC Potter, MD Erion, A Melnikov, O Zaborina, N Dhiman, BS Prabhakar, AM Chakrabarty, K Yuasa, F Mi-Ichi, T Kobayashi, M Yamanouchi, J Kotera, Y Shakur, HP de Koning, H Ke, J Kambayashi, T Seebeck, FS Menniti, WS Faraci, CJ Schmidt, A Sanon, C Tournaire-Arellano,

SY El Hage, C Bories, R Caujolle, AD Villela, ZA Sanchez-Quitian, RG Ducati, DS Santos, LA Basso, L Kucerova, M Matuskova, A Pastorakova, S Tyciakova, J Jakubikova, CH Lin, MY Chung, WB Chen, CH Chien, RL Bateman, P Bhanumoorthy, JF Witte, RW McClard, and M Grompe. Discovery of Anthelmintic Drug Targets and Drugs Using Chokepoints in Nematode Metabolic Pathways. *PLoS Pathogens*, 9(8):e1003505, aug 2013.

[208] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[209] Michelle S Teng, Martijn PJ Dekkers, Bee Ng, Suzanne Rademakers, Gert Jansen, Andrew G Fraser, and John McCafferty. Expression of mammalian GPCRs in C. elegans generates novel behavioural responses to human ligands. *BMC Biology*, 4(1):22, jul 2006.

[210] O. Thomas, et Al., Travis Beddoe, Glenn Anderson, Philip Skuce, and Terry W. Spithill. Triclabendazole resistant Fasciola hepatica in south-west Wales. *Vet. Rec.*, 146(6):200, 2000.

[211] Els Torreele, Martine Usdin, Pierre Chirac, By Jayasree K Iyer, Emily R Adams, and Paul R Klatser. Priority Medicines for Europe and the World: A Public Health Approach to Innovation. 2004.

[212] Santiago Trainor-Moss and Francisca Mutapi. Schistosomiasis therapeutics: whats in the pipeline? *Expert Review of Clinical Pharmacology*, 9(2):157–160, feb 2016.

[213] Cole Trapnell, Lior Pachter, and Steven L Salzberg. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, 25(9):1105–11, may 2009.

[214] Isheng J Tsai, Magdalena Zarowiecki, Nancy Holroyd, Alejandro Garciarrubio, Alejandro Sanchez-Flores, Karen L Brooks, Alan Tracey, Raúl J Bobes, Gladis Fragoso, Edda Sciutto, Martin Aslett, Helen Beasley, Hayley M Bennett, Jianping Cai, Federico Camicia, Richard Clark, Marcela Cucher, Nishadi De Silva, Tim A Day, Peter Deplazes, Karel Estrada, Cecilia Fernández, Peter W H Holland, Junling Hou, Songnian Hu, Thomas

Huckvale, Stacy S Hung, Laura Kamenetzky, Jacqueline A Keane, Ferenc Kiss, Uriel Koziol, Olivia Lambert, Kan Liu, Xuenong Luo, Yingfeng Luo, Natalia Macchiaroli, Sarah Nichol, Jordi Paps, John Parkinson, Natasha Pouchkina-Stantcheva, Nick Riddiford, Mara Rosenzvit, Gustavo Salinas, James D Wasmuth, Mostafa Zamanian, Yadong Zheng, Xuepeng Cai, Xavier Soberón, Peter D Olson, Juan P Laclette, Klaus Brehm, and Matthew Berriman. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*, 496(7443):57–63, apr 2013.

[215] G. E. Tusnady and I. Simon. The HMMTOP transmembrane topology prediction server. *Bioinformatics*, 17(9):849–850, sep 2001.

[216] C. L. L. Valentim, D. Cioli, F. D. Chevalier, X. Cao, A. B. Taylor, S. P. Holloway, L. Pica-Mattoccia, A. Guidi, A. Basso, I. J. Tsai, M. Berriman, C. Carvalho-Queiroz, M. Almeida, H. Aguilar, D. E. Frantz, P. J. Hart, P. T. LoVerde, and T. J. C. Anderson. Genetic and Molecular Basis of Drug Resistance and Species-Specific Drug Action in Schistosome Parasites. *Science*, 342(6164):1385–1389, dec 2013.

[217] Marieke J van der Werf, Sake J de Vlas, Simon Brooker, Caspar W N Looman, Nico J D Nagelkerke, J.Dik F Habbema, and Dirk Engels. Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Tropica*, 86(2-3):125–139, may 2003.

[218] Kevin J Vogel, Mark R Brown, and Michael R Strand. Phylogenetic investigation of Peptide hormone and growth factor receptors in five dipteran genomes. *Frontiers in endocrinology*, 4:193, jan 2013.

[219] Bo Wang, James J III, Collins, and Phillip A Newmark. Functional genomic characterization of neoblast-like stem cells in larval Schistosoma mansoni. *eLife*, 2:1–15, 2013.

[220] Wei Wang, Li Wang, and You-Sheng Liang. Susceptibility or resistance of praziquantel in human schistosomiasis: a review. *Parasitology research*, 111(5):1871–7, nov 2012.

[221] Yuying Wang, Joel M Stary, James E Wilhelm, and Phillip A Newmark. A functional genomic screen in planarians identifies novel regulators of germ cell development. *Genes & development*, 24(18):2081–92, sep 2010.

[222] Joanne P Webster, David H Molyneux, Peter J Hotez, and Alan Fenwick. The contribution of mass drug administration to global health: past, present and future. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1645):20130434, 2014.

[223] Vivian A Welch, Elizabeth Ghogomu, Alomgir Hossain, Shally Awasthi, Zulfiqar A Bhutta, Chisa Cumberbatch, Robert Fletcher, Jessie McGowan, Shari Krishnaratne, Elizabeth Kristjansson, Salim Sohani, Shalini Suresh, Peter Tugwell, Howard White, and George A Wells. Mass deworming to improve developmental health and wellbeing of children in low-income and middle-income countries: a systematic review and network meta-analysis. *The Lancet Global Health*, 5(1):e40–e50, jan 2017.

[224] Nicolas J Wheeler, Prince N Agbedanu, Michael J Kimber, Paula Ribeiro, Tim A Day, and Mostafa Zamanian. Functional analysis of Girardia tigrina transcriptome seeds pipeline for anthelmintic target discovery. *Parasites & vectors*, 8:34, jan 2015.

[225] Michael Whitaker. Genetically encoded probes for measurement of intracellular calcium. *Methods in cell biology*, 99:153–82, 2010.

[226] WHO. Schistosomiasis: population requiring preventive chemotherapy and number of people treated in 2010. *Weekly Epidemiological Record*, 87(4):37–44, 2012.

[227] WHO. Schistosomiasis: number of people treated worldwide in 2015. Technical report, 2016.

[228] WHO. WHO — WHO urges increased access to praziquantel as schistosomiasis treatment coverage improves, 2016.

[229] Hadley Wickham. ggplot2: Elegant Graphics for Data Analysis. aug 2009.

[230] Lothar Wissler, Jürgen Gadau, Daniel F Simola, Martin Helmkampf, and Erich Bornberg-Bauer. Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome biology and evolution*, 5(2):439–55, jan 2013.

[231] D YADEGARI, H TALAIE, and J MASSOUD. CLINICAL TRIAL OF TRICLABENDAZOLE ON HUMAN FASCIOLIASIS: LONG TERM FOLLOW UP. *Medical Journal of The Islamic Republic of Iran (MJIRI)*, 13(2):89–91, 1999.

[232] Mostafa Yakoot. A short review of the anthelmintic role of mirazid. *Arquivos de Gastroenterologia*, 47(4):393–394, dec 2010.

[233] Guangchuang Yu, David K. Smith, Huachen Zhu, Yi Guan, and Tommy Tsan-Yuk Lam. 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, jan 2017.

[234] Mostafa Zamanian, Prince N. Agbedanu, Nicolas J. Wheeler, Paul McVeigh, Michael J. Kimber, and Tim A. Day. Novel RNAi-mediated approach to G protein-coupled receptor deorphanization: Proof of principle and characterization of a planarian 5-HT receptor. *PLoS ONE*, 7(7), 2012.

[235] Mostafa Zamanian and Erik C Andersen. Prospects and challenges of CRISPR/Cas genome editing for the study and control of neglected vector-borne nematode diseases. *The FEBS journal*, 283(17):3204–21, sep 2016.

[236] Mostafa Zamanian, Michael J Kimber, Paul McVeigh, Steve A Carlson, Aaron G Maule, and Tim A Day. The repertoire of G protein-coupled receptors in the human parasite Schistosoma mansoni and the model organism Schmidtea mediterranea. *BMC genomics*, 12:596, jan 2011.

[237] Daniel R Zerbino and Ewan Birney. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, 18(5):821–829, may 2008.

[238] Jianyong Zhang, Zuoqing Yuan, Mingyue Zheng, Yuqian Sun, Youjun Wang, and Shudong Yang. Effects of N,N-dimethylformamide on behaviour and regeneration of planarian Dugesia japonica. *Toxicology and industrial health*, 29(8):753–760, sep 2013.

[239] Min Zheng, Daniel Messerschmidt, Benno Jungblut, and Ralf J Sommer. Conservation and diversification of Wnt signaling function during the evolution of nematode vulva development. *Nature genetics*, 37(3):300–4, mar 2005.