Chapter 6:

Genomic Sequencing of 5 *Steinernema* Nematodes (Rhabditida: Steinernematidae): Insights Into the Evolution of Parasitism*

*This chapter is not yet published but is currently in preparation and in its current form is written solely by Adler Dillman. It will ultimately include additional authors when published, including Ali Mortazavi, Marissa Macchietto, Byron J. Adams, Paul W. Sternberg, and possibly other others as well. Some of the data presented herein will change prior to publication, as additional sequencing data is acquired.

Abstract

Nematodes are amazing animals, both old and diverse. Among their diversity are many plant and animal parasites, many of which negatively affect humans. However, not all parasitic nematodes are bad and some are currently being used as organic alternatives to chemical pesticides for controlling damaging insect pests. Although there are many insect-parasitic nematodes, the entomopathogenic nematodes are the best studied of these and are remarkably different in their lifestyle and in their particular parasitism. Herein I discuss the difference between entomopathogenic nematodes and other insect parasites and what makes them so interesting and useful.

Introduction

When first looking for projects to propose for my graduate thesis, I was intrigued by the jumping abilities of some species of *Steinernema* and had hoped to explore this behavior in the context of foraging and host seeking. In addition to host seeking, I am interested in understanding the architecture of parasitism within the genome and how the genome of a free-living nematode might differ from that of a parasite. If we were to look at overviews of 2 genomes, could we tell just by the genomic content that one belonged to a parasite? I also thought it would be neat to understand how some species of *Steinernema* are capable of jumping but not others. Is this due to differences in physical structure or musculature, or does the difference lie deeper and hidden at the molecular level? Though I knew it was an impossibly risky proposal, it seemed that by sequencing the genomes of jumpers and non-jumpers, we might learn something about what

facilitates this amazing behavior. Admittedly this was a naïve supposition, but the EPNs are potentially well-suited to answering the question of parasitic architecture in the genome, since they are easily cultured within a short generation time and can be synchronized as IJs in, in addition to the abundance of ecological and behavioral data, as previously presented. However, there are currently no EPN genomes publicly available. There are over 70 species in the genus *Steinernema*, making it somewhat difficult to decide which species to sequence [1, 2]. Ultimately this decision was made based on availability of material, usefulness in biological control, and their phylogenetic position within the genus. I have sequenced and begun annotation on 5 steinernematids: *S. carpocapsae*, *S. scapterisci*, *S. monticolum*, *S. feltiae*, and *S. glaseri*. Several of these taxa were also included in behavioral studies detailed in Chapter 4 of this thesis. Among these 5, *S. carpocapsae* and *S. scapterisci* are known to be capable of jumping and are considered ambush foragers [3–5]. *S. monticolum* is reported as being capable of jumping but is thought to employ an intermediate foraging strategy [3, 5–7]. Originally I had wanted to include this species in my behavioral assays described in Chapter 4, but this nematode is not a very good jumper and it was not practical to use it in jumping assays. *S. feltiae* is not capable of jumping but is commonly used in biological control and is thought to use an intermediate foraging strategy, leaning toward the cruising side of the continuum [3, 5]. *S. glaseri* is not capable of jumping and is a classic example of a cruise forager among steinernematids [3, 5]. Sequencing these species also meant we were using taxa from clades II, III, and V of the five clades identified within the genus [8, 9].

Materials and Methods

Strain culturing and maintenance of *Steinernema* sp. *S. carpocapsae* were from the inbred strain ALL [10–12]. *S. glaseri* were from the inbred NC strain [13]. *S. scapterisci* were inbred from the FL strain [14]. *S. feltiae* were from the inbred SN strain [15]. *S. monticolum* were inbred from the originally isolated strain from Korea [7]. All nematodes were cultured as previously described [12]. Briefly, 5 last instar *Galleria mellonella* larvae (American Cricket Ranch, Lakeside, CA) were placed in a 5 cm Petri dish with a 55 mm Whatman 1 filter paper acting as a pseudo-soil substrate in the bottom of the dish. ≤ 250 ml containing 500–1000 IJs suspended in water was evenly distributed on the filter paper. After 7–10 days the insect cadavers were placed on White traps [16]. *Steinernema glaseri* was placed onto a modified White trap containing plaster of Paris as previously described [17]. Emerging IJs were harvested and rinsed 3 times with water. *S. scapterisci* was also cultured by infecting house crickets and mole crickets using similar techniques. IJs were stored harvested and used to isolate either total genomic DNA or stage specific RNA. To obtain *S. carpocapsae* stage-specific RNA for embryo, L1, and adult stages, nematodes were grown on lipid agar plates inoculated overnight with *Xenorhabdus nematophila* cultures [18]. Fresh bacterial lawns were inoculated with IJs and given three days to develop and reproduce. After 3 days, all nematodes were harvested and bleached for synchronization, then harvested at the appropriate times for stage-specific material.

Isolation of DNA and RNA. Once harvested, nematodes were frozen at −80°C until used. To extract nucleic acid, the nematodes were thawed and refrozen two to three times

to facilitate breaking the tough cuticle before extracting either genomic DNA or bulk RNA. Genomic DNA was extracted using a Promega Wizard® genomic DNA purification kit and following the protocol described in that kit. The genomic DNA was then treated with RNase A for digestion of any RNAs present in the sample. Bulk RNA was extracted using a Trizol® extraction as previously described [19].

Genomic and RNA-Seq library construction. Genomic library was constructed using Illumina Paired End DNA Sample Preparation Kit according to the manufacturer's instructions. Briefly, 3 μg of genomic DNA were fragmented using nebulization. The fragments were end repaired, 3' adenylated and ligated to Illumina's paired end adaptors. The ligation products were size selected on an agarose gel to yield fragments of approximate length of 350 bp and PCR amplified to produce the finished library. For *S. carpocapsae*, we also made a jumping library with in insert fragment length of 2kb to help facilitate a better assembly [20]. RNA-Seq library was created from 10 μg of total RNA. mRNA was purified using Dynal magnetic oligo(dT) beads (Invitrogen) and fragmented with 40mM Tris-acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc buffer for 4 min at 94°C. First and second cDNA strands were synthesized using random primers and SuperScript II RT (Invitrogen), and RNaseH and DNA Pol I, respectively. The rest of the procedure was identical to that used for the genomic library preparation, except that the gel cut for the RNA-seq library was ~ 300 bp. Libraries were quantified using Qubit fluorometer (Invitrogen) and size distributions were verified using Agilent Bioanalyzer and the High Sensitivity DNA Kit. Libraries were sequenced on Illumina Genome Analyzer IIx sequencer in paired-end mode with the read length of 100 nt.

Genome assembly and annotation. Both the genomic and the mixed-stage transcriptome libraries were built, sequenced, assembled, filtered, and repeat-masked as previously described [21] using Velvet 1.0.9. Genome and RNA-seq reads will be submitted to the public database once the assembly is complete. Assembled cDNA was used to train Augustus 2.5 [22] for protein-coding gene finding. Separately, RNA-seq reads were mapped onto the genome using TopHat 1.3.1 [23], assembled intro transcripts using Cufflinks 1.2.0 [24] and merged with the Augustus annotations using the RABT method [25]. Candidate SNVs in the genome and transcriptome mapped reads were called using the SAMtools [26] pileup and varFilter options. Candidate SNVs in the transcriptome that fell within 5 bp of exon junctions were filtered out as likely splicing artifacts.

Orthology analyses. To study the evolution of gene families across nematodes, we used the available predicted protein datasets from WormBase release WS225 (www.wormbase.org)—*Brugia malayi*, *Caenorhabditis elegans*, *Meloidogyne hapla*, *Pristionchus pacificus*, and *Trichinella spiralis*. We also included the *Ascaris suum* and *Bursaphelenchus xylophilus* predicted proteome data sets from WormBase release WS229. For outgroup and comparative analysis we used the predicted protein datasets of the *Arabidopsis thaliana* (vGNOMON 7/9/07), *Drosophila melanogaster* (v10/30/11), *Homo sapiens* (v9/7/11), *Mus musculus* (v3/4/11), *Nasonia vitripennis* (v1.2), *Saccharomyces cerevisiae* (v2/3/11), and *Tribolium castaneum* (vTcas 3.0) genome projects, obtained from the NCBI/NIH repository (ftp://ftp.ncbi.nih.gov/genomes).

Version 1.4 of the OrthoMCL pipeline was used to cluster proteins into families of orthologous genes, with default settings and the BLAST parameters recommended in the OrthoMCL documentation [27].

Protein domain analyses. To evaluate the prevalence of protein domains in the proteome of *S. carpocapsae*, *S. scapterisci*, *S. monticolum*, *S. feltiae*, and *S. glaseri* and other species, we used the HMMscan program from the latest version of HMMER (3.0) software package, which implements probabilistic profile hidden Markov models [28]. We set our threshold *E*-value criterion at 10^{-6}, so that no known false-positive matches would be detected in assigning Pfam domain identities. We ran this analysis on the proteomes mentioned above and filtered out splice isoforms from the *C. elegans* proteome.

Gene tree analyses. Some protein families were further explored by evaluating gene trees, either with whole protein sequences or by protein domain sequences. To do these analyses we aligned protein sequences using MUSCLE [29]. Aligned protein sequences were then evaluated by distance analysis using the JTT matrix and a subsequent neighbor-joining tree was created using the PHYLIP software package version 3.68 [30].

Results and Discussion

The *Steinernema* species selected proved to be amenable to short-read sequencing technology, resulting in assembled genomes between 82 and 114 Mb in size with variable contig sizes (Table 1). With the quality expectation of newly sequenced genomes on the

rise, we have decided to do additional sequencing prior to publishing these genomes, but Table 6.1 clearly indicates the usability of these genomes. The quality should dramatically improve with additional sequencing. In addition to high N50 values for several of these genomes, an analysis of 458 core eukaryotic genes reveals these genomes are largely complete: 98.67% for *S. carpocapsae*, 97.13% for *S. scapterisci*, 96.68% *S. monticolum*, 97.57% for *S. feltiae*, and 97.13% for *S. glaseri* [31].

Genome	Size (Mb)	# scaffolds	Max scaffold	N50	# genes predicted
S. carpocapsae	85.66	8,470	890.8 kb	93.15 kb	27,706
S. scapterisci	82.54	16,412	479.7 kb	46.0 kb	31,939
S. monticolum	114.25	82,427	347.9 kb	8.0 kb	41,294
S. glaseri	93.83	28,194	261.2 kb	27.6 kb	34,109
S. feltiae	101.04	66,553	566.99 kb	18.2 kb	36,178

Table 6.1 | *Steinernema* Genome Statistics. This table lists the 5 *Steinernema* genome species that were sequenced and provides several statistics. The size of the assemblies are given in Mb. The number of pieces or scaffolds that are in the assembly are reported, the fewer the pieces the better. Ideally there would be one scaffold per chromosome; obviously these genomes are not in the same state as the *C. elegans* genome. The max scaffold gives the size of the largest scaffold of the assembly. The N50 statistic provides the size of at least half of the scaffolds in that assembly. For example, at least half of the *S. carpocapsae* scaffolds are 93.15kb or larger in size, meaning that each scaffold is likely a multigenic-sized piece, making this a good quality genome. Also listed is the number of predicted genes, which is artificially high due to splice isoforms but will get better as annotation continues.

In order to evaluate the potential differences and similarities in genome architecture between a parasite and a 'free-living' nematode, a comparison was made between *S. carpocapsae* and *C. elegans*. The Pfam database was used to assign protein domain families to each of the full proteomes of these species. It is possible for some

proteins to have no recognizable protein domains while others can have several. I compared the prevalence of protein domains between these two species to see if anything stands out as being more abundant in the parasite or more abundant in the non-parasite, thus giving me a starting place to identify underlying differences in life strategy (Figure 6.1). I find that the *C. elegans* genome has an abundance of G-protein-coupled receptor (GPCR) domains, including members of the Srh, Sri, Srd, Str, and Srj families. Many olfactory receptors in nematodes are known to be GPCRs, which sparked my interest, especially considering the host-seeking studies I had performed previously, described in Chapters 3 and 4 [12, 32–34]. It is striking that Srh and Sri are much more abundant in *C. elegans*, given what is known about their expression. A particular promoter sequence called the E-box has been shown to be enriched in Srh and Sri GPCRs, and it is thought that this promoter drives expression in the chemosensory ADL neuron, making these GPCRs likely olfactory receptors [35]. Other Pfam domains that are much more abundant in *C. elegans* include F-box and F-box associated domains, which are involved in protein-protein interactions (Figure 6.1). On the other hand, I find trypsin inhibitor, aspartyl protease, and trypsin domains to be much more prevalent in the *S. carpocapsae* genome. This is not particularly surprising, given the potential utility of proteases and protease inhibitors in affecting insect immunity and tissue digestion. One surprising finding is the abundance of Srt family GPCRs in *S. carpocapsae*. Evolutionary analysis of this family has been limited and almost nothing is known about their function, but their expression pattern seems consistent with a role in chemosensation. This is a promising and unexpected expansion that could prove interesting regarding host seeking among steinernematids.

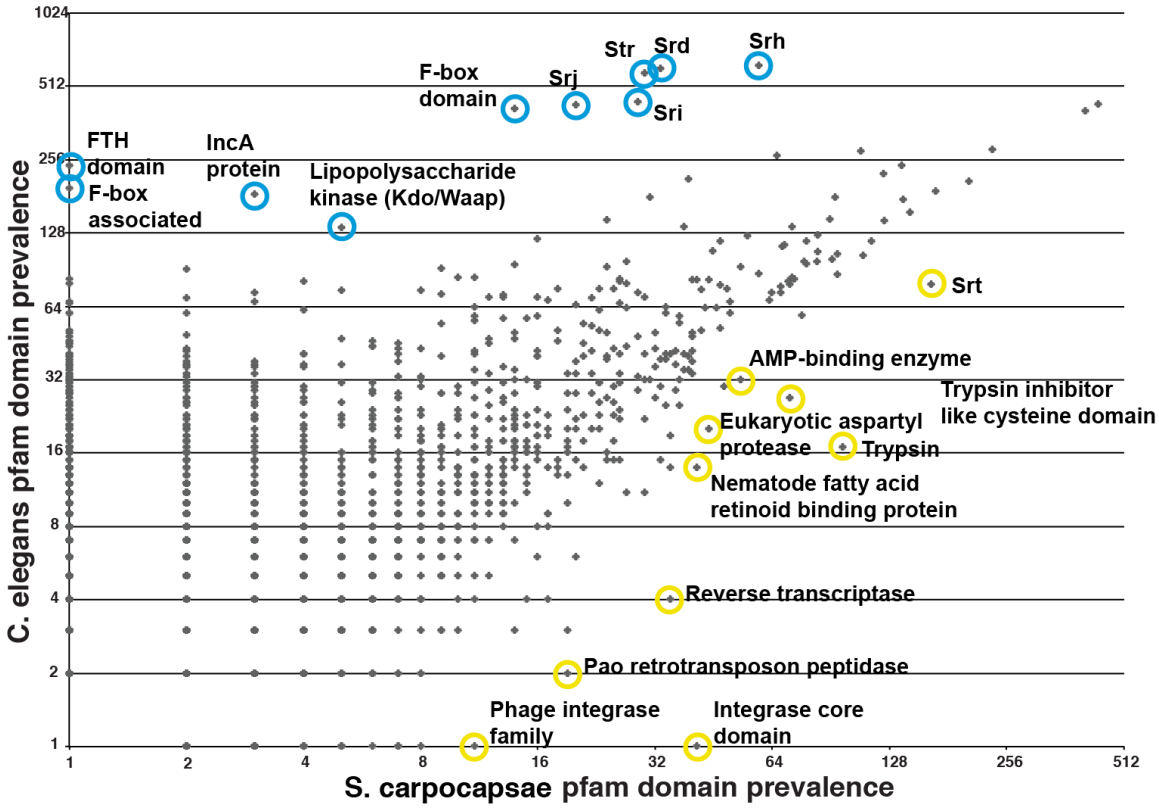

Figure 6.1 | Comparison of Pfam protein domain prevalence between *C. elegans* and *S. carpocapsae*. Protein domains that are in equal abundance in both species will show up on the diagonal axis, while those more abundant in *C. elegans* will cluster in the upper left and those more abundant in *S. carpocapsae* will appear toward the lower right. Several of the most divergently abundant protein domains have been highlighted in blue and yellow for those more abundant in *C. elegans* and *S. carpocapsae*, respectively.

These two findings, the abundance of GPCRs in *C. elegans* and the abundance of proteases and protease inhibitors, have shaped much of the rest of my genomic research, and I will discuss each in further detail below.

The abundance of GPCRs in *C. elegans* is interesting, since it is known that *C. elegans* is a fruit-dwelling nematode, not normally found in soil but in rotting fruit or

plant material. It spends its entire life cycle in a very complex environment avoiding predators while seeking resources and mates. It is unsurprising that an abundance of potential olfactory receptors would be useful for its lifestyle, but what of other nematodes? I searched for the abundance of potential olfactory receptors across available nematode genomes and found there is a common trend for GPCR abundance (Table 6.2).

	C. elegans	P. redivivus	B. xylophilus	P. pacificus	M. hapla	A. suum	B. malayi	T. spiralis	S. carpocapsae	S. scapterisci	S. monticolum	S. feltiae	S. glaseri
Number of proteins with multiple GPCR domains	1026	295	296	229	34	24	3	3	152	178	185	219	277
Number of proteins with only one GPCR domain	558	289	198	307	115	78	20	25	452	553	505	664	527
Total GPCR domain-containing proteins	**1654**	**584**	**494**	**536**	**149**	**102**	**23**	**28**	**604**	**731**	**690**	**883**	**804**

Table 6.2 | Total GPCRs Identified by Pfam Across Nematodes. This table lists total number of potential GPCRs as identified by a Pfam analysis.

	C. elegans	P. redivivus	B. xylophilus	P. pacificus	M. hapla	A. suum	B. malayi	T. spiralis	S. carpocapsae	S. scapterisci	S. monticolum	S. feltiae	S. glaseri
Serpentine type 7TM GPCR chemoreceptor Srh	625	208	221	154	2	9	0	0	58	90	90	200	220
Serpentine type 7TM GPCR chemoreceptor Srd	606	143	206	146	28	16	0	0	32	40	43	65	90
Serpentine type 7TM GPCR chemoreceptor Str	578	155	238	176	20	13	0	0	29	39	40	59	96
Serpentine type 7TM GPCR chemoreceptor Sri	439	127	90	61	0	1	0	0	28	27	26	74	84
Serpentine type 7TM GPCR chemoreceptor Srj	427	64	143	131	2	6	0	0	19	18	18	31	55
Serpentine type 7TM GPCR chemoreceptor Srx	241	33	23	50	9	20	9	2	136	154	172	201	172
Serpentine type 7TM GPCR chemoreceptor Srw	179	24	27	15	5	23	7	4	30	27	29	31	26
Serpentine type 7TM GPCR chemoreceptor Srsx	154	67	62	41	60	35	2	18	143	171	167	177	137
Srg family chemoreceptor	144	93	13	39	8	1	0	0	23	23	23	22	32
Serpentine type 7TM GPCR chemoreceptor Srbc	120	10	4	4	0	4	0	0	15	21	6	13	12
Serpentine type 7TM GPCR receptor class ab chemoreceptor	112	45	51	25	13	5	5	1	67	74	77	69	80
Sre G protein-coupled chemoreceptor	81	55	30	38	9	5	2	0	40	43	44	38	50
Serpentine type 7TM GPCR chemoreceptor Srv	80	49	10	36	1	2	0	4	25	38	31	31	45
Serpentine type 7TM GPCR chemoreceptor Srz	79	0	0	0	0	0	0	1	0	1	0	0	0
Serpentine type 7TM GPCR chemoreceptor Srt	78	39	33	35	21	3	0	1	163	192	153	191	138
Serpentine type 7TM GPCR chemoreceptor Sra	69	4	6	5	2	1	3	0	12	15	19	11	15
Serpentine type 7TM GPCR chemoreceptor Sru	68	1	0	1	1	1	0	0	1	3	2	1	12
Serpentine type 7TM GPCR chemoreceptor Srb	46	0	0	4	0	1	0	0	11	11	15	6	11
Serpentine receptor-like protein, class xa	41	1	0	0	0	0	0	0	1	1	2	1	1
Total Serpentine GPCR Domains	**4167**	**1128**	**1157**	**961**	**191**	**146**	**28**	**31**	**833**	**989**	**957**	**1221**	**1276**

Table 6.3 | All GPCR Families Identified by Pfam Across Nematodes. This table lists the number of all potential GPCRs as identified by a Pfam analysis and categorizes them by family [36, 37].

Nematode species that spend little to none of their foraging time in complex soil environments, such as the passively ingested vertebrate parasites *Ascaris suum*, *Brugia malayi*, and *Trichinella spiralis* have very few potential olfactory receptors (as GPCRs). While species that have free-living stages or forage in complex environments, including plant parasites, insect parasites, and free-living nematodes, have an abundance of potential olfactory receptors (Table 6.2).

In addition to looking at the total number of potential olfactory receptors, breaking these down into their respective families provides additional details about which GPCRs are highly conserved across all nematodes and which families have been expanded for particular use among the different lineages (Table 6.3). This analysis reveals that Srx, Srw, and Srsx GPCRs are the most highly conserved numerically across nematodes. Finding that Srsx GPCRs are conserved across nematodes is not a surprising finding and agrees with previous research indicating that the Srsx family of GPCRs seems evolutionarily stable [37]. This analysis also reveals that an abundance of Srt GPCRs is common among all of the steinernematids we sequenced and potentially all steinernematids. I suggest that it is the Srt GPCRs among *Steinernema* that have led to host preferences and specializations within the genus and that they merit further investigation. This analysis also reveals that many of the GPCRs in *C. elegans* are unique to it or the *Caenorhabditis* lineage and are not shared among other nematodes, such as the abundance of Sri, Srj, Srx, and Srw GPCRs (Table 6.3). The general trend that GPCR abundance seems to correlate with environmental foraging can also be observed within the steinernematids, with intermediate and cruise foraging seeming to have more potential olfactory receptors than ambush foragers, a trend that should be explored in

further detail (Tables 6.2 and 6.3). I was able to explore this trend in somewhat more detail, at least informatically. The Srh, Srd, Str, Sri, and Srj families of GPCRs are at least two-fold more abundant in intermediate and cruise foragers *S. feltiae* and *S. glaseri* than they are in the well-known ambushers *S. carpocapsae* and *S. scapterisci* (Table 6.3). I constructed a gene tree including all identified Srd GPCRs to examine the evolutionary dynamics of this particular gene family (Figure 6.2). Although you won't be able to read the names of individual Srd genes, by looking at the colors a trend is easily observed. There are regions where all five colors seem well represented, likely indicating conserved GPCRs across the species and other regions of abundant red or purple, indicating expansions in *S. glaseri* and *S. feltiae*, respectively. It appears that evolution is driving the expansion of this GPCR family among the cruise foragers, indicating that this GPCR family may be involved in the different odor preferences of these nematodes.

It is known that EPNs can affect their host's immune response, and several researchers have already implicated a handful of *Steinernema* proteases and protease inhibitors as influencing the insect immune system. However, the full complement of these proteins and their modes of action still remains unknown [13, 38–43]. Using the agnostic approach described above to identify differences in protein abundances in the steinernematids, I was led to explore in more detail the abundance of proteases among these species and which families seem to be expanded. I find that the metalloproteases and serine proteases are expanded in *Steinernema*, above what is seen in other non-parasitic species, as well as the abundance of protease inhibitors (Table 6.4). It is likely that these types of proteases are involved in insect parasitism and could play a role in host specificity.

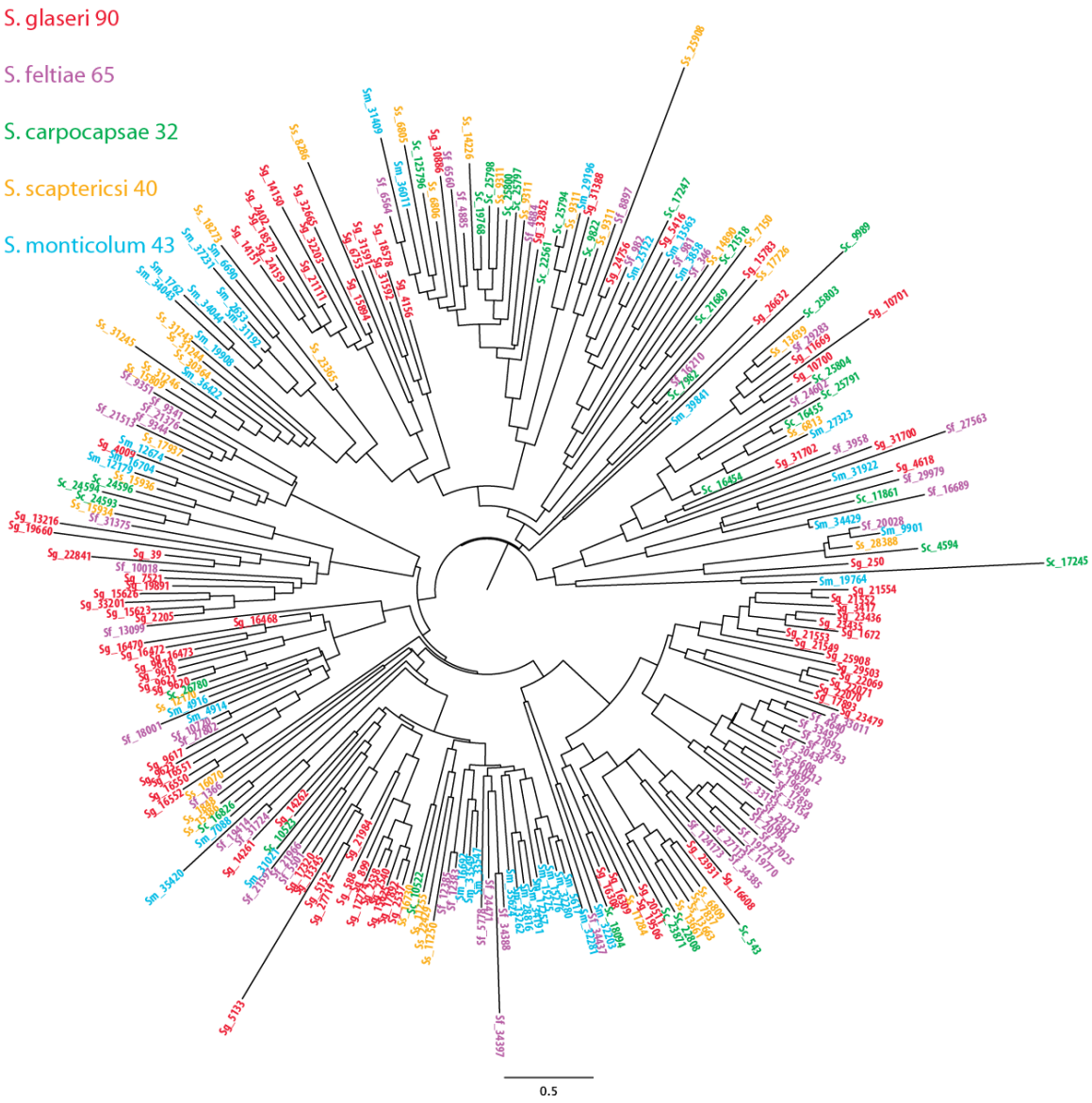

Figure 6.2 | Srd Family GPCRs Among Sequenced Steinernematids. This is a gene tree including all Srd GPCR family genes among the five steinernematids I sequenced. Although the gene names are too small to read, the pattern of conservation or expansion of genes is visible in the colors, with several apparent expansions in *S. glaseri* and *S. feltiae*.

These data lead me to believe that it could be the abundance and diversity of proteases that determines what kinds of insects an EPN is capable of infecting, while it is the abundance and diversity of GPCRs that determine which insects they are attracted to.

Clearly proteases, protease inhibitors, and GPCRs have played a significant role in the evolution of parasitism among steinernematids as well as niche partitioning among these species.

Proteases

Type of Protease	S. carpo	S. scapt	S. felti	S. glase	S. monti	B. xylop	P. rediv	C. japon	C. elega	C. brenn	C. reman	C. brigg	P. pacif
Aspartic	51	77	48	56	83	78	24	36	27	36	36	33	38
Cysteine	141	155	152	130	171	148	112	108	141	171	151	126	100
Metallo	232	306	365	360	439	209	232	209	203	275	233	185	215
Serine	264	370	333	288	531	174	227	134	156	169	168	156	237
Threonine	28	37	36	38	54	21	21	102	23	42	74	26	22
Unassigned	2	1	5	3	17	1	1	2	1	1	1	2	1
Total	718	946	939	875	1295	631	617	591	551	694	663	528	613
Inhibitors	189	225	231	158	191	60	91	80	94	132	119	102	64

Table 6.4 | Protease and Protease Inhibitors Across Selected Nematodes. This table displays the number of proteases in each of the subtypes. These data were assembled using the MEROPS protease database [44].

As mentioned earlier, these data are unpublished and remain unrefined. Additional sequencing is currently taking place and more analyses including the conservation of certain important biological pathways such as sex determination, RNA interference, dauer, and cell death pathways will be analyzed. I plan to do a more detailed analysis of the proteases, including those with signal peptides, to narrow down a list of potentially secreted proteases.

As a collected work this thesis contributes to our understanding of parasitism, host-seeking behavior, and the architecture of parasitism among nematodes. I have shown a conserved role for the BAG neurons in detected CO_2 in *C. elegans*, *H. bacteriophora*, and *S. carpocapsae*. This conservation spans considerable phylogenetic distance. I have

shown clear differences in olfactory preferences and virulence among EPNs and have demonstrated that this preference correlates with host suitability. I have placed these findings in the broader context of what it means to be an EPN and how these differ from other insect parasites.

I have discussed the role of genomics in nematology and exerted considerable effort to encourage genomic sequencing and analysis among nematologists and have been a driving force in steering the sequencing that is currently being done at Caltech. Though no genome papers have yet been published based on my work, I anticipate several significant contributions coming out in the next two years. It is clear that much of the information in my thesis builds on the work of others, but I have still conducted new research and contributed new knowledge of appreciable application across many fields.

It has been a pleasure to be involved in this work, and I have particularly enjoyed the conditions and working environment I experienced at Caltech. I close with my favorite Charles Darwin quote: “Doing what little one can to increase the general stock of knowledge is as respectable an object of life, as one can in any likelihood pursue.”

References

1. Nguyen, K.B., Hunt, D.J., and Mracek, Z. (2007). Steinernematidae: species and descriptions. In Entomopathogenic Nematodes: Systematics, Phylogeny and Bacterial Symbionts, K.B. Nguyen and D.J. Hunt, eds. (Boston: Brill), pp. 121–609.

2. Stock, S.P., and Goodrich-Blair, H. (2012). Nematode parasites, pathogens and associates of insects and invertebrates of economic importance. In Manual of techniques in invertebrate pathology, 2nd Edition, L.A. Lacey, ed. (Amsterdam: Elsevier Press), pp. 375–425.

3. Campbell, J.F., and E.E., L. (2002). Entomopathogenic nematode host-search strategies. In The Behavioral Ecology of Parasites, J.F. Campbell and M.V.K. Sukhdeo, eds. (New York: CABI), pp. 13–38.

4. Campbell, J.F., and Kaya, H.K. (1999). Mechanism, kinematic performance, and fitness consequences of jumping behavior in entomopathogenic nematodes (Steinernema spp.). Canadian Journal of Zoology *77*, 1947–1955.

5. Campbell, J.F., and Kaya, H.K. (2002). Variation in entomopathogenic nematode (Steinernematidae and Heterorhabditidae) infective-stage jumping behavior. Nematology *4*, 471–482.

6. Koppenhofer, A.M., Ganguly, S., and Kaya, H.K. (2000). Ecological characterisation of Steinernema monticolum, a cold-adapted entomopathogenic nematode from Korea. Nematology *2*, 407–416.

7. Stock, S.P., Choo, H.Y., and Kaya, H.K. (1997). An entomopathogenic nematode, Steinernema monticolum sp. n (Rhabditida: Steinernematidae) from Korea with a key to other species. Nematologica *43*, 15–29.

8. Spiridonov, S.E., Reid, A.P., Podrucka, K., Subbotin, S.A., and Moens, M. (2004). Phylogenetic relationships within the genus Steinernema (Nematoda: Rhabditida) as inferred from analyses of sequences of the ITSI-5.8S-ITS2 region of rDNA and morphological features. Nematology *6*, 547–566.

9. Adams, B.J., Peat, S.M., and Dillman, A.R. (2007). Phylogeny and evolution. In Entomopathogenic nematodes: Systematics, phylogeny, and bacterial symbionts., Volume 5, K.B. Nguyen and D.J. Hunt, eds. (Leiden-Boston: Brill), pp. 693–733.

10. Bilgrami, A.L., Gauger, R., Shapiro-Ilan, D.I., and Adams, B.J. (2006). Source of trait deterioration in entomopathogenic nematodes *Heterorhabditis bacteriophora* and *Steinernema carpocapsae* during *in vivo* culture. Nematology *8*, 397–409.

11. Chaston, J.M., Dillman, A.R., Shapiro-Ilan, D.I., Bilgrami, A.L., Gaugler, R., Hopper, K.R., and Adams, B.J. (2011). Outcrossing and crossbreeding recovers deteriorated traits in laboratory cultured Steinernema carpocapsae nematodes. Int J Parasitol *41*, 801–809.

12. Hallem, E.A., Dillman, A.R., Hong, A.V., Zhang, Y., Yano, J.M., DeMarco, S.F., and Sternberg, P.W. (2011). A sensory code for host seeking in parasitic nematodes. Current Biology *21*, 377–383.

13. Li, X., Cowles, E.A., Cowles, R.S., Gaugler, R., and Cox-Foster, D.L. (2009). Characterization of immunosuppressive surface coat proteins from Steinernema glaseri that selectively kill blood cells in susceptible hosts. Molecular and biochemical parasitology *165*, 162–169.

14. Nguyen, K.B., and Smart, G.C. (1991). Pathogenicity of Steinernema scapterisci to Selected Invertebrates. J. Nematol. *23*, 7–11.

15. Campos-Herrera, R., Escuer, M., Robertson, L., and Gutierrez, C. (2006). Morphological and Ecological Characterization of Steinernema feltiae (Rhabditida: Steinernematidae) Rioja Strain Isolated from Bibio hortulanus (Diptera: Bibionidae) in Spain. J Nematol *38*, 68–75.

16. White, G.F. (1927). A Method for Obtaining Infective Nematode Larvae from Cultures. Science *66*, 302–303.

17. Kaya, H.K., and Stock, S.P. (1997). Techniques in insect nematology. In Manual of techniques in insect pathology, L. Lacey, ed. (San Diego, CA: Academic Press Limited).

18. Vivas, E.I., and Goodrich-Blair, H. (2001). Xenorhabdus nematophilus as a model for host-bacterium interactions: rpoS is necessary for mutualism with nematodes. J Bacteriol *183*, 4687–4693.

19. Portman, D.S. (2006). Profiling C. elegans gene expression with DNA microarrays. WormBook, 1–11.

20. Collins, F.S., Drumm, M.L., Cole, J.L., Lockwood, W.K., Van de Woude, G.F., and Iannuzzi, M.C. (1987). Contruction of a general human chromosome jumping library, with application to cystic fibrosis. Science *235*, 1046–1049.

21. Mortazavi, A., Schwarz, E.M., Williams, B.A., Schaeffer, L., Antoshechkin, I., Wold, B., and Sternberg, P.W. (2010). Scaffolding a *Caenorhabditis* nematode genome with RNA-seq. Genome Res. *20*, 1740–1747.

22. Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics *24*, 637–644.

23. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

24. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology *28*, 511–U174.

25. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., and Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol *12*, R22.

26. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

27. Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res *13*, 2178–2189.

28. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res *39*, W29–37.

29. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

30. Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6.

31. Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061–1067.

32. Bargmann, C.I. (2006). Chemosensation in *C. elegans*. In WormBook, E.M. Jorgensen, ed.

33. Bargmann, C.I. (2006). Comparative chemosensation from receptors to ecology. Nature *444*, 295–301.

34. Dillman, A.R., Guillermin, M.L., Lee, J.H., Kim, B., Sternberg, P.W., and Hallem, E.A. (2012). Olfaction shapes host-parasite interactions in parasitic nematodes. Proc Natl Acad Sci U S A *109*, E2324–2333.

35. McCarroll, S.A., Li, H., and Bargmann, C.I. (2005). Identification of transcriptional regulatory elements in chemosensory receptor genes by probabilistic segmentation. Current biology: CB *15*, 347–352.

36. Robertson, H.M., and Thomas, J.H. (2006). The putative chemoreceptor families of C. elegans. WormBook, 1–12.

37. Thomas, J.H., and Robertson, H.M. (2008). The Caenorhabditis chemoreceptor gene families. BMC biology *6*, 42.

38. Allen, J.E., and MacDonald, A.S. (1998). Profound suppression of cellular proliferation mediated by the secretions of nematodes. Parasite Immunol *20*, 241–247.

39. Balasubramanian, N., Hao, Y.J., Toubarro, D., Nascimento, G., and Simoes, N. (2009). Purification, biochemical and molecular analysis of a chymotrypsin protease with prophenoloxidase suppression activity from the entomopathogenic nematode Steinernema carpocapsae. Int. J. Parasit. *39*, 975–984.

40. Balasubramanian, N., Toubarro, D., and Simoes, N. (2010). Biochemical study and in vitro insect immune suppression by a trypsin-like secreted protease from the nematode Steinernema carpocapsae. Parasite Immunol. *32*, 165–175.

41. Brivio, M.F., Moro, M., and Mastore, M. (2006). Down-regulation of antibacterial peptide synthesis in an insect model induced by the body-surface of an entomoparasite (Steinernema feltiae). Dev Comp Immunol *30*, 627–638.

42. Brivio, M.F., Pagani, M., and Restelli, S. (2002). Immune suppression of Galleria mellonella (Insecta, Lepidoptera) humoral defenses induced by Steinernema feltiae (Nematoda, Rhabditida): involvement of the parasite cuticle. Exp. Parasitol. *101*, 149–156.

43. Li, X.Y., Cowles, R.S., Cowles, E.A., Gaugler, R., and Cox-Foster, D.L. (2007). Relationship between the successful infection by entomopathogenic nematodes and the host immune response. Int. J. Parasit. *37*, 365–374.

44. Rawlings, N.D., Barrett, A.J., and Bateman, A. (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res *40*, D343–D350.