2010-06-18

Utilization of Phylogenetic Systematics, Molecular Evolution, and Comparative Transcriptomics to Address Aspects of Nematode and Bacterial Evolution

Scott M. Peat
Brigham Young University - Provo

Utilization of Phylogenetic Systematics, Molecular Evolution, and Comparative Transcriptomics to Address Aspects of Nematode and Bacterial Evolution

Scott M. Peat

A dissertation submitted to the faculty of
Brigham Young University
In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Byron J. Adams
Keith A. Crandall
Michael F. Whiting
Alan R. Harker
George O. Poinar

Department of Biology

Brigham Young University

August 2010

ABSTRACT

Utilization of Phylogenetic Systematics, Molecular Evolution, and Comparative Transcriptomics to Address Aspects of Nematode and Bacterial Evolution

Scott M. Peat

Department of Biology

Doctor of Philosophy

Both insect parasitic/entomopathogenic nematodes and plant parasitic nematodes are of great economic importance. Insect parasitic/entomopathogenic nematodes provide an environmentally safe and effective method to control numerous insect pests worldwide. Alternatively, plant parasitic nematodes cause billions of dollars in crop loss worldwide. Because of these impacts, it is important to understand how these nematodes evolve, and, in the case of entomopathogenic nematodes, how their bacterial symbionts evolve. This dissertation contains six chapters. Chapter one is a review of DNA markers and their use in the phylogenetic systematics of entomopathogenic and insect-parasitic nematodes as well as a review of phylogenetic, co-phylogenetic, and population genetic methodologies. Chapter two characterizes positive destabilizing selection on the *lux*A gene of bioluminescent bacteria. Our data suggests that bacterial ecology and environmental osmolarity are likely driving the evolution of the *lux*A gene in bioluminescent bacteria. Chapter 3 examines relationships among bacteria within the genus *Photorhabdus*. Our analyses produced the most robust phylogenetic hypothesis to date for the genus *Photorhabdus*. Additionally, we show that *glnA* is particularly useful in resolving specific and intra-specific relationships poorly resolved in other studies. We conclude that *P. asymbiotica* is the sister group to *P. luminescens* and that the new strains HIT and JUN should be given a new group designation within *P. asymbiotica*. Chapter 4 characterizes the morphology of the head and feeding apparatus of fungal feeding and insect infective female morphs of the nematode *Deladenus siricidicola* using scanning electron microscopy. Results showed dramatic differences in head, face, and stylet morphology between the two *D. siricidicola* female morphs that were not detected in previous studies using only light microscopy. Chapter five utilizes comparative transciptomics to identify putative plant and insect parasitism genes in the nematode *Deladenus siricidicola*. Results from this study provide the first transcriptomic characterization for the nematode *Deladenus siricidicola* and for an insect parasitic member of the nematode infraorder Tylenchomorpha. Additionally, numerous plant parasitism gene homologues were discovered in both *D. siricidicola* libraries suggesting that this nematode has co-opted these plant parasitism genes for other functions. Chapter six utilizes a phylogenomic approach to estimate the phylogeny of the nematode infraorder Tylenchomorpha.

Keywords: *Photorhabdus*, *lux*A, positive destabilizing selection, bioluminescent bacteria, phylogeny, *Deladenus siricidicola*, stylet, morphology, scanning electron microscopy, comparative transcriptomics, parasitism genes, phylogenomics, Tylenchomorpha

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to all of the individuals who have assisted and supported me in completing this dissertation. I would first like to thank my advisor, Byron Adams, for his advice, support, optimism, encouragement throughout this whole process. You were not just my advisor but also a mentor and a friend and I am forever greatful for the time I have been able to share with you and your family. To the members of my graduate committee, Keith Crandall, Michael Whiting, Alan Harker, and George Poinar, thank you for all of your feedback and guidance on each of my projects, and for pushing me to be a better scientist. I would like to thank Hilary Oldroy and Elaine Rotz for making sure I met all of my deadlines and for having an answer or solution to any issues I had regarding degree requirements and paperwork.

I would like to thank Ed Wilcox for sequencing assistance and troubleshooting and Josh Udall and Peter Maughan for providing invalueable suggestions on cDNA library construction techniques. Additionally, I would like to thank Matthew Bendall, Tyler Collette, and Daniel Standage for invaluable bioinformatics assistance. To all of the members of the nematode evolution lab, thank you for all of your support and assistance over the last five years.

I would like to thank the following agenecies for providing funding for this project: Brigham Young UniversityGraduate Studies, Department of Biology, Department of Molecular Biology and Microbiology, Charles Redd Center, United States Department of Agriculture, and the National Science Foundation.

Finally, I would like to express gratitude to my family for their love and support thoroughout this whole process. I would especially like to thank my wife for her unwavering

support, patience, and continuous encouragement, without which I would have never been able to make it through this process.

Table of Contents

List of Tables

List of Figures

Introduction

Nematodes are a relatively ancient group of organisms, with their origin believed to be sometime around the Cambrian or Precambrian period (Baldwin et al., 2004; Poinar et al., 2008). While most nematode diversity is represented by free-living nematodes in marine, soil, or freshwater environments, some of the more economically important nematodes are those that live a parasitic lifestyle. Nematodes parasitize a wide range of hosts from plants, to arthropods and vertebrates, with some insect pathogenic nematode utilizing the services of a symbiotic bacterium to aid in the taking advantage of their insect host. The exact origin of the parasitic lifestyle in nematodes is unknown, though a fossil of *Cretacimermis libani* parasitizing adult midges in 135 million year old amber demonstrates that the animal parasitic lifestyle was around at least during the Cretaceous period (Poinar, 2003; Poinar et al., 1994). Like their animal counterparts, the origin of plant parasitic nematodes (PPNs) is unknown. A recent discovery of eggs, juveniles, and adults of an early Devonian nematode within the plant tissue of the early land plant *Aglaophyton major* suggests that nematodes had already formed associations with plants some 396 million years ago (Poinar et al., 2008). Paleontological evidence indicates that nematodes have co-inhabited earth with plants and other animals for well over 100 million years, and as such it is not surprising that the parasitic lifestyle in nematodes has arisen multiple independent times throughout the evolution of the phylum Nematoda (Blaxter et al., 2000; Blaxter et al., 1998; Dorris et al., 1999; Holterman et al., 2006)

Some of the most economically important nematodes are insect parasitic/entomopathogenic nematodes. The relationship that insect associated nematodes have with their insect host varies from group to group and species to species. Some insect associated nematodes are considered entomopathogenic, as they are lethal to their insect host (Liu et al.,

2000), while others are phoretic, as the insect is only used as a mode of transportation for the nematode (Giblin-Davis, 1996). Still other insect associated nematodes, while inhibiting reproductive capabilities, do not appear to have a direct pathogenic effect on the host itself (Poinar Jr., 1991; poinar Jr. and Van Der Laan, 1972). Many of these insect parasitic and entomopathogenic nematodes have shown promise as biocontrol agents (Bedding and Akhurst, 1974; Bedding and Iede, 2005; Poinar Jr., 1979). To harness the true potential of these nematodes as biological control agents, it is important to have an in-depth understanding of not only their biology but also their evolutionary history. As such, chapter one of my dissertation provides a review of the DNA markers that have been used to study the evolutionary relationships of entomopathogenic and insect-parasitic nematodes as well as a review of phylogenetic, co-phylogenetic, and population genetic methodologies that can be used in combination with molecular data to effectively study nematodes.

The entomopathogenic nematodes *Heterorhabditis* and *Steinernema* utilize bacterial endosymbionts to aid in killing and utilizing larval insects as a food source (Poinar, 1990; Poinar Jr., 1979). The endosymbiont of heterorhabditid nematodes, *Photorhabdus*, is unique in that it is the only terrestrial bacteria known to exhibit bioluminescence (Gerrard et al., 2003). While the mechanism by which light is produced in *Photorhabdus* is understood, the functional significance for the production of light in *Photorhabdus* has yet to be discovered. As such, chapter two of my dissertation utilizes molecular biological methods to identify positive destabilizing selection acting on the *lux*A gene (one of the genes responsible for the production of light in bioluminescent bacteria) of bioluminescent bacteria to address questions regarding the functional significance of bioluminescence in *Photorhabdus*. Additionally, to fully investigate the functional significance of bioluminescence as well as test fundamental hypotheses regarding

the evolutionary history of Photorhabdus, a well resolved phylogenetic hypothesis for the genus Photorhabdus is essential. Previous studies have been unsuccessful at resolving the basal nodes within the *Photorhabdus* phylogeny. With this in mind, chapter three of my dissertation utilizes three molecular datasets and a suite of phylogenetic systematic tools to reconstruct the evolutionary history of the bacterial genus *Photorhabdus.* The resulting phylogeny was used to evaluate specific and sub-specific taxonomic statements within the genus *Photorhabdus*, to evaluate the utility of 16S rRNA, *gyrB*, and *glnA* in resolving relationships within the genus *Photorhabdus,* and to investigate the evolution of bioluminescent intensity through the genus *Photorhabdus*.

While entomopathogenic nematodes and their endosymbionts have shown great progress in the biological control arena, the insect parasitic nematode *Deladenus* (=*Beddingia*) *siricidicola* has been utilized extensively for well over 30 years in the control of the woodwasp, *Sirex noctilio* (Bedding, 1972, 1993; Bedding and Akhurst, 1974; Bedding and Iede, 2005). *Deladenus siricidicola* is somewhat unique in that it has two autonomous and trophically diverse life history stages; a mycetophagous (fungal feeding) life stage and an entomophagous (insect parasitic) life stage and it is part of an intermediate clade between fungal feeding and plant parasitic nematodes within the infraorder Tylenchomorpha (De Ley and Blaxter, 2002). As such, the biology and evolutionary history of *D. siricidicola make* it an excellent model system for studies addressing the origin and maintenance of plant parasitism genes and the identification of genes involved in the parasitism of insects by nematodes. To this end, chapters four, five, and six focus on the characterization of the unique morphology of *D. siricidicola* females, the characterization of the expressed genes in *D. siricidicola*, and a reconstructing evolutionary relationships within the infraorder Tylenchomorpha. Chapter four utilized scanning electron

microscopy to investigate morphological differences which allow the two female morphs of *D. siricidicola* (fungal feeding and insect parasitic) to inhabit two extremely diverse ecological niches. Chapter five utilized 454 pyrosequencing to characterize the transcriptomes of mycetophagous and entomophagous female morphs of the nematode *D. siricidicola,* and utilized the transcriptomic data to facilitate the identification of potential insect and plant parasitism genes. Work from chapter five provides the first characterization of the transcriptome of an insect parasitic member of the nematode infraorder Tylenchomorpha.

In order to effectively study the origin and maintenance of the plant and insect parasitism gene data that was generated from the transcriptomic analyses discussed in chapter five, a robust phylogenetic hypothesis for the infraorder Tylechomorpha is required. Previous phylogenetic systematic studies on the infraorder Tylenchomorpha (Bert et al., 2008; Holterman et al., 2009; Subbotin et al., 2006) have shown strong support for terminal clades, though many of the deeper nodes remain poorly supported and/or unresolved, and more genetic data from different loci are needed to fully resolve the relationships within Tylenchomorpha. As such, chapter six utilizes a phylogenomic approach to estimate the phylogeny of the nematode infraorder Tylenchomorpha, in an attempt to provide better resolution at the deeper nodes of the phylogeny.

References

Baldwin, J.G., Nadler, S.A., Adams, B.J., 2004. Evolution of plant parasitism among nematodes. Annual Review of Phytopathology 42, 83-105.

Bedding, R.A., 1972. Biology of *Deladenus siricidicola* (Neotylenchidae), an entomophagous-mycetophagous nematode parasitic on siricid woodwasps. Nematologica 18, 482-493.

Bedding, R.A., 1993. Biological control of *Sirex noctilio* using the nematode *Deladenus siricidicola*. In: Bedding, R.A., Akhurst, R.J., Kaya, H.K. (Eds.), Nematodes and the Biological Control of Insect Pests. CSIRO, East Melbourne, Australia, pp. 11-20.

Bedding, R.A., Akhurst, R.J., 1974. Use of the nematode *Deladenus siricidicola* in the biological control of *Sirex noctilio* in Australia Journal of the Australian Entomological Society 13, 129-135.

Bedding, R.A., Iede, E.T., 2005. Application of *Beddingia siricidicola* for Sirex Woodwasp Control. In: Grewal, P.S. (Ed.), Nematodes As Biocontrol Agents. CABI Publishing, Oxfordshire, UK, pp. 389-399.

Bert, W., Leliaert, F., Vierstraete, A.R., Vanfleteren, J.R., Borgonie, G., 2008. Molecular phylogeny of the Tylenchina and evolution of the female gonoduct (Nematoda : Rhabditida). Molecular Phylogenetics and Evolution 48, 728-744.

Blaxter, M., Dorris, M., De Ley, P., 2000. Patterns and processes in the evolution of animal parasitic nematodes. Nematology 2, 43-55.

Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., Vida, J.T., Thomas, W.K., 1998. A molecular evolutionary framework for the phylum Nematoda. Nature 392, 71-75.

De Ley, P., Blaxter, M., 2002. Systematic position and phylogeny. In: Lee, D.L. (Ed.), Biology of Nematodes. Taylor & Francis, Florence, KY, pp. 1 - 30.

Dorris, M., De Ley, P., Blaxter, M.L., 1999. Molecular analysis of nematode diversity and the evolution of parasitism. Parasitology Today 15, 188-193.

Gerrard, J.G., McNevin, S., Alfredson, D., Forgan-Smith, R., Fraser, N., 2003. *Photorhabdus* species: Bioluminescent bacteria as emerging human pathogens? Emerg. Infect. Dis 9, 251-254.

Giblin-Davis, R.M., 1996. Phoresy between diplogasterid, aphelenchid, and tylenchid nematodes and insects. Nematropica 26, 205.

Holterman, M., Karssen, G., van den Elsen, S., van Megen, H., Bakker, J., Helder, J., 2009. Small Subunit rDNA-Based Phylogeny of the Tylenchida Sheds Light on Relationships Among Some High-Impact Plant-Parasitic Nematodes and the Evolution of Plant Feeding. Phytopathology 99, 227-235.

Holterman, M., van der Wurff, A., van den Elsen, S., van Megen, H., Bongers, T., Holovachov, O., Bakker, J., Helder, J., 2006. Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. Molecular Biology and Evolution 23, 1792-1800.

Liu, J., Poinar Jr., G.O., Berry, R.E., 2000. Control of insect pests with entomopathogenic nematodes the impact of molecular biology and phylogenetic reconstruction. Annual Review of Entomology 45, 287-306.

Poinar, G., 2003. Trends in the evolution of insect parasitism by nematodes as inferred from fossil evidence. Journal of Nematology 35, 129-132.

Poinar, G., Kerp, H., Hass, H., 2008. *Palaeonema phyticum* gen. n., sp n. (Nematoda : Palaeonematidae fam. n.), a Devonian nematode associated with early land plants. Nematology 10, 9-14.

Poinar, G.O., Acra, A., Acra, F., 1994. Earliest fossil nematode (Mermithidae) in Cretaceous lebanese amber. Fundamental and Applied Nematology 17, 475-477.

Poinar, G.O.J., 1990. Biology and taxonomy of Steinernematidae and Heterorhabditidae. In: Gaugler, R., Kaya, H.K. (Eds.), Entomopathogenic Nematodes in Biological Control. CRC Press, Boca Raton, FL, pp. 23-62.

Poinar, G.O., 1979. Nematodes for Biological Control of Insects. CRC Press, Boca Raton, Florida.

Poinar, G.O., 1991. The mycetophagous and entomophagous stages of *Iotonchium californicum* n. sp. (Iotonchiidae: Tylenchida). . Revue Nematol 14, 565-580.

Poinar, G.O., Van Der Laan, P.A., 1972. Morphology and life history of *Sphaerularia bombi*. Nematologica 18, 239-252.

Subbotin, S.A., Sturhan, D., Chizhov, V.N., Vovlas, N., Baldwin, J.G., 2006. Phylogenetic analysis of Tylenchida Thorne, 1949 as inferred from D2 and D3 expansion fragments of the 28S rRNA gene sequences. Nematology 8, 455-474.

Manuscript published in Molecular Approaches and Techniques for the Study of Insect Pathogens

Chapter 1

Phylogenetics and Population Genetics of Entomopathogenic and Insect-Parasitic Nematodes

SCOTT M. PEAT[1], BRADLEY C. HYMAN[2] AND BYRON J. ADAMS[1]

[1] *Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT, USA 84602-5253*

[2] *Department of Biology, University of California, Riverside, CA 92521*

Introduction

Phylogenetics, and particularly molecular systematics, has played a key role in numerous advances in the study of entomopathogenic, entomophilic, and insect parasitic nematodes. The contribution of molecular systematics and population genetics to both applied and fundamental research on these organisms is most evident in taxonomic endeavors, but has also been integral to expanding knowledge of their biodiversity, geographic distributions, host ranges, ecology, behavior, and coevolution (Adams et al., 2006, Campbell et al., 2003). In this chapter we present a brief introduction to phylogenetics, population genetics and DNA barcoding, and discuss the genetic loci and analytical tools (software) that are relevant to applying them to entomogenous nematodes.

Phylogenetics

Phylogenetic systematics, the study and process of recovering the historical relationships among species and taxonomic groups, has greatly aided in the study of insect parasitic and pathogenic nematode diversity and their evolution. Phylogenies not only reveal the hierarchical relationships among taxa, but they can also be used as a contextual framework to study the evolution of life history traits, morphological traits, behavior, pathogenicity, and any other characteristic of entomopathogenic nematodes. Phylogenies are also the foundation of research programs in historical biogeography, phylogeography, historical ecology and coevolution (Avise, 2000, 2004; Brooks and McLennan, 1991, 2002). A reconstructed past, in the form of a phylogeny, provides the historical context required for inferences of evolutionary change to be tested, shedding light on the origin, mode, tempo, and maintenance of entomogenous nematode diversity that we see today.

The development of phylogenetic systematics emerged as biologists began to embrace Darwin's notion that classifications should reflect evolutionary relationships (Andrássy, 1976), with pheneticists and cladists in conflict over which philosophical approach was most appropriate to use for constructing evolutionary relationships. Pheneticists considered overall similarity to be the best indicator of phylogeny, whereas cladists argue that only shared, derived characters appropriately and accurately reflected evolutionary relationships. The cladists prevailed, in terms of their logical and empirical arguments, as well as having their methods widely accepted and adopted by subsequent researchers. Today, a wide variety of methods exist for building phylogenetic trees. These include phenetic methods such as UPGMA or neighbor-joining, parsimony, and model-based methods, like maximum likelihood and Bayesian analysis.

Parsimony and model-based methods have been shown to perform best in simulations (Huelsenbeck and Hillis, 1993, Huelsenbeck, 1995, Siddall, 2001, Siddall, 1998), but more contemporary methods have been developed for neighbor joining analyses that improve their performance (Gascuel, 1997; Steel et al., 2000)(for further discussion see (Hillis et al., 1992; Hillis et al., 1994; Huelsenbeck and Hillis, 1993; Siddall, 1998; Swofford et al., 2001).

Prior to the advent of molecular tools, systematists primarily utilized morphological characters to construct phylogenies and infer evolutionary relationships. The relatively conservative morphologies of nematodes have made this exercise difficult. For example, *Caenorhabditis elegans* and *C. briggsae* are virtually indistinguishable based on their morphologies, yet comparative analyses of their genomes suggest that they last shared a common ancestor 80-110 million years ago - comparable to the divergence times of *Anopheles* and *Drosophila* (Stein et al., 2003). Clearly, the benefits of molecular data in nematode phylogenetics are numerous. The advent and refinement of PCR and DNA sequencing techniques have made it possible to produce millions of characters in a matter of days (Hudson, 2007).

Early studies of phylogenetic relationships of insect parasitic and pathogenic nematodes primarily utilized randomly amplified polymorphic DNA (RAPD) (Liu and Berry, 1996) and morphology and bionomics (Poinar, 1993). Today, phylogenetic relationships of insect parasitic and pathogenic nematodes can be inferred using sequence data from nuclear and mitochondrial genetic loci, often with a special emphasis on nuclear ribosomal DNA genes, partial mRNA copies of protein-coding genes (Expressed Sequence Tags [ESTs]) and more recently, whole genomes (phylogenomics).

One of the more challenging problems facing systematists is finding the optimal solution given the number of possible phylogenetic trees that can be produced. As the number of taxa increase in a phylogenetic systematic study, the number of possible phylogenetic solutions drastically increases (Felsenstein, 1978). For example, in an analysis with four taxa, there are 15 possible rooted phylogenetic reconstructions. In an analysis of only 10 taxa, the number of possible rooted phylogenetic reconstructions drastically increases to approximately 34,459,425. For 135 taxa there are approximately 10^{265} possible rooted binary trees, more than the number of electrons in the known universe (Penny et al., 1995).

Population Genetics

Population genetics remains one of the most understudied aspects of this important group of organisms. The roots of population genetics can be traced back to Darwin and Wallace's theory of natural selection as well as Mendel's explanation of the genetic mechanisms of inheritance. From Mendel's ideas, Hardy and Weinberg developed one of the simplest models of population genetics, which has become the null model in describing genetic attributes of a population (Templeton, 2006). Other models/theories exploring changes in allele frequencies within populations, mutation, and inbreeding were pioneered by Fisher, Haldane, and Wright. Their work provided the framework for a quantitative analysis of Mendelian genetics (Thompson, 1990), and a synthesis of Mendelian heredity and natural selection into the science of population genetics (Provine, 2001). Methods for inferring population processes from genetic patterning have increased tremendously of the past 20 years. In this chapter we present some of the most relevant to entomogenous nematodes.

DNA Barcoding

Barcoding (see Chapter 4) has not been developed to infer deep relationships or group species into kingdoms, phyla, or classes, as this is the job of phylogenetic systematics (though MOTU [Molecular Operational Taxonomic Unit] data can be used in phylogenetic analyses [Floyd et al., 2002; Powers, 2004]). However, the information gained from barcoding can be used with pre-existing phylogenies to answer questions that barcoding could not answer by itself. DNA barcoding, taxonomy, and systematics should not be thought of as mutually exclusive research pathways. There is overlap between these three disciplines and together they can work to build a better system for the identification of species, inferring diversity, and determining relationships between and among a variety of taxonomic groups.

DNA Markers considered for phylogenetic and population genetics studies

Ribosomal DNA

Nuclear ribosomal DNA has proven extremely useful and has been employed extensively to study nematode systematics at the molecular level (See Chapter 4). The variability of evolutionary rates observed among different genes and spacers within an rDNA transcription unit is useful in that specific segments can be chosen based on the taxonomic or organismal level of study. For example, for a phylogenetic analysis at the species level, small subunit (SSU) rRNA has been considered suboptimal when attempting to differentiate closely related species (Liu et al., 1997), while the less conserved regions of the large ribosomal subunit (LSU) (Stock et al., 2001, Nadler et al., 2006b) and the internal transcribed spacer (ITS) that separates rDNA coding regions (Adams et al., 2006, Nguyen et al., 2001) have proven to be more informative. For

studies involving deeper nodes among more distantly related taxa, the more conserved SSU and LSU gene regions are more appropriate.

The single biggest obstacle in using rRNA genes in phylogenetic reconstruction is that the gene product can vary in length without compromising functionality within the ribosome. Whereas the length and composition of protein coding genes are generally subject to selection by codon usage, rRNA genes are not. For some rDNA regions, insertion and deletion events are as frequent as transitions and transversions. In some cases, insertion and deletion events (indels) can involve blocks of multiples of nucleotides (Adams et al., 1998, Nguyen et al., 2001). Indel events can result in substantial rDNA size differences between sequences (taxa), which complicate the process of generating multiple sequence alignments and reduces confidence in the homology statements for each nucleotide in the multiple sequence alignment. Because phylogenetic reconstruction tests accurate homology statements (i.e., the thymidine at position 123 in the multiple sequence alignment in taxon A is homologous to the thymidine in taxon B at the same position), alignment ambiguity can result in spurious phylogenies.

In fact, this aspect has been explored for *Heterorhabditis* and *Steinernema*, and results suggest there is more variation in tree topology due to differences in the multiple sequence alignment than there is from the different methods used to generate the trees (i.e. parsimony, maximum likelihood, and neighbor-joining) (Adams et al., 1998, Nadler et al., 2006a, Nguyen et al., 2001, Spiridonov et al., 2004). Approaches to addressing this problem require thoughtful consideration and include visually inspecting the sequences and removing the alignment-ambiguous regions based on an *a-priori* metric (i.e. remove ambiguous indels that lie between invariant regions), direct optimization (discussed below in the section on phylogenetic methods), comparison of secondary structure based on minimum energy models, and minimum posterior

probabilities among alternative placements of nucleotides (characters) in the alignment (see discussion of alignment methods below).

Alternatives to rRNA genes include single copy mitochondrial sequences (discussed in more detail below), nuclear protein coding genes, and intron sequences. The majority of the single copy nuclear protein-coding genes thus far explored for phylogenetic utility are fairly conserved, and found to be most useful for resolving very deep nodes among distantly related taxa. Some of these include heat shock protein HSP90 *(daf*-21), RNA polymerase II (*ama*-1) and actin (*act*-1/3,2,4) (Skantar and Carta, 2004, Kovaleva et al., 2004, Baldwin et al., 1997). Genes encoding ribosomal proteins (rather than rRNA) are promising for the resolution of fairly deep nodes, such as among genera of entomogenous nematodes, but it should be noted that rRNA genes and their highly expressed associated ribosomal proteins appear to evolve in a concerted fashion (Longhorn et al., 2007). Similar to ITS rDNA, introns generally have high rates of nucleotide substitution, and contain numerous indels. Thus, ITS and intron sequences present similar multiple sequence alignment challenges. Intron sequences can be very useful for population genetic studies and phylogenetic analyses among very closely related taxa, although these are only beginning to be explored in entomogenous nematodes (Rolston et al., 2004).

Small Subunit Ribosomal DNA (SSU or 18S)

The SSU rRNA gene has been the most frequently utilized genetic marker for nematodes, as it has proven useful in studies of deep phylogenetic relationships because of its slow evolutionary rate of change. The conservative nature of SSU rRNA also allows for the development of universal primers which can be used to amplify DNA from groups of nematodes for which little to no molecular data exists (Hillis and Dixon, 1991). Blaxter et al. (1998)

utilized small subunit sequences from 53 nematode taxa to construct a phylogeny from which they could study the evolution of the phylum Nematoda. Though the SSU is highly conserved, Blaxter et al. were able to use the SSU to differentiate between major nematode groups, and use their newly constructed evolutionary framework to bring into question the monophyly of previously proposed groups. Holterman et al. (2006) conducted a similar study using 339 nematode taxa. The additional taxa resulted in the proposal that the phylum Nematoda be divided into 12 clades rather than the five clades proposed by Blaxter et al. (1998). Furthermore, Holterman et al. suggested that 18S rRNA may be suitable for differentiation at the species level due to the acceleration of substitution rates in plant and animal parasitic clades.

Large Subunit Ribosomal DNA (LSU or 28S)

While SSU rRNA is primarily used to examine evolutionary events that occurred in the Precambrian time period, LSU rRNA is used primarily to examine evolutionary events which occurred through the Paleozoic and Mesozoic time periods (Hillis and Dixon, 1991). For example, Stock et al. (2001) utilized 28S rRNA sequence and morphology to investigate phylogenetic relationships among 21 *Steinernema* species. The use of a combined data set and a larger sampling of taxa within the genus allowed for the construction of a robust evolutionary framework for the genus *Steinernema*, upon which hypotheses of species boundaries and the evolution of morphological features were assessed. A striking example of the explanatory power of comparative methods applied to entomogenous nematodes is that of Campbell et al. (2003). Using the relationships inferred from the Stock et al. (2001) phylogenetic study of *Steinernema*, Campbell et al. (2003) mapped behavioral, ecological, and morphological characters onto a *Steinernema* tree to assess the origin, maintenance, and evolution of inter-specific variation in

these traits. Mapping of host finding strategies supported the inference that the ancestral *Steinernema* species was an intermediate forager, and that two other feeding strategies, ambush and cruise foraging, evolved only once.

Though not as extensively studied as their plant parasitic counterparts, relationships within and among insect parasitic tylenchid (Hexatylina) genera have been investigated using both ribosomal and mitochondrial DNA loci. To investigate the relationships within the genus *Fergusobia*, Ye et al. (2007) utilized SSU data to determine that *Howardula* was the sister taxon to *Fergusobia*. Subsequently, LSU, mitochondrial cytochrome *c* oxidase subunit I (COI), and a combined analysis of both LSU and COI sequences were employed to construct a phylogenetic framework for *Fergusobia* spp., facilitating an investigation of how plant-host associations evolved. Analyses of relationships within *Fergusobia* provide substantial evidence for host switching within the genus *Fergusobia* with gall types being a labile feature (Ye et al., 2007). While it has been shown that the LSU alone has been unable to adequately resolve relationships between the three subfamilies of Hexatylina first proposed by Chizhov (2004) (Subbotin et al., 2006a), the LSU region may be best suited for resolving relationships at the species level (Ye et al., 2007).

Internal Transcribed Spacer (ITS) Region

Investigations into the utility of ITS regions embedded within the rRNA transcription unit, ITS-1 and ITS-2, indicate that they evolve at a much higher rate than the 18S and 28S genes, making these regions ideal for phylogenetic studies at the species and population levels, population genetic studies, as well as taxonomic identification (Powers et al., 1997, Ferris et al., 1993, Chilton et al., 1995, Cherry et al., 1997). Furthermore, presence of conserved flanking

regions encoding the 18S, 5.8S, and 28S rRNA gene products allow for the reliable amplification of both of the ITS regions (Hillis and Dixon, 1991).

Multiple studies have utilized all or a portion of ITS rDNA to investigate phylogenetic relationships of a number of different entomopathogenic and other parasitic nematode genera with varying success. Spirodonov et al. (2004) utilized the whole ITS rDNA region to analyze relationships between groups and species of steinernematids using a phylogenetic framework constructed using parsimony. Spirodonov et al. (2004) concluded that while providing new information about the composition of five main clades within the genus *Steinernema*, the ITS rDNA region was of little value for resolving relationships between clades. A similar study by Nguyen et al. (2001) utilized fewer taxa to evaluate the utility of the ITS rDNA region in identifying species, reconstructing evolutionary histories, and delimiting species within the genus *Steinernema*. It was concluded that while suitable for species identification, ITS rDNA is too variable to resolve relationships between all *Steinernema* species. Adams et al. (1998) utilized the ITS1 region to infer phylogenetic relationships among *Heterorhabditis* spp. The study indicated that ITS1 sequences resolved relationships among sister *Heterorhabditis* taxa better than it resolved larger clades within the genus. Finally, the high rate of sequence evolution within *Howardula* data caused significant alignment difficulties for Perlman et al. (2003), and as such 18S and COI data were primarily used in the inference of interspecific relationships for this genus. Each of these examples point to the need for multiple loci exhibiting varying levels of variation to reliably infer relationships among deep and shallow nodes of phylogenetic trees.

Mitochondrial DNA

Application of mitochondrial DNA (mtDNA) sequence analysis to the study of nematode population and evolutionary biology was first reviewed by Hyman (Hyman et al., 1988). Since that time, 27 complete nematode mitochondrial genome sequences have been deposited in GenBank. These circular molecules, ranging in size from 12.6-39 kb (Tang and Hyman, 2007) typically encode 12 protein-coding genes (cox1-cox3, nad1-nad6, nad4L, cob, atp6), 22 transfer RNAs, and two ribosomal RNAs (rrnS and rrnL) (Hu and Gasser, 2006). The vertebrate parasitic nematode, *Trichinella spiralis* mtDNA encodes an additional protein coding gene, atp8 (Lavrov and Brown, 2001). The mitochondrial genome maintained within *Globodera pallida* (a plant-plant parasitic nematode) is not a single circular molecule but instead the mtDNA is multipartite in architecture, with these same mitochondrial genes distributed among several sub-genomic circles (Armstrong et al., 2000).

Favorable aspects of mtDNA analysis for population and evolutionary studies include an accelerated rate of nucleotide substitution at levels measured to be 10-100 times that nuclear DNA, asexual transmission through maternal lineages, and infrequent recombination events. Once thought to be absent, nematodes were among the first systems in which animal mitochondrial DNA recombination was demonstrated (Lunt and Hyman 1997, Piganeau et al., 2004,). Unlike many taxa, nematode mitochondrial gene orders can vary considerably. While the Chromadorean nematodes show some degree of syntenic relationships among mitochondrial genes, no two Enoplean nematodes share the same gene order (Tang, 2006). Given an accelerated degree of mtDNA rearrangement coupled with considerable nucleotide substitution among mitochondrial gene orthologs (Powers et al., 1993), it is difficult to design universal primers for amplification across a wide range of taxa.

Mitochondrial genes are evolving at different rates; therefore, it becomes necessary to identify loci that diverge at a rate that provides signal useful to the question being addressed. MtDNA loci that evolve slowly, such as COI is best suited to deeper lineage phylogeny, such as affinities between genera, as more rapidly evolving genes would obscure ancestral affinities. However, more rapidly evolving mitochondrial genes can often distinguish between congeners.

Steinernematidae and Heterorhabditidae

With respect to entomopathogenic nematodes, Liu et al. (Liu et al., 1999) evaluated the mitochondrial ND4 gene as a phylogenetically informative marker for 15 heterorhabditid isolates representing five species. Seven mtDNA sequence haplotypes were identified that could be catalogued into four distinct groupings; the study concluded that ND4 is able to reveal inter and intra-specific differences within the genus *Heterorhabditis,* and that the molecular phylogeny constructed using ND4 divergence supports an existing, morphology-based taxonomic framework. Furthermore, the level of ND4 sequence substitution is such that geographic variation within certain species of *Heterorhabditis* (*H. megidis* and *H. indica*) can be resolved. Finally, the 1999 study by Liu et al. also suggests that the ND4 region may provide more robust phylogenetic information as compared to the ITS1 region, due to the requirement for gap insertion necessary to align the ITS1 region.

Intraspecific variation with the heterorhabditid ND4 locus has also been exploited to study the genetic structure of *H. marelatus* isolates (Blouin et al., 1999). Four ND4 sequence haplotypes were identified among some 60 total individuals representing six populations distributed along the Oregon and California coasts. ND4 nucleotide sequence diversity was analyzed by standard population genetic methodologies to estimate gene flow and effective

population sizes. The *H. marelatus* populations were genetically structured; a high proportion of the genetic diversity could be apportioned between populations, suggesting small population sizes and minimal gene flow, characters expected of insect parasites with little opportunity for migration.

The complete mitochondrial genome of the entomopathogenic nematode *Steinernema carpocapsae* has recently been determined (Montiel et al., 2006). *Steinernema carpocapsae* and its congeners parasitize a wide variety of insect pests, and are often used in biological control strategies. The mtDNA sequence was determined for the purpose of identifying genetic markers for use in field applications (Montiel et al., 2006). When placed in a phylogenetic context, the *S. carpocapsae* mtDNA sequence reveals more affinity to that of *Ascaris suum* (a vertebrate parasite) and *Caenorhabditis elegans* (free-living bacteribore nematode) relative to *Strongyloides stercoralis* (a vertebrate parasite), a result that stands in contrast to their phylogenetic position based on nuclear SSU rDNA data. The recent availability of the *S. carpocapsae* mitochondrial genome sequence has not yet enabled its use in population studies.

Interestingly, the mitochondrial protein and rRNA gene order of *S. carpocapsae* is identical to that of its fellow rhabditids *Ancylostoma duodenale, C. briggsae, C. elegans, Cooperia oncophora*, *Necator americanus*, and *Haemonchus contortus*, as well as the ascarids *Anisakis simplex* and *Ascaris suum*. However, *S. carpocapsae* is not completely syntenic to the only other entomopathogenic rhabditid mtDNA characterized to date, *H. bacteriophora.* Rather, these two insect parasitic nematodes diverge at a few gene junctions (Tang, 2006). As such, attempting to map life history traits onto a phylogeny based on mitochondrial gene order is not always a simple exercise, likely a result of the rapid changes in mitochondrial gene order discussed earlier.

Beyond nucleotide sequence divergence, short repeated sequences, often involving mtDNA non-coding regions, have proven to be markers useful for nematode population genetics. The first such application of variable number tandem repeats (VNTR) to population genetic analysis (Whipple et al., 1998) involved measuring the copy number a 63 base pair (bp) tandemly repeated sequence within the *M. incognita* mitochondrial genome. The number of 63 bp repeat copies ranged from 1-21 within individual mitochondrial genomes, thus defining 21 different alleles. Hierarchical statistical treatment of allele frequencies revealed that most of the genetic diversity resides within individuals, with little differentiation among populations. Diversity of mtDNA molecules within individuals is primarily due to an elevated rate of "mutation" to different copy numbers as nematodes progress through their life stages. Hypervariation at this level has also been observed within representative genera of the nematode family Mermithidae.

Mermithidae

The Mermithidae is the only taxonomic order within the Enoplea that have evolved obligate invertebrate parasitism. Mermithid nematodes parasitize a wide range of invertebrates, with insects being the most common hosts. Several have been used as biological control agents with a strong emphasis on mosquito management, and as such are considered entomopathogens.

Within the Mermithidae, mtDNA variation is not a consequence of simple VNTR copy number changes. Rather, lengthy (>1 kilobase) expanses have become repeated, often incorporating mitochondrial gene coding sequences. In the absence of selective pressure, loss-of-function mutations can accumulate in all but one gene copy. These alterations are in the form

of base substitutions, deletions, and inversions to form degenerated pseudogene copies. All mermithid nematode mtDNAs characterized to date contain such large repeat regions.

Such a complex locus resides within the mitochondrial genome of the isopod parasitic nematode *Thaumamermis cosgrovei* (Tang and Hyman, 2007). Most mitochondrial genes are mapped to a common skeleton shared by all *T. cosgrovei* individuals. The remainder of the mtDNA is occupied by a hypervariable region containing duplicated pseudogene copies of the ATP6 and ND4 genes, four mitochondrial tRNA genes, and one or more functional and pseudogene copies of the small ribosomal rRNA (rrnS) gene. These intact or degenerated gene copies are interspersed with a variety of non-coding sequences that themselves have been duplicated and have accumulated substitutions. Deletions and inversions, along with bases substitutions have resulted in a seemingly endless ensemble of variations, detectable as different banding patterns on electrophoretic gels after restriction enzyme digestion of rolling circle amplified mtDNA (Tang and Hyman, 2005) from individual nematodes.

Haplotype hypervariation can be used to better understand interesting questions in nematode life history and population structure. Typically, isopod hosts are infected by a single *T. cosgrovei* individual. Between 5 and 10% of the hosts are multiply infected with 2 – 16 nematodes. Little is known of the mechanism by which multiple infections occur. Do these represent independent spatial or temporal parasitism events by genetically unrelated nematodes? Are the parasites genetically related, indicating simultaneous infection of the host?

When isopod hosts are parasitized by multiple nematodes, individuals infecting the same host share the same mtDNA haplotype, indicating that they are derived from the same maternal lineage (Tang and Hyman, 2007). Sharing of mtDNA haplotypes has been observed in 80% of the cohorts dissected from multiply infected hosts. Kaiser (Kaiser, 1991) described two general

routes for multiple infection by mermithids that would include the entomopathogens in this nematode family. One proposed mechanism is passive infection that involves host ingestion of an egg clutch containing unhatched J1-stage nematodes. A second possible route to parasitism, termed active infection, suggests that hatched, J2-stage infectious individuals independently infect a single host. That 80% of the parasitized hosts contain individuals with identical mtDNA haplotypes, the only documented occurrence of shared mitochondrial haplotypes within *T. cosgrovei*, indicates that passive infection is the most frequent mode of infection, though active infection can infrequently occur.

It will be exciting to learn of additional examples of mitochondrial genome hypervariation within other entomopathogenic nematodes. It is anticipated that application of mtDNA variation to population structure and life histories of entomopathogens will find an important role in integrated pest management regimes.

Methodology

Alignment Strategies

When conducting a phylogenetic analysis of molecular data, generating a robust multiple sequence alignment is critical to inferring accurate relationships. The alignment is a statement of positional homology, and tree topology is often more sensitive to alignment methodology than to the method of phylogenetic tree reconstruction that is chosen (Phillips et al., 2000, Morrison and Ellis, 1997). Multiple methods exist for the alignment of sequence data and are briefly described below.

Visual Inspection

This was the first and probably most common alignment method. This method consists of using a word processor or other sequence visualization program (i.e. MacClade, BioEdit, etc.) to view sequences and manually move lines of sequence left or right and/or insert gaps, until the investigator is satisfied with the alignment. The problem with this method is that there are no discernable criteria by which the investigator decides upon a suitable alignment, thus making this method highly subjective. While this method may work well for sequences that are very similar, attempts to align dissimilar sequence data will result in highly subjective and irreproducible alignments.

Pairwise alignment

Pairwise alignments can be subdivided into two methods, local and global. Local methods are typically used in a database searching and retrieval function, where the alignment tries to determine if a portion or portions of one sequence is present in another sequence (Phillips et al., 2000). This is the type of method that is used in BLAST searches that are conducted in GenBank. Global methods, which are typically conducted for alignments that will be used in phylogenetic analysis, compare the entire sequence of taxon A to the entire sequence of taxon B.

Global pairwise alignment relies on the assignment of costs to changes (transition and transversion) and gaps. The best alignment is the one that minimizes the cost. An extension of the pairwise alignment is multiple sequence alignment (below). When multiple alignment methods are used, the same cost minimization idea is applied to n-sequences in n-dimensions (Phillips et al., 2000).

Multiple alignments and most common software

Numerous multiple sequence alignment programs exist, each with its particular advantages and disadvantages. These algorithms are summarized below.

CLUSTAL

This is one of the most utilized multiple alignment programs (Thompson et al., 1997, Thompson et al., 1994). CLUSTAL utilizes a progressive alignment algorithm that first estimates a distance tree (Wheeler, 2001), which is then used to construct pairwise alignments of subtrees within the original guide tree (Edgar, 2004b). Advantages of CLUSTAL include the speed at which alignments are constructed, a friendly graphical user interface (GUI), and an output of only one multiple sequence alignment. Another helpful feature is the profile alignment mode, which allows users to align individual sequences to an already established multiple sequence alignment. Limitations to the method include a lack of guarantee that the minimum cost alignment is found, practicality in that a finite number of sequences can be analyzed (typically 500, but varies with computer platform and computational power), and the output represents a single alignment when many equally scoring alternatives may exist. Furthermore, most CLUSTAL alignments often require readjustments by eye.

MALIGN

Similar to CLUSTAL, MALIGN (Wheeler and Gladstein, 1994) also utilizes a guide tree to chaperon the alignment process. MALIGN furthers the idea of using a guide tree by searching multiple guide trees in an attempt to find an optimally minimized cost (parsimonious) alignment. Multiple guide trees are searched in a manner similar to a phylogenetic tree search, where branchswapping and random addition of taxa are used. While MALIGN will generally find more

optimal solutions than CLUSTAL, considerably more computational power is employed in constructing the optimal alignment. Occasionally both methods will recover multiple alignments of equal score.

MAFFT

MAFFT was developed to increase efficiency (speed of computation and accuracy) of the multiple alignment process (Katoh et al., 2002). Using a fast Fourier transform (FFT), MAFFT is able to quickly identify homologous regions of DNA sequence. Along with the FFT, MAFFT also employs a scoring system that is different from earlier alignment programs such as CLUSTAL and T-COFFEE. The MAFFT scoring system is touted as enabling the accurate alignment of sequence data with large insertions or extensions, as well as highly divergent sequence data of similar length (Katoh et al., 2002).

MUSCLE

MUSCLE also utilizes a progressive alignment algorithm, but refines the alignment procedure by applying a horizontal process to its initial progressive alignment (Edgar, 2004a, Edgar, 2004b). This improves the initial guide tree in an attempt to find the lowest scoring guide tree for use in directing the alignment process. The improvements made to the progressive alignment process allow MUSCLE to compute alignments of large numbers of taxa (several thousand) in a shorter amount of time, with enhanced biological accuracy relative CLUSTAL and MAFFT. One drawback of MUSCLE is its use of a command line driven interface, making it less user-friendly than CLUSTAL.

Direct optimization (DO)

First proposed by Wheeler (1996), DO strays from the typical multiple alignment procedure where the alignment is conducted in one step and is followed by the construction of a phylogenetic tree in a separate step. Instead, DO constructs the alignment and the phylogenetic tree concurrently, thus applying the same optimality criterion to alignment and tree construction, a feature that is lacking in all other alignment and tree reconstruction methods. DO completely eliminates the multiple alignment procedure, and instead constructs alignments at each node of the phylogenetic tree (Wheeler, 1996). Due to large numbers of possible topologies and possible ways in which sequences can be optimized to those topologies, DO presents an exceedingly computationally complex problem (Terry and Whiting, 2005). The DO method for the construction of phylogenies can be carried out using the software package POY (Wheeler, 2003).

ProAlign

Another approach with promise for analysis of rDNA sequences is ProAlign (Loytynoja and Milinkovitch, 2003). This approach uses a hidden Markov model, a progressive alignment algorithm (above), and a nucleotide substitution model to identify the minimum posterior probability of each homology statement. Thresholds can be explored and established that allow the user to exclude characters with a low posterior probability from the multiple sequence alignment. This approach reduces the potential for investigator bias regarding the identification and removal of alignment-ambiguous characters, and also allows for clever comparisons among datasets where the effects of indel inclusion/excision are concerned (Nadler et al., 2006a).

Secondary structure models

Two-dimensional secondary structure modeling of rRNAs can be used to inform homology statements. Ribosomal RNAs have base paired stem and unpaired loop regions that result in different constraints and substitution rates among bases that comprise these two structural features. There are several methods for inferring secondary structure, including minimum free energy, base pair probabilities, and comparing secondary structures as they vary across a broad range of energies (Hofacker, 2003). Since each of these approaches can yield different optimal secondary structure models, there remains a continuum of multiple sequence alignments that should be evaluated prior to phylogenetic analysis. A library of nematode sequences based on modeled secondary structures for nematodes (ITS and LSU D2, D3 regions) is available at http://www.nemamex.ucr.edu/rna/ (Subbotin et al., 2006b). Phylogenetic analyses of these can be run in Phase, which uses explicit models to account for substitution events that are influenced by secondary structure (Telford et al., 2005).

Phylogenetic Reconstruction Methods

Parsimony

Phylogenetic reconstruction methods can be divided into two main categories; parsimony methods and model-based methods. Parsimony, the most widely used method, is based on the principle that the simplest explanation is the explanation best supported by the current data. Accordingly, the optimal phylogenetic tree is the one that minimizes the number of ad-hoc hypotheses required to explain the data. Model-based methods, such as maximum likelihood and Bayesian inference assume a model of DNA sequence evolution and then find the tree that best fits the model.

As mentioned, the first and most important step in the construction of phylogenies is the alignment, as this is the homology statement. Everything that occurs after the alignment is directly dependent on the accuracy of this step. When conducting analyses under the parsimony criterion, the most widely used software package is PAUP* 4.0b10 (Swofford, 2002). PAUP* features a GUI, which allows users to quickly specify parameters and run analyses without the need for in-depth knowledge of command line prompts and keywords. NEXUS files that have been directly outputted from alignment programs or exported from MacClade (Maddison and Maddison, 2002) can be loaded directly into PAUP*. Parsimony trees are built via a multi-step process that utilizes Fitch optimization, a method by which the cost of a tree is calculated. The number of character state changes is calculated for each tree to determine which tree has the lowest score. The tree with the lowest score is the most parsimonious (optimal) tree.

Parsimony analyses can be refined through the specification of a number of different tree search parameters, including the type of search algorithm employed, the form of branch swapping that is conducted, and how taxa are added to each reconstruction. PAUP* offers heuristic, branch and bound, and exhaustive search algorithms. The exhaustive search guarantees that the globally optimal topology will be discovered by examining every possible topology in the landscape of trees. This is a very computationally expensive method, and thus is only practical for a data set of 20 taxa or less. The branch and bound algorithm also guarantees obtaining the most parsimonious solution, but is much faster because it bypasses known suboptimal topologies. An even less computationally intensive method is the heuristic algorithm. The heuristic search algorithm takes samples (local optima) from the tree landscape, with the idea that if you sample enough local optima, one of them will likely be the global optimum. The heuristic method is useful when working with data sets with more than 20 taxa.

The heuristic search method utilizes multiple methods of branch swapping and tree construction methods to search tree space in an attempt to find the globally optimal topology in a vast forest of trees. To begin a parsimony analysis, an initial tree is constructed using multiple options for the formation of the initial tree provided within PAUP*. Starting tree construction options include neighbor joining, a distance based method of tree construction, and stepwise addition. The choice of stepwise addition offers further taxa addition options, including asis, simple, closest, and random. Asis adds taxa to the initial tree in the order they are listed in the data matrix. The preferred method of stepwise addition is random addition, which randomly adds taxa to the tree. It is generally suggested that a minimum of 1000 random addition replicates be used in the construction of the initial tree.

Occasionally, a heuristic search will get stuck on a locally optimal solution within the tree space. Branch swapping provides a method by which new reconstructions are proposed, thus enabling new parts of the tree space to be explored and the discovery of trees that are more optimal than previous reconstructions. PAUP* offers three main branch swapping methods; nearest neighbor interchange (NNI), subtree pruning and re-grafting (SPR), and tree bisection and reconnection (TBR). Nearest neighbor interchange utilizes subtrees that make up the larger tree. Each subtree has two neighbors, and by swapping a neighbor from one subtree with a neighbor from an adjacent subtree (Felsenstein, 2004), a new arrangement of the taxa is produced. SPR is similar to NNI in that it also utilizes subtrees to form new topologies. However, SPR removes one branch, along with its subtree, and forms new arrangements by re-inserting the removed subtree in all possible places within the tree (Felsenstein, 2004). TBR breaks a tree into two separate trees, and subsequently proceeds to assemble new tree rearrangements by attaching a branch from one tree to a branch from the other (Felsenstein,

2004). While TBR is the more computationally intensive of the three methods described above, it generally finds a more optimal solution.

Branch swapping methods, while useful for smaller data sets (< 200 taxa), do not work as well for large data sets, which produce large numbers of suboptimal trees, many of which are similar in topology and length (Nixon, 1999). These large groups of trees are known as "islands of trees", and branch-swapping methods can get marooned on these large islands. The parsimony ratchet (Nixon, 1999), which is implemented in the software package TNT (Goloboff et al., 2008) and can also be run in PAUP*, reduces the problem of becoming trapped on "islands", and finds optimal trees for very large datasets (> 500 taxa) much faster than other methods. By maximizing starting points, reducing the amount of time spent swapping on each starting point, and retaining structure from the existing solution at each point, the parsimony ratchet allows for the majority of the computing time to be spent breaking out of tree islands and improving the current tree (Nixon, 1999).

While a phylogeny alone shows inferred relationships, estimates of support can be generated for each node (monophyletic group) in the tree. Multiple categories of branch support methods exist, including incongruence, support indices, and statistical resampling methods. Non-parametric bootstrapping is a method that relies on re-sampling characters, with replacement, to estimate confidence limits of internal branches (Hillis and Bull, 1993) and is incorporated in many phylogeny reconstruction software packages. Bremer support, also called the decay index, is a measure of the number of extra steps beyond the most parsimonious tree that must be allowed before trees are found which do not include the monophyletic group of interest (DeBry, 2001). Bremer support can also be defined as the number of steps required to dissolve a node. While Bremer support values can be informative, their interpretation can be misleading. The

number of informative characters present at each node will have a large effect on the interpretation of a Bremer support value (DeBry, 2001), and thus should always be taken into account when evaluating the utility or measure of support that is indicated by a specific Bremer support value. In analyses that incorporate multiple data sets (i.e. sequence data for multiple loci), partitioned Bremer support values, or Bremer support values for each data set, can be calculated. Bremer support values can be calculated using PAUP*, AutoDecay, TreeRot, and many other programs. Partitioned Bremer support values, calculated using the program TreeRot (Sorenson, 1999), allow for the relative amount of support each data set contributes to each node to be assessed. Thus, one can tell if a particular gene gives more support to terminal taxa or to internal nodes of the tree. This information may be useful when examining conflicts among partitions.

Maximum Likelihood

Maximum likelihood is a model-based method of phylogenetic reconstruction. That is, it relies on a specific model of sequence evolution to infer the probability of the data given a particular phylogenetic tree. Critics of likelihood methods assert that even the most general and parameter rich models cannot possibly capture all of the processes that generated a particular sequence or sequences (Sullivan and Swofford, 2001), thus hampering the ability of the method to recover the 'true' tree. Likelihood analysis of molecular data begins by constructing a suitable alignment of the sequence data. The alignment is then input into MODELTEST (Posada and Crandall, 1998), or other model selection programs that determine the best-fit model of evolution for the specified group of sequences. This step is important, as the selection of an incorrect model may lead to the recovery of an incorrect tree (Posada and Buckley, 2004). MODELTEST

selects models using two different criteria, the likelihood-ratio test (LRT) and the Akaike information criterion (AIC). The LRT, while possessing unfavorable qualities such as a propensity to always select the most complex model, has nonetheless been the most widely used model selection criteria. Recently, the AIC has gained popularity, likely due to its ability to calculate an AIC for each model in isolation and the inclusion of penalties for over parameterization (Sullivan and Joyce, 2005). Model parameters output from ModelTest are inserted at the end of a PAUP* block of the corresponding aligned sequence data, and executed in PAUP*. Likelihood analysis proceeds by simply switching the optimality criterion to likelihood, and selecting a tree searching method. Likelihood analyses conducted in PAUP* are typically very time intensive, with analyses running for weeks or even months before completion. Other, less time intensive software for conducting likelihood analyses include GARLI (http://www.zo.utexas.edu/faculty/antisense/Download.html), a program that utilizes a genetic algorithm approach to efficiently and accurately find the most likely tree, TREEFINDER (Jobb, 2008), a fast method that allows the use of partitioned data and bootstrap calculations, and PHYML (Guindon and Gascuel, 2003) which builds an initial tree using a simple hill climbing algorithm, then modifies the topology and branch lengths simultaneously and progressively to find the optimal tree.

Bayesian Analysis

Bayesian phylogenetic inference is another model-based tree reconstruction approach. Bayesian inference utilizes the Metropolis coupled Markov chain Monte Carlo (MCMCMC) method to randomly explore the tree space so as to settle down into an equilibrium distribution of trees with the desired distribution (Felsenstein, 2004). Bayesian phylogenetic inference has been

lauded for its ability to incorporate priors, to handle large numbers of taxa, and efficiently sample tree space.

MRBAYES (Huelsenbeck and Ronquist, 2001) is a command line driven program that can be used to conduct Bayesian analysis of phylogenies. MRBAYES reads aligned DNA or amino acid sequences in a NEXUS formatted file. As with a likelihood analysis, the specific parameters for the best fit model of evolution should be included at the end of the alignment file, along with the number of chains to be used, total generations, number of runs, and frequency with which trees are saved. Upon completion of a Bayesian analysis, the burn-in value (defined below) must be calculated and all trees falling within the burn-in phase of the analysis must be removed prior to calculating a Bayesian tree.

The burn-in refers to all of the tree topologies prior to the point at which the equilibrium distribution of trees is reached. To determine when the analysis reaches equilibrium, the likelihood values file (file will end in .p) is opened in Excel or any other spreadsheet program. The numbers in the first two columns are plotted as a scatter plot to identify the point on the plot where the distribution levels out. The x value of this point is the burn-in value. Since the burn-in contains trees that are not in the desired distribution, these trees will be eliminated from the tree file (file will end in .t) following the completion of a full Bayesian run. This process can also be conducted in Tracer v1.3 (Rambaut and Drummond, 2003). Tracer allows the user to input multiple ".p" files (from different independent runs) at the same time, allowing visualization of the point where all runs reach stationarity as well enabling the user to ensure that all runs converged to a similar likelihood value. While one could build a Bayesian tree from a single run, a more thorough method would be to conduct multiple runs, ensure all runs converge on a similar likelihood score and calculate the burn-in for each run, eliminate the burn-in from

the tree files (.t file), combine the tree files from each independent run into a single tree file, and then compute the final Bayesian tree (majority rule consensus tree).

The number of generations that are considered suitable for an optimal run in a Bayesian analysis varies based on individual datasets, though no less than 10 to 20 million should probably be run. This depth allows for sufficient search of the tree space, and detects whether a second set of desired tree distributions exists. Similarly, the number of runs and the number of chains used in an analysis will also vary based on knowledge of the dataset in question as well as available computing power, though more runs and more chains (>4) will only improve the analyses.

MrBayes 3 (Ronquist and Huelsenbeck, 2003), a rewritten and restructured version of MrBayes, enables the incorporation of mixed models into a Bayesian analysis. This strategy is beneficial for studies that contain heterogeneous sequence data. The use of mixed models in a simultaneous analysis allows for the partitioning of a gene or genes based on the best fit model of evolution of each gene or gene region, thus allowing each data set to be analyzed independently according to its specific best fit model.

To illustrate the benefit of a mixed models analysis in MrBayes 3 over a single model analysis in MrBayes, consider a hypothetical situation where a phylogeny of *Heterorhabditis* is constructed using 18S rRNA, ITS1 rRNA, and ND4 data. If this analysis were run in MrBayes, ModelTest would first be conducted on the combined data set of the three genes to identify a single best-fit model. To conduct a mixed models Bayesian analysis, the data set for each individual gene is run through ModelTest, resulting in a specific model for each gene. Since both nuclear ribosomal gene and a mitochondrial protein-coding locus are being used in the hypothetical study, the chance that a single model sufficiently accounts for the rates of nucleotide evolution in all three genes is low. Thus, an analysis that applies a best-fit model to

each gene has a higher probability of recovering the 'true' tree than an analysis than one that utilizes one model for all three genes.

Bayesian analysis produces a unique form of branch support known as posterior probabilities. A posterior probability represents the number of trees (represented as a percentage in decimal form) that supported the grouping of the clade of interest. Care should be taken when interpreting Bayesian posterior probability values, as posterior probabilities can be potentially inflated, especially relative to bootstrap support values (Pérez-Losada et al., 2007).

Co-Phylogenesis and Cospeciation

General Concepts

Cospeciation, the joint speciation of two organisms living in close association with one another, has been detected in numerous systems, including parasitic (Perlman et al., 2003, Hafner and Nadler, 1988) and mutualistic (Clark et al., 2000) relationships. Assessment of cospeciation is carried out through the use of phylogenetics, and more specifically co-phylogenetic analyses. Through the process of a co-phylogenetic analysis, phylogenies of both the host and associate are compared and the amount of congruence between the two phylogenies is assessed. The null hypothesis in cospeciation studies is one of strict congruence or strict cospeciation. As such, the phylogeny of the associate mirrors the phylogeny of the host, though this is a rare occurrence (Downie and Gullan, 2005, Johnson et al., 2003). A lack of correspondence between host and associate phylogenies can be explained by numerous other events. These events include host switching, duplications, and sorting events. Sorting events are events that account for the absence of a parasite lineage on a host. These can include instances of 'missing the boat' or

extinction events. Duplications refer to an occasion where the parasite speciates in the absence of host speciation.

Methodology

Numerous methods exist to assess the amount of phylogenetic congruence between host and associate phylogenies. One of the earliest methods is Brooks parsimony analysis or BPA (Brooks, 1981), a method that utilizes parsimony and the Wagner algorithm, as well as additive binary coding of the parasite phylogeny to detect cospeciation (a strategy also used in studies of historical biogeography) (Dowling, 2002). BPA analysis of coevolution begins by converting all terminal taxa and nodes of the parasite phylogenetic tree into a binary character matrix (Brooks, 1981). The parasite names in the character matrix are then replaced with the name of the host known to associate with each parasite. The end result is a phylogeny that minimizes the number of host-switching events and extinctions (Dowling, 2002). This earliest version of BPA has been modified several times since its inception to further the development of this method for conducting cospeciation and historical biogeographical analyses (Brooks, 1990, Wiley, 1988a, Wiley, 1988b).

Software

COMPONENT (Page, 1993)

This algorithm is utilized for the comparison of phylogenetic trees. Thus, this software is not specific for cospeciation studies, and like BPA, can also be used to conduct studies of historical biogeography. COMPONENT employs the "tree reconciliation" method developed by Page (1990), which uses duplication and loss events to fit the parasite tree to the host tree

(Slowinski, 1993). One criticism of COMPONENT is its inability to allow for host switches (Charleston, 1998).

TreeMap (Page, 1994)

Similar to COMPONENT, this software program is also a reconciliation-based method of cophylogenetic analysis (Page, 1994). As with all other cophylogenetic analysis methods, the goal of TreeMap is to maximize the number of cospeciation events, while minimizing the number of ad-hoc hypotheses in the form of duplications and sorting events. The reconciliation process used in TreeMap is executed by labeling all internal nodes and terminal tips of both the host and associate tree. Each node on the parasite tree is mapped onto a corresponding node in the host tree, and paths are drawn to trace the path between the corresponding host tree and the parasite tree (Dowling, 2002). From this process, the number of cospeciation, duplication, and sorting events are determined. Included in the TreeMap v1.0 program is an option that allows for testing of the least costly reconciliation against a number of randomly generated trees, to determine if the result obtained is statistically significant from chance alone.

Some common criticisms of TreeMap include its prohibition of host-switching, the requirement of input trees that are fully resolved, and its overestimation of duplications and sorting events (Dowling, 2002, Ronquist, 1995). A later release of TreeMap, version 2.0 (Charleston and Page, 2002), made major improvements to the tree reconciliation process through the use of a method called "jungles". One of the improvements is the inclusion of more event types, including host-switching, lineage sorting, duplications, and of course, cospeciation events (Charleston, 1998).

Case Studies

Perlman et al. (2003) investigated associations between *Drosophila* and its nematode parasite *Howardula*. Following construction of both host (COI, II, and III) and parasite (18S, ITS, and COI) phylogenies, tree reconciliation analysis performed using TreeMap v1.0 demonstrated that *Drosophila* and *Howardula* phylogenies were not congruent, as indicated by a lack of statistically significant cospeciation events. Perlman et al. (2003) conclude that incongruence between host and parasite phylogenies is most likely attributed to a relatively high degree of host switching and infection of novel hosts other than *Drosophila*. Conversely, preliminary analyses of cospeciation between *Heterorhabditis* and its bacterial endosymbiont *Photorhabdus* (Peat et al., 2006), indicates the presence of statistically significant cospeciation within this highly specialized symbiotic system.

Entomopathogenic nematodes and their symbiotic bacteria

The association all entomopathogens have with insects provides unbounded opportunities to test the hypothesis that host and associate phylogenies mirror one another. Furthermore, the close associations of *Heterorhabditis* and *Steinernema* with their respective bacterial endosymbionts (*Photorhabdus* and *Xenorhabdus*) also provide research avenues that could uncover clues to the origin and maintenance of these nematode/bacterium associations. Numerous resources and opportunities exist for studying the coevolution of entomopathogenic nematodes and their associates (insects and/or bacteria), driving this nascent but rapidly growing area of research.

Population Genetics Methods

Software and analysis of data

The following section describes population genetics software that can be used to address questions focusing on population structure, demography, genetic diversity, gene flow, linkage disequilibrium, and selection on genomes. When using any of these programs, it is imperative that the assumptions used in each algorithm are fully understood to enable accurate and meaningful interpretations of data outputs. As a complete review of all contemporary, computational approaches to population genetics is beyond the scope of the present discussion, we suggest (Labate, 2000) and (Excoffier and Heckel, 2006) for a more exhaustive review of population genetics software applications.

Studies investigating properties of populations often begin by using descriptive statistics. Common approaches include measures of genetic diversity, linkage disequilibrium (LD), and tests of Hardy-Weinburg equilibrium (HWE). Genetic diversity, a measure of variation within populations, measures the numbers of polymorphic loci, catalogues distinct haplotypes and allele frequencies, and proportions of heterozygotes. Several software programs, including Arlequin (Excoffier and Schneider, 2005), DnaSP (Rozas and Rozas, 1999, Rozas and Rozas, 1997, Rozas and Rozas, 1995), GDA (Lewis and Zaykin, 2001) and Genepop (Raymond and Rousset, 1995), offer calculations of these and other measures of diversity. Arlequin and Genepop provide tests of HWE and LD, and DnaSP and GDA compute LD indices (Excoffier and Heckel, 2006). Detection of population structure and measures of population subdivision (using F-statistics) are common in a number of broad spectrum and specific population genetics software programs.

STRUCTURE (Falush et al., 2003, Pritchard et al., 2000)

This program was designed to detect genetic structure present in a set of individuals in the absence of user-defined population information, and can be utilized to determine the number of unique populations that exist from a data set of individuals. It can assign individuals of unknown origin to a pre-defined population, or to identify cryptic population structure (Pritchard et al., 2000). Other programs available to detect population structure and/or population subdivision include GDA, Arlequin, and DnaSP.

LAMARC v 2.0.2 (Kuhner et al., 2005)

This program couples the primary abilities of four different programs (MIGRATE, FLUCTUATE, COALESCE, and RECOMBINE) into one interface. LAMARC estimates effective population sizes, population exponential growth rates, divergence times, recombination rates, and past migration rates for one to n populations using single nucleotide polymorphism (SNP), microsatellite, DNA or RNA data (Kuhner et al., 2005). A useful feature provided by LAMARC is the estimation of population size or migration rates between more than two populations (Excoffier and Heckel, 2006), a feature not offered in most population genetics programs. Estimation of migration rates requires the separation of each population into its own separate data file using either PHYLIP or MIGRATE file formatting.

IM (Nielsen and Wakeley, 2001)

IM is another program well-suited for inference of population size, divergence times, and migration rates. IM furthers the estimation of migration rates by jointly estimating divergence times and migration rates from DNA sequence data (Nielsen and Wakeley, 2001). Other

programs that can be used to estimate migration rates and divergence times include Arlequin, Genepop, and DnaSP.

User interface, computation time, input file format, accepted data types, and model assumptions vary across all population genetics software. It is important to understand what population related questions will be addressed and what programs can be used to answer proposed research questions, prior to selecting a particular population genetics program. If only descriptive statistics are required, Arlequin is the most appropriate, leaving STRUCTURE for more sophisticated analyses. Defining hypotheses prior to beginning population genetic analyses will facilitate proper selection of appropriate tests, narrowing the choice of software, and allowing for a more refined selection of appropriate tests/software based on data type (i.e. microsatellite, SNP, etc.) and fit of assumptions to the system being analyzed.

We have discussed only a few of the phylogenetic and population genetic analysis programs available; all computer programs mentioned in this paper, and many more, are available online from the following websites:

http://evolution.genetics.washington.edu/phylip/software.html

http://www.biology.lsu.edu/general/software.html

http://evonet.sdsc.edu/ROADS/subject-listing/softwrpopgen.html

http://pritch.bsd.uchicago.edu/software/structure2_1.html

References

Adams, B. J., Burnell, A. M. & Powers, T. O. (1998) A phylogenetic analysis of *Heterorhabditis* (Nemata: Rhabditidae) based on internal transcribed spacer 1 DNA sequence data. *Journal of Nematology,* 30**,** 22-39.

Adams, B. J., Fodor, A., Koppenhöfer, H. S., Stackebrandt, E., Stock, S. P. & Klein, M. G. (2006) Biodiversity and systematics of nematode-bacterium entomopathogens. *Biological Control,* 37**,** 32-49.

Andrassy, I. (1976) *Evolution as a basis for the systematization of nematodes,* London and San Francisco, Pitman Publishing.

Armstrong, M. R., Blok, V. C. & Phillips, M. S. (2000) A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. *Genetics,* 154**,** 181-192.

Avise, J. C. (2000) *Phylogeography : the history and formation of species,* Cambridge, Mass., Harvard University Press.

Avise, J. C. (2004) *Molecular markers, natural history, and evolution,* Sunderland, Mass., Sinauer Associates.

Baldwin, J. G., Frisse, L. M., Vida, J. T., Eddleman, C. D. & Thomas, W. K. (1997) An evolutionary framework for the study of developmental evolution in a set of nematodes related to *Caenorhabditis elegans*. *Molecular Phylogenetics And Evolution,* 8**,** 249-259.

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T. & Thomas, W. K. (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature,* 392**,** 71-75.

Blouin, M. S., Liu, J. & Berry, R. E. (1999) Life cycle variation and the genetic structure of nematode populations. *Heredity,* 83**,** 253-259.

Brooks, D. R. (1981) Hennig's parasitological method: a proposed solution. *Systematic Zoology,* 30**,** 229-249.

Brooks, D. R. (1990) Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. *Systematic Zoology,* 39**,** 14-30.

Brooks, D. R. & McLennan, D. A. (1991) *Phylogeny, Ecology, And Behavior,* Chicago, the University of Chicago Press.

Brooks, D. R. & McLennan, D. A. (2002) *The Nature of Diversity : An Evolutionary Voyage of Discovery,* Chicago, Ill., University of Chicago Press.

Campbell, J. F., Lewis, E. E., Stock, S. P., Nadler, S. & Kaya, H. K. (2003) Evolution of host search strategies in entomopathogenic nematodes. *Journal of Nematology,* 35**,** 142-145.

Charleston, M. A. (1998) Jungles: a new solution to the host/parasite phylogeny reconciliation problem. *Mathematical Biosciences,* 149**,** 191-223.

Charleston, M. A. & Page, R. D. M. (2002) TREEMAP v2.0 Application for Apple Macintosh.

Cherry, T., Szalanski, A. L., Todd, T. C. & Powers, T. O. (1997) The internal transcribed spacer region of *Belonolaimus* (Nemata: Belonolaimidae). *Journal of Nematology,* 29**,** 23-29.

Chilton, N. B., Gasser, R. B. & Beveridge, I. (1995) Differences in a Ribosomal DNA-Sequence of Morphologically Indistinguishable Species within the Hypodontus-Macropi Complex (Nematoda, Strongyloidea). *International Journal for Parasitology,* 25**,** 647-651.

Chizhov, V.N. (2004). Entomogenous nematodes from the order Tylenchida (Nematoda). In: Sonin, E.D. (Ed.). *Parasitic nematodes of plants and insects*. Moscow, Nauka, pp. 277-293.

Clark, M. A., Moran, N. A., Baumann, P. & Wernegreen, J. J. (2000) Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution,* 54**,** 517-525.

Debry, R. W. (2001) Improving interpretation of the decay index for DNA sequence data. *Systematic Biology,* 50**,** 742-752.

Dowling, A. P. G. (2002) Testing the accuracy of TreeMap and Brooks parsimony analyses of coevolutionary patterns using artificial associations. *Cladistics,* 18**,** 416-435.

Downie, D. A. & Gullan, P. J. (2005) Phylogenetic congruence of mealybugs and their primary endosymbionts. *Journal of Evolutionary Biology,* 18**,** 315-324.

Edgar, R. C. (2004a) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics,* 5**,** 1-19.

Edgar, R. C. (2004b) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32**,** 1792-1797.

Excoffier, L. & Heckel, G. (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics,* 7**,** 745-758.

Excoffier, L. G. L. & Schneider, S. (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online,* 1**,** 47-50.

Falush, D., Stephens, M. & Pritchard, J. K. (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics,* 164**,** 1567-1587.

Felsenstein, J. (1978) The number of evolutionary trees. *Systematic Zoology,* 27**,** 27-33.

Felsenstein, J. (2004) *Inferring Phylogenies,* Sunderland, Mass., Sinauer Associates.

Ferris, V. R., Ferris, J. M. & Faghihi, J. (1993) Variation in spacer ribosomal DNA in some cyst-forming species of plant parasitic nematodes. *Fundamental & Applied Nematology,* 16**,** 177-184.

Floyd, R., Abebe, E., Papert, A. & Blaxter, M. (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology,* 11**,** 839-850.

Gascuel, O. (1997) BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution,* 14**,** 685-695.

Goloboff, P.A., Farris, J.S. Nixon, K.C (2008) TNT, a free program for phylogenetic analysis. *Cladistics* 24, 774-786.

Guindon, S. & Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52, 696–704.

Hafner, M. S. & Nadler, S. A. (1988) Phylogenetic trees support the coevolution of parasites and their hosts. *Nature,* 332**,** 258-259.

Hillis, D. M. & Bull, J. J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology,* 42**,** 182-192.

Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R. & Molineux, I. J. (1992) Experimental Phylogenetics - Generation of a Known Phylogeny. *Science,* 255**,** 589-592.

Hillis, D. M. & Dixon, M. T. (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology,* 66**,** 411-453.

Hillis, D. M., Huelsenbeck, J. P. & Cunningham, C. W. (1994) Application and Accuracy of Molecular Phylogenies. *Science,* 264**,** 671-677.

Hofacker, I. L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Research,* 31**,** 3429-3431.

Holterman, M., Van Der Wurff, A., Van Den Elsen, S., Van Megen, H., Bongers, T., Holovachov, O., Bakker, J. & Helder, J. (2006) Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Molecular Biology and Evolution,* 23**,** 1792-1800.

Hu, M. & Gasser, R. B. (2006) Mitochondrial genomes of parasitic nematodes - progress and perspectives. *Trends in Parasitology,* 22**,** 78-84.

Hudson, M. E. (2007) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Notes,* doi: 10.1111/j.1471-8286.2007.02019.

Huelsenbeck, J. P. (1995) Performance of phylogenetic methods in simulation. *Systematic Biology,* 44**,** 17-48.

Huelsenbeck, J. P. & Hillis, D. M. (1993) Success of phylogenetic methods in the four-taxon case. *Systematic Biology,* 42**,** 247-264.

Huelsenbeck, J. P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics,* 17**,** 754-755.

Hyman, B. C., Beck, J. L. & Weiss, K. C. (1988) Sequence amplification and gene rearrangement in parasitic nematode mitochondrial DNA. *Genetics,* 120**,** 707-712.

Jobb, G. (2008) TREEFINDER version of March 2008. Munich, Germany. Distributed by the author at www.treefinder.de

Johnson, K. P., Adams, R. J., Page, R. D. M. & Clayton, D. H. (2003) When do parasites fail to speciate in response to host speciation? *Systematic Biology,* 52**,** 37-47.

Kaiser, H. (1991) Terrestrial and semi-terrestrial Mermithidae. IN Nickle, W. R. (Ed.) *Manual of agricultural nematology.* New York, M. Dekker.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research,* 30**,** 3059-3066.

Kovaleva, E. S., Masler, E. P., Skantar, A. M. & Chitwood, D. J. (2004) Novel matrix metalloproteinase from the cyst nematodes *Heterodera glycines* and *Globodera rostochiensis*. *Molecular and Biochemical Parasitology,* 136**,** 109-112.

Kuhner, M. K., Yamato, J., Beerli, P. & Felsenstein, J. (2005) LAMARC. 2.0.2

Labate, J. A. (2000) Software for population genetic analyses of molecular marker data. *Crop Science,* 40**,** 1521-1528.

Lavrov, D. V. & Brown, W. M. (2001) *Trichinella spiralis* mtDNA: A nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. *Genetics,* 157**,** 621-637.

Lewis, P. O. & Zaykin, D. (2001) Genetic data analysis: Computer program for the analysis of allelic data.

Liu, J. & Berry, R. E. (1996) Phylogenetic analysis of the genus *Steinernema* by morphological characters and randomly amplified polymorphic DNA fragments. *Fundamental & Applied Nematology,* 19**,** 463-469.

Liu, J., Berry, R. E. & Blouin, M. S. (1999) Molecular differentiation and phylogeny of entomopathogenic nematodes (Rhabditida : Heterorhabditidae) based on ND4 gene sequences of mitochondrial DNA. *Journal of Parasitology,* 85**,** 709-715.

Liu, J., Berry, R. E. & Moldenke, A. F. (1997) Phylogenetic relationships of entomopathogenic nematodes (Heterorhabditidae and Steinernematidae) inferred from partial 18S rRNA gene sequences. *Journal of Invertebrate Pathology,* 69**,** 246-252.

Longhorn, S. J., Foster, P. G. & Vogler, A. P. (2007) The nematode-arthropod clade revisited: phylogenomic analyses from ribosomal protein genes misled by shared evolutionary biases. *Cladistics,* 23**,** 130-144.

Loytynoja, A. & Milinkovitch, M. C. (2003) A hidden Markov model for progressive multiple alignment. *Bioinformatics,* 19**,** 1505-1513.

Lunt, D. H. & Hyman, B. C. (1997) Animal mitochondrial DNA recombination. *Nature,* 387**,** 247-247.

Maddison, W. P. & Maddison, D. R. (2002) *MacClade version 4.0.5,* Sunderland, Massachusetts, Sinauer.

Montiel, R., Lucena, M. A., Medeiros, J. & Simoes, N. (2006) The complete mitochondrial genome of the entomopathogenic nematode *Steinernema carpocapsae*: Insights into nematode mitochondrial DNA evolution and phylogeny. *Journal of Molecular Evolution,* 62**,** 211-U15.

Morrison, D. A. & Ellis, J. T. (1997) Effects of nucleotide sequence alignment on phylogeny estimation: A case study of 18S rDNAs of Apicomplexa. *Molecular Biology and Evolution,* 14**,** 428-441.

Nadler, S. A., Bolotin, E. & Stock, S. P. (2006a) Phylogenetic relationships of *Steinernema* Travassos, 1927 (Nematoda : Cephalobina : Steinernematidae) based on nuclear, mitochondrial and morphological data. *Systematic Parasitology,* 63**,** 161-181.

Nadler, S. A., De Ley, P., Mundo-Ocampo, M., Smythe, A. B., Stock, S. P., Bumbarger, D., Adams, B. J., De Ley, I. T., Holovachov, O. & Baldwin, J. G. (2006b) Phylogeny of Cephalobina (Nematoda): Molecular evidence for recurrent evolution of probolae and

incongruence with traditional classifications. *Molecular Phylogenetics and Evolution,* 40**,** 696-711.

Nguyen, K. B., Maruniak, J. & Adams, B. J. (2001) Diagnostic and phylogenetic utility of the rDNA internal transcribed spacer sequences of *Steinernema*. *Journal of Nematology,* 33**,** 73-82.

Nielsen, R. & Wakeley, J. (2001) Distinguishing Migration From Isolation: A Markov Chain Monte Carlo Approach. *Genetics* 158**,** 885–896.

Nixon, K. C. (1999) The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics,* 15**,** 407-414.

Page, R.D.M. (1990) Component analysis: a variant failure? *Cladistics* 6, 119-136

Page, R. D. M. (1993) COMPONENT: Tree comparison software for Microsoft Windows, version 2.0. London, The Natural History Museum.

Page, R. D. M. (1994) Parallel phylogenies: reconstructing the history of host-parasite assemblages. *Cladistics,* 10**,** 155-173.

Peat, S., ffrench-Constant, R., Waterfield, N. and Adams, B.J. (2006) The effect of rooting and outgroup choice in recovering significant cospeciation between *Heterorhabditis* and *Photorhabdus*. *Journal of Nematology* 38, 286.

Penny, D., Steel, M., Waddell, P. J. & Hendy, M. D. (1995) Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*. *Molecular Biology and Evolution,* 12**,** 863-882.

Pérez-Losada, M., Porter, M.L., Tazi, L., & Crandall, K.A. (2007) New methods for inferring population dynamics from microbial sequences. Infection, Genetics, and Evolution, *7, 24-43.*

Perlman, S. J., Spicer, G. S., Shoemaker, D. D. & Jaenike, J. (2003) Associations between mycophagous *Drosophila* and their *Howardula* nematode parasites: a worldwide phylogenetic shuffle. *Molecular Ecology,* 12**,** 237-249.

Phillips, A., Janies, D. & Wheeler, W. (2000) Multiple sequence alignment in phylogenetic analysis. *Molecular Phylogenetics and Evolution,* 16**,** 317-330.

Piganeau, G., Gardner, M. & Eyre-Walker, A. (2004) A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution,* 21**,** 2319-2325.

Poinar, G. O., Jr. (1993) Origins and phylogenetic relationships of the entomophilic rhabditids, *Heterorhabditis* and *Steinernema*. *Fundamental and Applied Nematology,* 16**,** 333-338.

Posada, D. & Buckley, T. R. (2004) Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology,* 53**,** 793-808.

Posada, D. & Crandall, K. A. (1998) ModelTest: testing the model of DNA substitution. *Bioinformatics,* 14**,** 817-818.

Powers, T. O. (2004) Nematode molecular diagnostics: From bands to barcodes. *Annual Review of Phytopathology,* 42**,** 367-383.

Powers, T. O., Harris, T. S. & Hyman, B. C. (1993) Mitochondrial DNA sequence divergence among *Meloidogyne incognita*, *Romanomermis culicivorax*, *Ascaris suum*, and *Caenorhabditis elegans*. *Journal of Nematology,* 25**,** 564-572.

Powers, T. O., Todd, T. C., Burnell, A. M., Murray, P. C. B., Fleming, C. C., Szalanski, A. L., Adams, B. A. & Harris, T. S. (1997) The rDNA internal transcribed spacer region as a taxonomic marker for nematodes. *Journal of Nematology,* 29**,** 441-450.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics,* 155**,** 945-959.

Provine, W. B. (2001) *The origins of theoretical population genetics,* Chicago, University of Chicago Press.

Rambaut, A. & Drummond, A.J. (2003) Tracer v1.3 http://evolve.zoo.ox.ac.uk/software.html?id=tracer.

Raymond, M. & Rousset, F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity,* 86**,** 248-249.

Rolston, A. N., Boyle, S., Kakouli-Duarte, T., Griffin, C. T. & Downes, M. J. (2004) Distribution of, and intraspecific variation between, populations of entomopathogenic nematodes from Bull Island, Republic of Ireland. *Journal of Nematology,* 36**,** 344.

Ronquist, F. (1995) Reconstructing the history of host-parasite associations using generalised parsimony. *Cladistics,* 11**,** 73-89.

Ronquist, F. & Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-1574.

Rozas, J. & Rozas, R. (1995) DnaSP, DNA sequence polymorphism: An interactive program for estimating population genetics parameters from DNA sequence data. *Computer Applications in the Biosciences,* 11**,** 621-625.

Rozas, J. & Rozas, R. (1997) DnaSP version 2.0: A novel software package for extensive molecular population genetics analysis. *Computer Applications in the Biosciences,* 13**,** 307-311.

Rozas, J. & Rozas, R. (1999) DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics,* 15**,** 174-175.

Siddall, M. E. (1998) Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris Zone. *Cladistics,* 14**,** 209-220.

Siddall, M. E. (2001) Philosophy and phylogenetic inference: A comparison of likelihood and parsimony methods in the context of Karl Popper's writings on corroboration. *Cladistics,* 17**,** 395-399.

Skantar, A. M. & Carta, L. K. (2004) Molecular characterization and phylogenetic evaluation of the Hsp90 gene from selected nematodes. *Journal of Nematology,* 34, 466-480.

Slowinski, J. B. (1993) Component, Version 2.0. *Cladistics,* 9**,** 351-353.

Sorenson, M. (1999) TreeRot Version 2. Boston, MA Boston University.

Spiridonov, S. E., Reid, A. P., Podrucka, K., Subbotin, S. A. & 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.

Steel, M., Huson, D. & Lockhart, P. J. (2000) Invariable sites models and their use in phylogeny reconstruction. *Systematic Biology,* 49**,** 225-232.

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'eustachio, P., Fitch, D. H. A., Fulton, L. A., Fulton, R. E., Griffiths-Jones, S., Harris, T. W., Hillier, L. W., Kamath, R., Kuwabara, P. E., Mardis, E. R., Marra, M. A., Miner, T. L., Minx, P., Mullikin, J. C., Plumb, R. W., Rogers, J., Schein, J. E., Sohrmann, M., Spieth, J., Stajich, J. E., Wei, C., Willey, D., Wilson, R. K., Durbin, R. & Waterston, R. H. (2003) The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. *PLOS Biology,* 1**,** 166-192.

Stock, S. P., Campbell, J. F. & Nadler, S. A. (2001) Phylogeny of *Steinernema* Travassos, 1927 (Cephalobina : Steinernematidae) inferred from ribosomal DNA sequences and morphological characters. *Journal of Parasitology,* 87**,** 877-889.

Subbotin, S.A., Sturhan, D., Chizhov V.N., Vovlas, N., & Baldwin, J.G. (2006a) Phylogenetic analysis of Tylenchida Thorne, 1949 as inferred from D2 and D3 expansion fragment of the 28S rRNA gene sequences. *Nematology,* 8**,** 455-474.

Subbotin, S. A., Sturhan, D., Vovlas, N., Castillo, P., Tanyi Tambe, L., Moens, M. & Baldwin, J. G. (2006b) Application of secondary structure model of rRNA for phylogeny: D2-D3 expansion segments of the LSU gene of plant-parasitic nematodes from the family Hoplolaimidae Filipjev, 1934. *Molecular Phylogenetics and Evolution,* 43**,** 881-890.

Sullivan, J. & Joyce, P. (2005) Model selection in phylogenetics. *Annual Review of Ecology Evolution and Systematics,* 36**,** 445-466.

Sullivan, J. & Swofford, D. L. (2001) Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Systematic Biology,* 50**,** 723-729.

Swofford, D. L. (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods),* Sunderland, Massachusetts, Sinauer Associates.

Swofford, D. L., Waddell, P. J., Huelsenbeck, J. P., Foster, P. G., Lewis, P. O. & Rogers, J. S. (2001) Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology,* 50**,** 525-539.

Tang, S. (2006). Comparative mitochondrial genomics of the Nematoda: The family Mermithidae represents rapid Enoplean mitochondrial DNA evolution. Ph.D. Dissertation, University of California-Riverside, 252 pages

Tang, S. & Hyman, B. C. (2005) Rolling circle amplification of complete nematode mitochondrial genomes. *Journal of Nematology,* 37**,** 236-241.

Tang, S. & B. C. Hyman, 2007. Mitochondrial genome haplotype hypervariation within the isopod parasitic nematode *Thaumamermis cosgrovei*. *Genetics* 176,1139-1150.

Telford, M. J., Wise, M. J. & Gowri-Shankar, V. (2005) Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the bilateria. *Molecular Biology and Evolution,* 22**,** 1129-1136.

Templeton, A. R. (2006) *Population genetics and microevolutionary theory,* Hoboken, N.J., Wiley-Liss.

Terry, M. D. & Whiting, M. F. (2005) Comparison of two alignment techniques within a single complex data set: POY versus Clustal. *Cladistics,* 21**,** 272-281.

Thompson, E. A. (1990) Fisher, R.A. Contributions to Genetic Statistics. *Biometrics,* 46**,** 905-914.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research,* 25**,** 4876-4882.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) Clustal W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22**,** 4673-4680.

Wheeler, W. (1996) Optimization alignment: The end of multiple sequence alignment in phylogenetics? *Cladistics,* 12**,** 1-9.

Wheeler, W. (2001) Homology and the optimization of DNA sequence data. *Cladistics,* 17**,** S3-S11.

Wheeler, W. & Gladstein, D. (1994) MALIGN. 1.99 ed. New York, American Museum of Natural History.

Wheeler, W. C. (2003) Implied alignment: a synapomorphy-based multiple-sequence alignment method and its use in cladogram search. *Cladistics,* 19**,** 261-268.

Whipple, L. E., Lunt, D. H. & Hyman, B. C. (1998) Mitochondrial DNA length variation in *Meloidogyne incognita* isolates of established genetic relationships: utility for nematode population studies. *Fundamental and Applied Nematology,* 21**,** 265-271.

Wiley, E. O. (1988a) Parsimony analysis and vicariance biogeography. *Systematic Zoology,* 37**,** 271-290.

Wiley, E. O. (1988b) Vicariance biogeography. *Annual Review of Ecology and Systematics,* 19**,** 513-542.

Ye, W., Giblin-Davis, R.M., Davies, K.A., Purcell, M.F., Scheffer, S.J., Taylor, G.S., Center, T.D., Morris, K., & Thomas, W.K. (2007) Molecular phylogenetics and the evolution of host plant associations in the nematode genus *Fergusobia* (Tylenchida: Fergusobiinae). *Molecular Phylogenetics and Evolution,* 45**,** 123-141.

Manuscript published in *Symbiosis*

Chapter 2

Natural Selection on the *lux*A Gene of Bioluminescent Bacteria

Scott M. Peat and Byron J. Adams

Microbiology & Molecular Biology Department, and Evolutionary Ecology Laboratories,

Brigham Young University, Provo, UT, USA 84602-5253,

Tel: +1-801-422-8723, Fax: +1-801-422-0519, Email: speat@byu.net; bjadams@byu.edu

Abstract: Despite a growing literature of *Vibrio, Photobacterium, Shewanella,* and *Photorhabdus* biology, little is known of the function bioluminescence provides to these light-emitting bacteria. Proposed benefits of bioluminescence include evasion of predators or attraction of prey for symbiotic bacterial hosts through a distraction, a method of oxygen consumption to suffocate a host or reduce competition from obligate aerobes, a mechanism that stimulates DNA repair, or as a redox sink. We tested for the presence or absence of destabilizing selection on 31 physicochemical properties of the *lux*A gene of bacterial luciferase in relation to a phylogenetic hypothesis and the location within the protein structure, in an attempt to further understand the evolution of bacterial bioluminescence and its importance to symbiosis. We show that amino acid properties most influenced by destabilizing selection include the power to be at the C-terminal, chromatographic index, and isoelectric point. The location of destabilizing selection for isoelectric point within a phylogenetic context indicates that bacterial ecology has had an effect on the evolutionary history of *lux*A, while the presence of destabilizing selection for chromatographic index supports previous findings that bioluminescence in these species is

sensitive to environmental osmolarity.

Introduction

Bioluminescence, the production and emission of light by a living organism as a result of a chemical reaction, occurs in an array of organisms including fish, insects, jellyfish, and bacteria. Production of light by bacteria is unique in that luminous bacteria continuously produce light at a wavelength of 490 nm, while higher organisms (i.e. insects and jellyfish) display only intermittent flashes of light (Haygood, 1993). Many marine fish species are bioluminescent due to the presence of bioluminescent bacterial symbionts that inhabit the fishes light organ. Bioluminescent bacteria are the most abundant and widely distributed of all light-emitting organisms (Meighen, 1994), occupying a wide variety of ecological niches (fish light organs, mammalian gut, nematode gut) and habitats (marine, freshwater, terrestrial, and symbiotic within a host). Currently, only four genera of bacteria are known to naturally bioluminesce: *Vibrio, Photobacterium*, *Shewanella,* and *Photorhabdus*.

Most luminous *Vibrio cholerae* strains are found in aquatic environments (Colwell et al., 1981, Garay et al., 1985, Falcao et al., 1998) commonly associated with zooplankton (Colwell, 1996). *Vibrio fischeri* is known to form a symbiotic relationship with squid as well as being found in fish in shallow temperate waters (Madigan and Martinko, 2005) and in planktonic environments (Ruby and Nealson, 1978, Ruby and Lee, 1998). In contrast, *Vibrio harveyi*, best known for causing milky ocean, a phenomenon where the ocean glows white at night due to large *V. harveyi* populations, is primarily a free-living bacterium found in the water column of marine environments. Other bioluminescent bacterial genera inhabiting aquatic environments include *Photobacterium* spp., which can be found on the surface of fish, as a symbiont in the light organs of deep water marine fish (Madigan and Martinko, 2005), and in coastal and open-ocean sea water (Ast and Dunlap, 2005), while *Shewanella* is commonly found free living in

freshwater environments (Haygood, 1993). *Photorhabdus* spp., gut endosymbionts in juveniles of entomopathogenic nematodes from the genus *Heterorhabditis*, are the only terrestrial bacteria known to exhibit bioluminescence (Gerrard et al., 2003).

The general bioluminescent reaction is a complicated process requiring the cooperation of multiple genes. The enzyme luciferase interacts with $FMNH_2$ to form an EFH_2 complex, which subsequently reacts with O_2 to yield an oxygenated enzyme-flavin complex. This complex interacts with aldehyde (RCOH) to form a luciferase-FH_2 O_2 -RCOH complex (Stabb, 2005; Li and Tu, 2005). Decay of this complex goes to completion with the emission of blue-green light at 490 nm (Haygood, 1993; Valkova et al., 1999).

In bacteria, *lux* genes are responsible for the production of light (Kuwabara et al., 1965; Friedland and Hastings, 1967; Baldwin et al., 1989). Bioluminescent bacteria have at least five *lux* genes, each with similar functions across taxa. *lux*AB genes are the genes that code for luciferase, while *lux*CDE genes are the fatty acid reductase complex, and are responsible for synthesizing the fatty aldehyde substrate for the luminescence reaction (O'Kane and Prasher, 1992; Stabb, 2005). While similarity in function of both the *lux*AB and *lux*CDE genes exists across bacterial species, the organization of the *lux* operon varies in each bacterial species (Kasai et al., 2007; Meighen and Szittner, 1992; O'Kane and Prasher, 1992; Fig. 1).

A great deal is known as to how light is produced in bioluminescent bacteria, though the question of why these bacteria emit light remains unanswered. For bacteria that form a symbiotic relationship with fish, luminescence may provide a distraction that allows the host fish to elude predators or catch prey (Szpilewska et al., 2003), and as such is necessary to maintain a successful relationship. The natural ability of bioluminescent bacteria to reduce molecular oxygen through the oxidation of luciferin led to the proposal by McElroy and Seliger (1962) that

light production evolved as a mechanism of oxygen consumption. By consuming oxygen in the surrounding environment, bioluminescent bacteria can out-compete obligate aerobic bacteria (Timmins et al., 2001) as well as slow a host animal's ability to produce toxic oxygen radicals (Stabb, 2005). The production of light by certain bacteria has also been speculated to function as a redox sink, whereby light production acts as a mechanism to reduce excess NADH, which has built up due to growth conditions, to NAD^+ (Stabb, 2005). Stimulation of DNA repair is a more recent idea that has been proposed to explain the evolution of bioluminescence (Czyz et al., 2003). Under this scenario, bioluminescence, even when present at very low levels, activates a photoreactivation reaction, which could act to repair DNA (Czyz et al., 2000). Thus, in some environments the ability to produce even low levels of light could give a luminescent bacterium an advantage over a non-luminescent bacterium (Czyz et al, 2003).

While understanding why bacteria bioluminesce is important, it may be equally important to know if luminescence plays a role in the maintenance of symbiosis. In the relationship between *Photorhabdus* and *Heterorhabditis*, it appears that symbiosis does benefit from luminescence. *Photorhabdus* bacteria are known to exhibit two phases: primary phase, characteristic of bacteria with the ability to bioluminesce and produce antibiotics and extracellular enzymes, and secondary phase, which lacks all of the aforementioned characteristics. Phase I *Photorhabdus* variants can support nematode growth and colonize the intestinal tract of *Heterorhabditis* infective juveniles (IJ's) while phase II variants cannot. It has been shown that those traits which differ between the two phases (bioluminescence, production of antibiotics, etc.) represent factors that facilitate symbiosis, termed symbiosis factors (Joyce and Clark, 2003). Joyce and Clark (2003) go on to show that the presence of a *hex*A homologue in phase II *Photorhabdus* represses these symbiosis factors and that insertion into the *hex*A gene

of secondary phase *Photorhabdus* restores symbiosis factors, allowing said mutant to support nematode growth and development. This suggests that the lux pathway may be necessary in the maintenance of symbiosis in this system.

As mentioned earlier, *lux*A and *lux*B are the two genes responsible for the production of luciferase, the enzyme that drives the bioluminescence reaction. *lux*A codes for the alpha subunit of luciferase, which is primarily responsible for the kinetic properties of luciferase (Madvar, et al., 2005). While the high quantum yield bioluminescent reaction requires a heterodimer of both the alpha and beta subunits (*lux*A and *lux*B), the active center of bacterial luciferase is found on the alpha subunit (Noland et al., 1999). Furthermore, the position and presence of the alpha subunit of bacterial luciferase within the *lux* operon appears to be conserved across taxa (Fig. 1), making *lux*A a suitable target to investigate selection across bioluminescent bacterial species.

Traditionally, selection on a protein coding gene was calculated using the ratio of nonsynonymous (d_N) to synonymous (d_S) substitutions, though it has been shown that some of the assumptions made by this method are too conservative (Crandall et al., 1999; Woolley et al., 2003). Furthermore, while the d_N/d_S ratio may indicate the presence of selection on a gene, it does not specify how the selection affects the structure and/or function of the protein (Taylor et al., 2005). By evaluating the presence or absence of selection among particular physicochemical properties of amino acids in relation to a phylogenetic hypothesis and the location of selection within the protein structure, we can more accurately detect the presence of destabilizing selection in an attempt to further understand the evolution of bacterial bioluminescence and its importance to symbiosis. Thus, we tested for the presence or absence of destabilizing selection on 31 physicochemical properties of the *lux*A gene of bacterial luciferase. We then mapped these

properties on a phylogenetic tree to determine if selection on specific physicochemical properties could account for differences in the ecology as well as the function of bioluminescence in each of the sampled bacterial species.

Materials and Methods

*lux*A cds sequences were obtained from Genbank for seven bacterial species representing four genera, including *Vibrio fischeri* strain ES114 (NC_006841) a mutualistic symbiont from the bobtailed squid (Ruby et al., 2005), *Vibrio harveyi* strain NBRC 15364 (DQ436496), *Vibrio cholerae* strain TP (AY876056) from plankton (Purdy et al., 2005), *Photobacterium leiognathi* strain lleuc.1.1 (AY341070) from the light organ of a leiognathid fish (Ast and Dunlap, 2004), *Photobacterium phosphoreum* strain ATCC 11040 (AY341063) from the skin of a fish (Ast and Dunlap, 2005), *Photorhabdus luminescens* strain TT01 (NC_005126), and *Shewanella hanedai* strain NCIMB 2157 (AB261992). The longest open reading frame for each sequence was determined prior to alignment of the sequences using BioEdit (Hall, 1999). As *lux*A is a coding region, AlignmentHelper 1.2 (http://inbio.byu.edu/faculty/dam83/cdm) was utilized to convert nucleotide sequences into amino acids prior to alignment. Furthermore, AlignmentHelper allows for the fate of each amino acid to be tracked during the alignment process, allowing codon conformations to remain intact following conversion back to nucleotide data. Following multiple alignment of amino acid sequences in MUSCLE 3.3 (Edgar, 2004), sequences were re-input into AlignmentHelper for conversion of the amino acid sequences back into nucleotide data. Phylogenetic relationships of the seven species were inferred from previously published trees as well as a parsimony analysis, conducted in PAUP*4.0b10 (Swofford, 2000), of 16S rRNA for all seven species using 1000 random addition sequences and TBR branch swapping.

TreeSAAP v3.0 (Woolley et al., 2003) was utilized to measure selection based on changes in 31 physicochemical amino acid properties. Each property change was classified into one of eight categories based on the magnitude of change, where categories 1- 3 indicates a conservative change, with conservative changes representing stabilizing selection, and categories 6 - 8 signifies a radical change, with radical changes indicating destabilizing selection. TreeSAAP uses inferred evolutionary relationships as well as user provided sequence data to calculate an expected random distribution of possible amino acid changes for each category. Significant deviations are detected by comparing the expected distribution to the observed number of amino acid replacements in the data set given the phylogenetic relationships. A z-score is calculated for each category and significant selection is measured at an alpha of 0.001. Radical changes with a z-score of < 0.001 indicate destabilizing selection. TreeSAAP data outputs were mapped onto a linearized and flattened version of the 3-D structure of *lux*A, allowing for visualization of the exact parts of the 3-D structure (i.e. loop, stem, etc.) where selection is taking place, and the effects selection for a particular property has on protein function (Woolley et al., 2003, Taylor et al., 2005).

Results and Discussion

Of the 31 amino acid properties tested, 27 exhibited some degree of positive destabilizing selection, including alpha-helical tendency, average number of surrounding residues, beta-structure tendency, buriedness, chromatographic index, coil tendency, composition, equilibrium constant, helical contact area, hydropathy, isoelectric point, long-range nonbonded energy, mean r.m.s. fluctuation displacement, molecular weight, normalized consensus hydrophobicity, partial specific volume, polar requirement, polarity, power to be at the C-terminal, power to be at the

middle of the alpha helix, power to be at the N-terminal, refractive index, short-range and medium-range nonbonded energy, solvent accessible reduction ratio, surrounding hydrophobicity, total nonbonded energy, and turn tendency. Properties that were most influenced by destabilizing selection included power to be at the C-terminal, isoelectric point, and chromatographic index. A few codons showed selection for multiple properties including codons 15, 28, 29, 65, and 145, indicating that certain properties may be correlated.

In the present study, selection for isoelectric point (pI), the pH at which a molecule carries no net electrical charge, occurred most often on the branch separating *Vibrio fischeri*, a symbiotic bacterium living within a light organ, from *Vibrio harveyi* and *Vibrio cholerae,* two aquatic/planktonic bacteria (Fig. 2). From these results one might suggest that difference in environmental pH may be the primary factor that is driving selection for isoelectric point in these bacteria, though this is probably not the case. The products of *lux* operon expression operate within the cytoplasm of the bacterial cell, and most bacteria can maintain an intracellular pH within a range of values (though the range varies for acidophiles, neutrophiles, and alkiphiles, and the intracellular pH values can be considerably different than the pH of the surrounding environment [Booth, 1985; Dilworth and Glenn, 1999]). A typical neutrophile will usually maintain a pH between 7.6 and 7.8 (Booth, 1985; Dilworth and Glenn, 1999), though studies have shown that very few proteins have an isoelectric point close to 7.4 (Kiraga et al., 2007). This can be explained by the fact that proteins are most insoluble, least reactive and unstable in pH close to their isoelectric point (Kiraga et al., 2007). Thus, the maintenance of a fairly homeostatic pH indicates that selection for isoelectric point is not driven by environmental pH. Instead, the presence of significant destabilizing selection on the branch separating a symbiotic bacterium (*V. fischeri*) from two free-living/planktonic bacteria, along with previous data from

Kiraga et al. (2007), indicate that selection for isoelectric point is probably driven in part by the ecology of the bacteria.

Photorhabdus bacteria are known to exhibit two phases: primary phase, characteristic of bacteria that are found in insect cadavers where osmolarity and bacterial biomass is high, and secondary phase, characteristic of *Photorhabdus* found in the intestines of infective dauerlarvae where osmolarity and biomass are low. Presence of high osmolarity and rich nutrients, as in the insect cadaver, appears to stabilize the phase I bioluminescent variants of *Photorhabdus* (Krasomil-Osterfeld, 1997). Variation in bioluminescent intensity has also been shown in *Vibrio fischeri* when the bacteria were subjected to high and low osmolarity concentrations, though the limiting factor causing the disparity in light output was revealed to be the aldehyde substrate (Stabb, 2004). The present study reveals the presence of significant destabilizing selection for chromatographic index on the evolutionary lineages leading to *Photorhabdus* and *P. phosphoreum* as well as the branch separating *Vibrio fischeri* from *Vibrio harveyi* and *Vibrio cholerae* (Fig. 2). Chromatographic index is defined as the hydropathy of a residue based on interactions of solute, solvent, and hydrophobic absorbent (Prabhakaran, 1990). The typical osmolarity of sea water is 1,000 mosM (Stabb et al., 2004), while the osmolarity in cephalopods is typically greater than sea water (Robertson, 1965; Stabb et al., 2004). Thus, bacterial cells within cephalopod light organs are probably subjected to higher salinities than bacteria that are free living in the ocean. We believe that the increased solute concentration is the reason destabilizing selection for chromatographic index was detected on the branch separating *V. fischeri* (a symbiont of squid) from *V. harveyi* (free living in marine environments) and *V. cholerae* (planktonic). Inversely, teleost fish maintain blood osmolarities that are less than the osmolarity of sea water (Fange et al., 1976; Stabb et al., 2004). Thus, in the *Photobacterium*

clade we see the presence of destabilizing selection for chromatographic index on the branch that separates the *P. leiognathi* lineage from the *P. phosphoreum* lineage. We attribute the change in chromatographic index to natural selection in response to the difference in osmolarity. Consequently, our data suggests that osmolarity has had an effect on the evolutionary history of the *lux*A gene of luminescent bacteria.

Power to be at the C-terminal is loosely defined as the propensity of the C-terminus of the alpha helix to interact with other residues (Prabhakaran and Ponnuswamy, 1979). While the property 'power to be at the C-terminal' is currently not well understood, selection for this property in the current study is associated with certain features of the secondary structure of the alpha subunit of luciferase. The active site of luciferase is located in a pocket near the C-terminal end of the alpha subunit (Li and Tu, 2005). Adjacent to the opening of the active site lies a 29-residue mobile loop, not present in the beta subunit, from α258 to α286 (Fig. 3). This loop is believed to be important to the gating of the active site and essential to luciferase light-emitting activity (Li and Tu, 2005). Mapping statistically significant destabilizing selection for power to be at the C-terminal onto a linearized version of the 3-D structure of luciferase (Fig. 3), reveals the presence of statistically significant destabilizing selection occurring in the α258 to α286 region. While a better understanding of the property 'power to be at the C-terminal' is needed to elucidate the role that selection on this property plays in the regulation of light emmitance, we reason that the multiple occurrences of significant selection within the α258 to α286 region of luciferase provides further evidence that this mobile loop region may be important to the luciferase light-emitting activity. Additionally, detection of multiple instances of destabilizing selection associated with key features (three turn helix, four turn helix, and hydrogen bonded turns) of the luciferase secondary structure signify that these regions may be

interacting with the active site of luciferace and as such may also be critical in the emission of light, though further analyses of these residues are needed to confirm this hypothesis. Further study into selection on other properties such as hydrophobicity, bulkiness, and alpha helical tendencies will reveal information on the importance of the mobile loop region of the alpha subunit of luciferase on the production of light in bacteria.

The unique ability of certain bacteria to bioluminesce compels many observers to generate scenarios for the origin and maintenance of light production in these bacteria. Each time a new explanation/hypothesis is proposed, it is assumed that bioluminescence confers some type of fitness benefit to bioluminescent bacteria, and often precludes the idea that bioluminescence might not have a direct function. In their critique of the adaptationist program, Gould and Lewontin (1979) note that evolutionary biologists tend to focus exclusively on immediate adaptations while ignoring phylogenetic legacies and constraints. As with the exemplar spandrels of St. Mark's Cathedral in Venice, bioluminescence, particularly as it exists in symbiotic relationships, provides a design "so elaborate, harmonious and purposeful that we are tempted to view it as the starting point of any analysis, as the cause in some sense to the surrounding architecture (Gould and Lewontin, 1979)." Instead, it may well be that bioluminescence in some of these species is analogous to an architectural constraint, a necessary secondary effect which originated from some other purpose or function.

Using the preceding two scenarios (bioluminescence confers some benefit to its possessor; bioluminescence as a byproduct), one can effectively evaluate numerous scenarios surrounding the evolutionary origin and maintenance of bioluminescence. If we evaluate the hypothesis that bioluminescence is used as an attractant or deterrent, we see that this hypothesis seems logical for those symbiotic bacteria that inhabit light organs of fish, but fails to account for

luminescence in *Photorhabdus*. *Photorhabdus*, the bacterial symbiont of the nematode *Heterorhabditis* is typically confined to the gut of its host and the hemocoel of larval insects, with both hosts inhabiting soil environments. As the phase of *Photorhabdus* that glows is typically only found in insect cadavers, there is no intuitive benefit of glowing to attract prey or distract a predator, as all necessary resources for survival and reproduction are present in the insect cadaver.

Since *Photorhabdus* probably does not utilize bioluminescence as an attractant, an adaptationist might propose that the alternative idea of a redox sink provides a more logical explanation as to why *Photorhabdus* bioluminesces. Furthermore, one might reference the reduction in the *lux* operon of *Photorhabdus* (Fig. 1) for support of an alternative hypothesis. A reduction in the genes utilized by *Photorhabdus* to produce light could indicate that lack of need for one function (i.e. attraction or repulsion) has caused a reduction of genes in the operon through evolutionary time, and a transition of these genes to a novel function (i.e. redox sink). To test the idea of operon reduction leading to an alternative function, more taxa are needed in the present analysis beyond those bacteria that possess *lux* genes, yet do not bioluminesce. Furthermore, from a biomass perspective, the redox sink hypothesis gains credence, as greater biomass of bacteria and nutrients exist in the insect cadaver than is found in the nematode gut. Subsequent reduction of excess NADH (which has accumulated due to high biomass) to NAD+ would cause an excess production of light, leading to the increased bioluminescence that is generally observed with *Photorhabdus* when inside the insect cadaver.

Finally, it should also be considered that bioluminescence in *Photorhabdus* has no primary function, and that the bacteria bioluminesce only because they possess the genes that allow for luminescence. Two points lend support to this idea. First is the generally accepted

view that *Photorhabdus*, like the luminescent bacteria *Shewanella*, acquired its *lux* operon through horizontal gene transfer from *Vibrio* (Kasai et al, 2007). Thus, if this is a non-functional trait that has been recently acquired via horizontal gene transfer rather than a trait that has been passed vertically over millions of years, it may not have had sufficient time, evolutionarily speaking, to have been completely lost, and as such light is still emitted without providing any real advantage to the organism. If this is the case, the intensity of light emitted might be expected to decrease over time in *Photorhabdus*, when compared to its luminescent counterparts. Experiments by Meighen (1999) lend support to this idea, finding that *Photorhabdus* emits a light intensity that is considerably lower than *Photobacterium leiognathi, Photobacterium phosphoreum, Vibrio fischeri,* and *Vibrio harveyi.*

Second is the fact that the closely related taxon *Xenorhabdus,* a bacterium which inhabits an almost identical niche (the gut of the nematode *Steinernema,* and the insect hemocoel), does not possess a *lux* operon. So if *Xenorhabdus*, which encounters similar environmental conditions within its nematode and insect hosts thrive without bioluminescing, then why would *Photorhabdus* need to bioluminesce? In this case a compensatory mechanism in *Xenorhabdus* (i.e. a redox sink analogue) could support the idea that bioluminescence as a redox sink is beneficial to *Photorhabdus.* Alternatively, absence of an analog in *Xenorhabdus* would lend support to the non-functional hypothesis. While the lack of function scenario assumes that negative selection pressure has been absent throughout the evolution of this bioluminescent bacterium, further tests evaluating the energetic costs of light production on *Photorhabdus* fitness need to be conducted to resolve this notion.

Acknowledgements

We gratefully acknowledge David McClellan for assistance with TreeSAAP and Nicole Lewis-Rogers for helpful discussion on TreeSAAP and data visualization. This work was supported in part by The United States Department of Agriculture (CSREES NRI 2005-35302-16089), and a Mentored Environment Grant from Brigham Young University.

References

Ast, J.C., and Dunlap, P.V. 2004. Phylogenetic analysis of the *lux* operon distinguishes two evolutionarily distinct clades of *Photobacterium leiognathi*. Archives of Microbiology 181: 352–361.

Ast, J.C., and Dunlap, P.V. 2005. Phylogenetic resolution and habitat specificity of members of the *Photobacterium phosphoreum* species group. Environmental Microbiology 7: 1641–1654.

Baldwin, T.O., Devine, J.H., Heckel, R.C., Lin, J.W., and Shadel, G.S. 1989. The complete nucleotide sequence of the *lux* regulon of *Vibrio fischeri* and the *lux*ABN region of *Photobacterium leiognathi* and the mechanism of control of bacterial bioluminescence. Journal of Bioluminescence and Chemiluminescence 4: 326-341.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. Nucleic Acids Research 28: 235-242.

Booth, I.R. 1985. Regulation of cytoplasmic pH in bacteria. Microbiological Reviews 49: 359–378.

Colwell, R.R. 1996. Global climate and infectious disease: the cholera paradigm. Science 274: 2025–2031.

Colwell, R. R., Seidler, R.J., Kaper, J., Joseph, S.W., Garges, S., Lockman, H., Maneval, D., Bradford, H., Roberts, N., Remmers, E., Huq, I., and Huq, A. 1981. Occurrence of *Vibrio cholerae* serotype O1 in Maryland and Louisiana estuaries. Applied and Environironmental Microbiology 41: 555–558.

Crandall, K.A., Kelsey, C.R., Imamichi, H., and Salzman, N.P. 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. Molecular Biology and Evolution 16: 372-382.

Czyz, A., Plata, K., and Wegrzyn, G. 2003. Stimulation of DNA repair as an evolutionary drive for bacterial luminescence. Luminescence 18:140-144.

Czyz, A., Plata, K., and Wegrzyn, G. 2000. Vibrio harveyi bioluminescence plays a role in stimulation of DNA repair. Microbiology 146: 283-288.

Dilworth, M.J., and Glenn, A.R. 1999. Problems of adverse pH and bacterial strategies to combat it. In: Bacterial responses to pH. Chadwick, D.J., and Cardew, G., eds. Wiley, Chichester, pp 4-14.

Edgar, R.C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

Falcao, D.P., Lustri, W.R., and Bauab, T.M. 1998. Incidence of non-O1 *Vibrio cholerae* and *Aeromonas* spp. in fresh water in Araraquara, Brazil. Current Microbiology 37: 28–31.

Fange, R., Lidman, U., and Larsson, A. 1976. Comparative studies of inorganic substances in the blood of fishes from the Scagerac Sea. *Journal of Fish Biology* 8: 441-448.

Friedland, J., and Hastings, J.W. 1967. Nonidentical subunits of bacterial luciferase: Their isolation and recombination to form active enzyme. Proceedings of the National Academy of Sciences of the United States of America 58: 2336-2342.

Garay, E., Arnau, A., and Amaro, C. 1985. Incidence of *Vibrio cholerae* and related vibrios in a coastal lagoon and seawater influenced by lake discharges along an annual cycle. Applied and Environmental Microbiology 50: 426–430.

Gerrard, J.G., McNevin, S., Alfredson, D., Forgan-Smith, R., and Fraser, N. 2003. *Photorhabdus* species: bioluminescent bacteria as emerging human pathogens? Emerging Infectious Diseases 9: 251-254.

Gould, S.J., and Lewontin, R.C. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proceedings of the Royal Society of London B 205: 581-598.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41: 95-98.

Haygood, M.G. 1993. Light organ symbioses in fishes. Critical Reviews in Microbiology 19: 191-216.

Joyce, S.A., and Clarke, D.J. 2003. A *hexA* homologue from *Photorhabdus* regulates pathogenicity, symbiosis and phenotypic variation. Molecular Microbiology 47: 1445-1457.

Kabasch, W., and Sander, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577-637.

Kasai, S., Okada, K., Hoshino, A., Iida, T., and Takeshi, H. 2007. Lateral transfer of the *lux* gene cluster. Journal of Biochemistry 141: 231-237.

Kiraga, J., Mackiewicz, P., Mackiewicz, D., Kowalczuk, M., Biecek, P., Polak, N., Smolarczyk, K., Dudek, M.R., and Cebrat, S. 2007. The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. BMC Genomics 8: 163.

Krasomil-Osterfeld, K.C. 1997. Phase II variants of *Photorhabdus luminescens* are

induced by growth in low-osmolarity medium. Symbiosis 22: 155-165.

Kuwabara, S., Cormier, M.J., Dure, L.S., Kreiss, L.S., and Pfuderer, P. 1965. Crystalline bacterial luciferase from *Photobacterium fischeri*. Proceedings of the National Academy of Sciences of the United States of America 53: 822-828.

Li, C.H., and Tu, S.C. 2005. Active site hydrophobicity is critical to the bioluminescence activity of *Vibrio harveyi* luciferase. Biochemistry 44: 12970-12977.

Madvar, A.R., Hosseinkhani, S., Khajeh, K., Ranjbar, B., and Asoodeh, A. 2005. Implication of a critical residue (GLU175) in structure and function of bacterial luciferase. FEBS Letters 579: 4707-4706.

Madigan, M., and Martinko, J. 2005. Brock Biology of Microorganisms, 11th ed., Prentice Hall.

McElroy W.D., and Seliger, H.H. 1962. Origin and evolution of bioluminescence. In: Horizons in biochemistry. Kasha, M., and Pullman, B., eds. Academic Press, New York, pp 91-101.

Meighen, E.A. 1999. Autoinduction of light emission in different species of bioluminescent bacteria. Luminescence 14: 3-9.

Meighen, E.A. 1994. Genetics of bacterial bioluminescence. Annual Review of Genetics 28: 117-139.

Meighen, E.A., and Szittner, R.B. 1992. Multiple repetitive elements and organization of the *lux* operons of luminescent terrestrial bacteria. Journal of Bacteriology 174: 5371-5381.

Noland, B.W., Dangott, L.J., and Baldwin, T.O. 1999. Folding, stability, and physical properties of the α subunit of bacterial luciferase. Biochemistry 38: 16136-16145.

O'Kane, D.J., and Prasher, D.C. 1992. Evolutionary origins of bacterial bioluminescence. Molecular Microbiology 6: 443-449.

Prabhakaran, M. 1990. The distribution of physical, chemical, and conformational properties in signal and nascent peptides. Biochemical Journal 269: 691-696.

Prabhakaran, M., and Ponnuswamy, P.K. 1979. The spatial distribution of physical, chemical, energetic and conformational properties of amino acid residues in globular proteins. Journal of Theoretical Biology 80: 485-504.

Purdy, A., Rohwer, F., Edwards, R., Azam, F., and Bartlett, D.H. 2005. A glimpse into the expanded genome content of *Vibrio cholerae* through identification of genes present in environmental strains. Journal of Bacteriology 187: 2992-3001.

Robertson, J.D. 1965. Studies on the chemical composition of muscle tissue. III. The mantle muscle of cephalopod mollusks. *Journal of Experimental Biology* **42**: 153-175.

Ruby, E.G., and Lee, K. 1998. The *Vibrio fischeri-Euprymna scolopes* light organ association: current ecological paradigms. Applied and Environmental Microbiology 64: 805-812.

Ruby, E.G., and Nealson, K.H. 1978. Seasonal changes in the species composition of luminous bacteria in nearshore seawater. Limnology and Oceanography 23: 530–533.

Ruby, E.G., Urbanowski, M., Campbell, J., Dunn, A., Faini, M., Gunsalus, R., Lostroh, P., Lupp, C., McCann, J., Millikan, D., Schaefer, A., Stabb, E., Stevens, A., Visick, K., Whistler, C., and Greenberg, E.P. 2005. Complete genome sequence of *Vibrio fischeri*: a symbiotic bacterium with pathogenic congeners. Proceedings of the National Academy of Sciences of the United States of America 102: 3004– 3009.

Stabb, E.V. 2005. Shedding light on the bioluminescence "paradox". ASM News 71: 223-229.

Stabb, E.V., Butler, M.S., and Adin, D.M. 2004. Correlation between osmolarity and luminescence of symbiotic *Vibrio fischeri* Strain ES114. Journal of Bacteriology 186: 2906-2908.

Swofford, D.L., 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Szpilewska, H., Czyz, A., and Wegrzyn, G. 2003. Experimental evidence for the physiological role of bacterial luciferase in the protection of cells against oxidative stress. Current Microbiology 47: 379-382.

Taylor, S.D., Dittmar de la Cruz, K., Porter, M.L., and Whiting M.F. 2005. Characterization of the long-wavelength opsin from mecoptera and siphonaptera: does a flea see? Molecular Biology and Evolution 22: 1165-1174.

Timmins, G.S., Jackson, S.K., and Swartz, H.M. 2001. The evolution of bioluminescent oxygen consumption as an ancient oxygen detoxification mechanism. Journal of Molecular Evolution 52: 321-332.

Valkova, N., Szittner, R., and Meighen, E.A. 1999. Control of luminescence decay and flavin binding by the *lux*A carboxyl-terminal regions in chimeric bacterial luciferases. Biochemistry 38: 13820-13828.

Woolley, S., Johnson, J., Smith, M.J., Crandall, K.A., and McClellan, D.A. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19: 671-672.

Figure Legends:

Figure 1: Organization of the *lux* operons of *Vibrio harveyi*, *Vibrio fischeri*, *Vibrio cholerae*, *Photobacterium leiognathi*, *Photobacterium phosphoreum*, *Shewanella hanedai* and *Photorhabdus luminescens*. The arrangement of *lux* A, B, C, and D are conserved across all bioluminescent taxa. *V. fischeri* and *S. hanedai* have two regulatory genes, *lux*I and *lux*R upstream of the *lux*C gene. A significant reduction in the number of *lux* genes can be seen in *Photorhabdus luminescens* in relation to the ancestral states, indicating selection for a decrease in the *lux* operon over time as well as possible selection for an alternative function of bioluminescence in *Photorhabdus* when compared to other bioluminescent bacteria. Figure adapted from Kasai et al., 2007, Meighen and Szittner, 1992, and O'Kane and Prasher, 1992.

Figure 2: Phylogenetic tree of bioluminescent bacterial relationships with destabilizing selection for isoelectric point and chromatographic index mapped onto corresponding branches. Black and white boxes indicate the number of times destabilizing selection was detected on each lineage for designated properties in the *lux*A gene, with each box representing a single occurrence of statistically significant destabilizing selection. The tree was generated using 16S rRNA sequences.

Figure 3: Location of statistically significant destabilizing selection (depicted as black bars) for the properties chromatographic index, isoelectric point, and power to be at the C-terminal, in relation to the amino acid sequence and secondary structure of bacterial luciferase. Destabilizing selection was detected at two distinct codon positions within residues 258 to 286, the mobile loop region (indicated by a gap in the secondary structure diagram) adjacent to the proposed active site of *lux*A. Multiple instances of destabilizing selection for power to be at the C-terminal were also found to be associated with key features of the luciferase secondary structure. Amino acid, secondary structure sequence (Kabasch and Sander, 1983), and secondary structure diagram (labeled DSSP) were obtained from the RCSB Protein Databank (Berman et al., 2000).

Figure 1

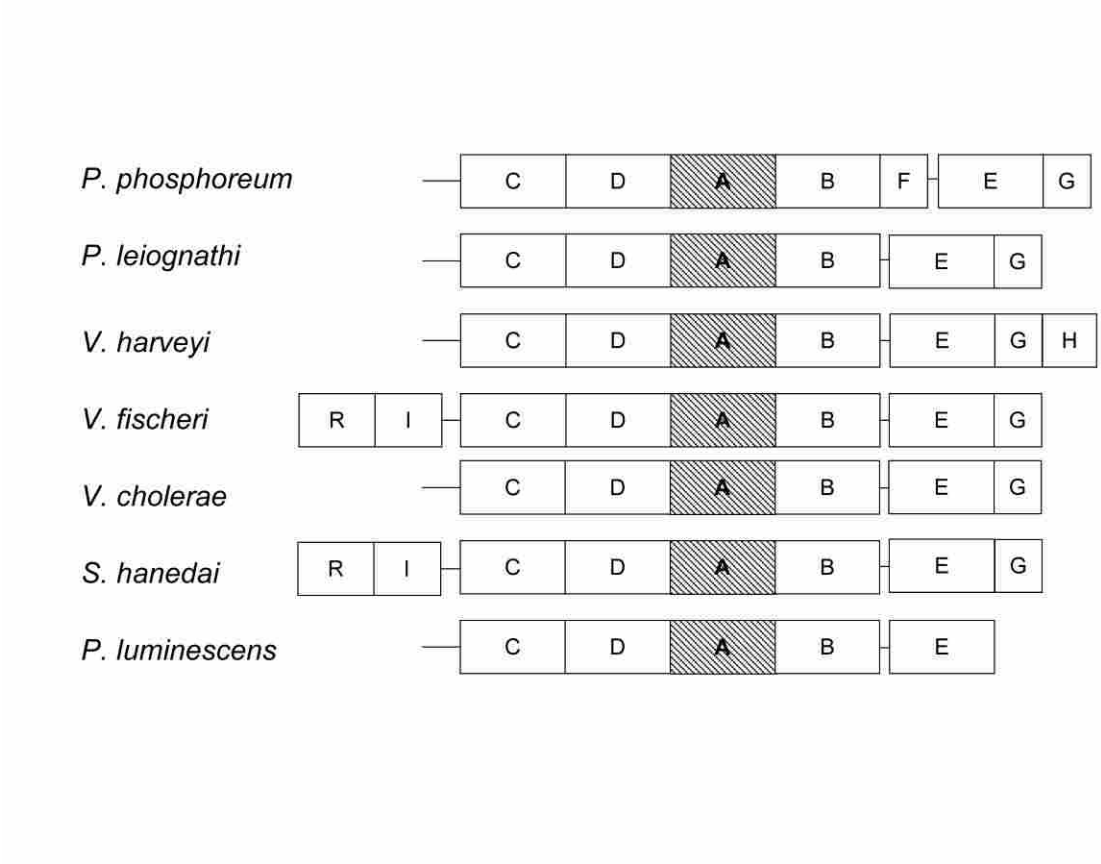

P. phosphoreum
C D A B F E G
P. leiognathi
C D A B E G
V. harveyi
C D A B E G H
V. fischeri
R I C D A B E G
V. cholerae
C D A B E G
S. hanedai
R I C D A B E G
P. luminescens
C D A B E

Figure 2

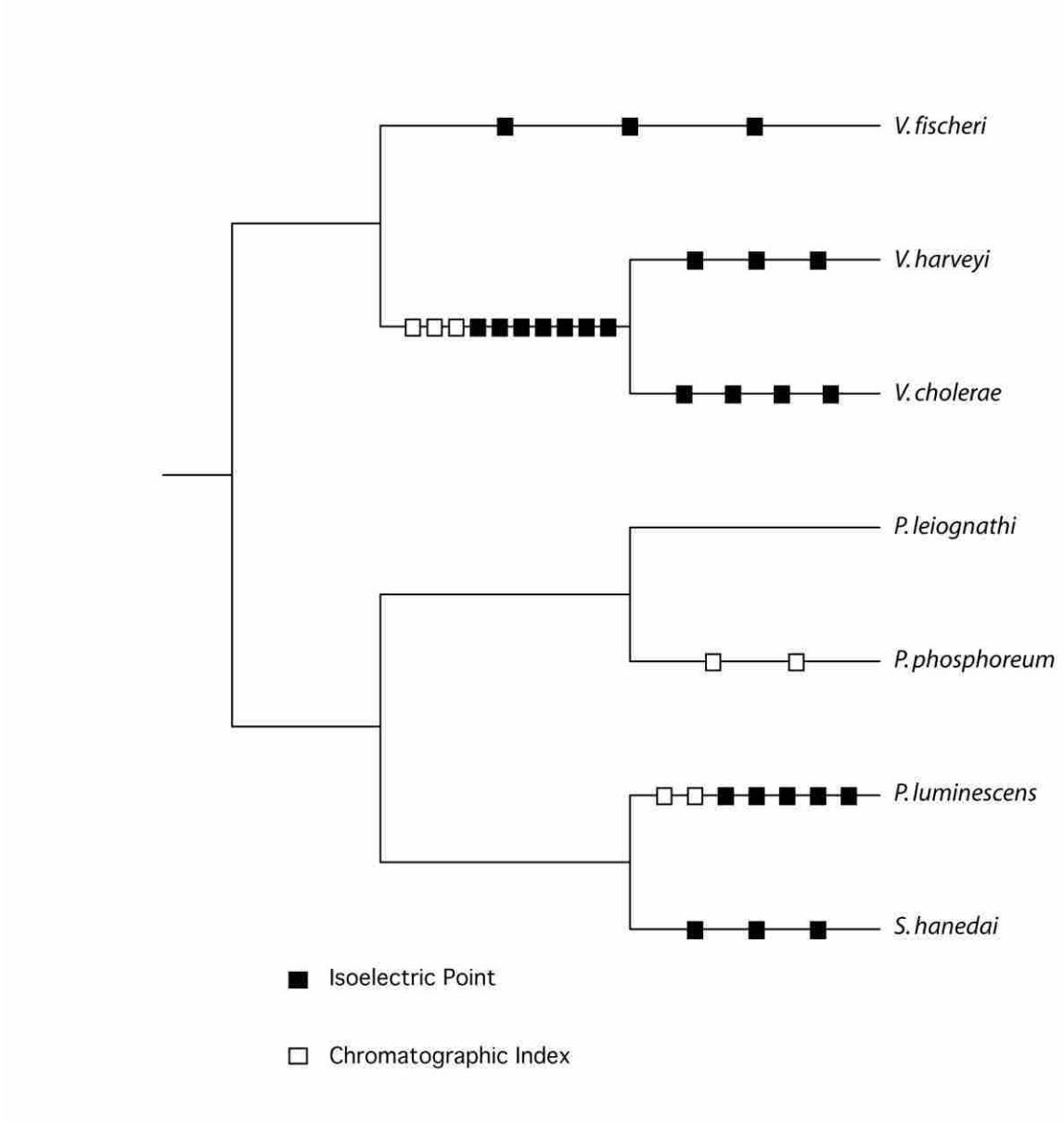

V. fischeri
V. harveyi
V. cholerae
P. leiognathi
P. phosphoreum
P. luminescens
S. hanedai
Isoelectric Point
Chromatographic Index

Figure 3

Chromatographic Index
Isoelectric Point
Power to be at C-Terminal
DSSP
Secondary Structure
Amino Acids
Chromatographic Index
Isoelectric Point
Power to be at C-Terminal
Key to Secondary Structure Codes
G = 3-turn helix
H = 4-turn helix
I = 5-turn helix
T = hydrogen bonded turn
E = beta sheet
B = residue in isolated beta bridge
S = bend

Manuscript submitted to *Molecular Phylogenetics and Evolution*

Chapter 3

A Robust Phylogenetic Framework for the Bacterial Genus *Photorhabdus* and its use in Studying the Evolution and Maintenance of Bioluminescence: A Case for 16S, *gyrB*, and *glnA*

Scott M. Peat[a*], Richard H. ffrench-Constant[b], Nick R. Waterfield[c], Judit Marokházi[d], Andras Fodor[e], and Byron J. Adams[a]

[a]Department of Biology, Brigham Young University, 401 WIDB, Provo, UT, USA 84602-5253
[b]School of Biosciences, University of Exeter in Cornwall, Penryn, TR10 9EZ, UK
[c]Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK
[d]Department of Biochemistry, Faculty of Natural Sciences, Eötvös University, H-1117, Budapest, Hungary
[e]Institute of Plant Protection, Georgikon Faculty, Pannonia University, Keszthely, Hungary

E-mail addresses: speat@byu.net (S.M. Peat), byron_adams@byu.edu (B.J. Adams), rf222@exeter.ac.uk (R.H. ffrench-Constant), bssnw@bath.ac.uk (N.R. Waterfield), juttka@gmail.com (J. Marokházi), fodorandras@yahoo.com (A. Fodor)

*Author for correspondence:
Scott Peat
E-mail: speat@byu.net
Tel: 1-801-422-8723
Fax: 1-801-422-0090

Abstract

Photorhabdus spp., the only known bioluminescent terrestrial bacteria, are well known for their symbiotic association with heterorhabditid nematodes. This association, along with their ability to kill insects, has aroused interest in the evolutionary relationships within this bacterial group. Currently, three species are recognized within the genus *Photorhabdus; P. temperata* and *P. luminescens,* which are endosymbionts of *Heterorhabditis* spp., and *P. asymbiotica,* which has been isolated from human wounds and has recently been shown to also have a heterorhabditid nematode vector. To examine phylogenetic relationships among these taxa, we utilize total evidence Bayesian, likelihood, and parsimony based analyses of three genetic loci (16S rRNA, *gyrB,* and *glnA*) to construct a robust evolutionary hypothesis for the genus *Photorhabdus.* Here we use this phylogeny to evaluate existing specific and sub-specific taxonomic statements within the genus, identify previously undescribed *Photorhabdus* strains, test the utility of 16S rRNA, *gyrB*, and *glnA* in resolving various levels of relationships within the genus, and, finally, to investigate the evolution of bioluminescence. The genes examined produced the most robust phylogenetic hypothesis to date for the genus *Photorhabdus,* as indicated by strong bootstrap and posterior probability values at previously unresolved or poorly resolved nodes. We show that *glnA* is particularly useful in resolving specific and intra-specific relationships poorly resolved in other studies. We conclude that *P. asymbiotica* is the sister group to *P. luminescens* and that the new strains HIT and JUN should be given a new group designation within *P. asymbiotica*. Furthermore, we reveal a pattern of decline in bioluminescent intensity through the evolution of *Photorhabdus*, suggesting that this may be a trait acquired and maintained under previous ecological (aquatic) selection pressures that is now gradually being lost in its terrestrial environment.

Keywords: *Photorhabdus*, Bioluminescence, Phylogenetic Systematics, *Heterorhabditis*, 16S rRNA, *gyrB*, *glnA*

Introduction

Photorhabdus spp., the only terrestrial bacteria known to exhibit bioluminescence (Gerrard et al., 2003), are motile, gram negative bacteria which are gut endosymbionts in juveniles of entomopathogenic nematodes from the genus *Heterorhabditis*. The close symbiotic relationship between *Photorhabdus* and *Heterorhabditis* has gained much attention due to their ability to work together to kill their insect host. Upon locating a suitable insect host, *Heterorhabditis* penetrates through natural openings (mouth, anus, spiracles) (Boemare, 2002), or directly into the hemocoel of the larval insect via the integument (Akhurst and Dunphy, 1993; Forst et al., 1997; Poinar, 1990), subsequently releasing bacteria into the hemolymph (Forst et al., 1997). Once in the hemolymph, *Photorhabdus* begins multiplying, simultaneously releasing toxins virulent enough to kill the insect within 24 hours (Ciche and Ensign, 2003; Forst et al., 1997). All *Photorhabdus* strains are considered highly entomopathogenic, with an LD_{50} of <100 cells per insect (Boemare, 2002). Following death of the insect and consumption of all available nutrients, *Photorhabdus* and *Heterorhabditis* re-assimilate, leaving the dead insect in search of another insect host (Forst and Nealson, 1996).

Initially classified as *Xenorhabdus luminescens*, the genus *Photorhabdus* would later be proposed by Boemare et al. (1993) based on the examination of phenotypic characters and DNA relatedness studies. Specific and subspecific taxonomic designations within the genus *Photorhabdus* are based on phenotypic data including morphological, biochemical and physiological characters (Akhurst et al., 1996; Fischer-Le Saux et al., 1999), DNA-DNA

hybridization (Akhurst et al., 1996; Farmer et al., 1989), sequencing of a portion of 16S rRNA (Liu et al., 1997), sequencing of the complete 16S region of rRNA (Fischer-Le Saux et al., 1999), sequencing of the *gyrB* gene (Akhurst et al., 2004), and a multilocus sequence typing analysis of *recA, gyrB, dnaN, gltX* (Tailliez et al., 2009). Akhurst et al. (1996) concluded that phenotypic data alone could only separate two groups of *Photorhabdus*, the symbionts and the clinical strains. Clinical infections with *Photorhabdus* have been reported in a number of humans within the United States (Farmer et al., 1989) and Australia (Gerrard et al., 2003; Peel et al., 1999).

Liu et al. (1997) developed a phylogeny of *Photorhabdus* and another closely related bacterial endosymbiont of nematodes, *Xenorhabdus.* The study used 13 *Photorhabdus* isolates, most of which had no species designation. Based upon maximum likelihood analysis of a portion of 16S rRNA, Liu et al. showed four well supported major clades within the one recognized clade, supporting the possibility that more than one species of *Photorhabdus* exists. Through a polyphasic approach utilizing 16S rRNA phylogenetic inference, phenotypic characterization, and DNA-DNA hybridization data, Fisher-Le Saux et al. (1999) proposed the existence of three separate species of *Photorhabdus: P. luminescens, P. temperata, and P. asymbiotica*. Furthermore, the study went on to propose the existence of three subspecies within *P. luminescens*. A second polyphasic approach utilizing phenotypic characterization, DNA-DNA hybridization, and two molecular markers, *gyrB* and 16S rRNA proposed the separation of *P. asymbiotica* into two subspecies (Akhurst et al., 2004). Hazir et al. (2004) used riboprint analysis, metabolic properties, and a distance analysis of 16S rRNA, to propose two new subspecies (*P. luminescens* ssp. *thracensis* and *P. luminescens* ssp. *kayaii*). More recently, the sub-species *P. temperata cinerea* has been proposed based on *gyrB* data (Toth and Lakatos,

2008) and Tailliez et al. (2009) used *recA, gyrB, dnaN, gltX* to propose four new subspecies (*P. luminescens caribbeanensis, P. luminescens hainanensis, P. temperata khanii,* and *P. temperata tasmaniensis*) and the renaming of another (*P. temperata thracensis*).

Historically, 16S rRNA has been the marker of choice when classifying/naming *Photorhabdus* species and subspecies. It has been suggested that *Photorhabdus* species may be subjected to a higher evolutionary rate than that of its sister taxon *Xenorhabdus* (Rainey et al., 1995) based on analysis of 16S rRNA data. Conversely, Tailliez et al. (2009) suggest that *Xenorhabdus* evolves at a faster rate than *Photorhabdus* based on the analysis of dN/dS ratios for five different genes (*rplB, recA, gyrB, dnaN, and gltX*). While 16S rRNA has been shown as useful in identifying many bacteria to the generic level (Fukushima et al., 2002; Wang et al., 1994), evidence of potential lateral gene transfer of 16S rRNA exists in numerous bacterial genera including *Photorhabdus* (Tailliez et al., 2009), giving 16S rRNA the potential to confound bacterial species relationships rather than resolving them, especially when using 16S alone. As such, other genes, particularly protein coding genes, may provide a clearer representation of the species relationships within the genus *Photorhabdus*. One such marker, *gyrB*, encodes the subunit B protein of DNA gyrase. DNA gyrase functions in the regulation of supercoiling of double stranded DNA. This enzyme is ubiquitous among all bacterial species (Yamamoto and Harayama, 1995). Previous studies have indicated that *gyrB* might prove to be more useful in identifying bacteria to the species level due to its higher rates of molecular evolution (Fukushima et al., 2002; Yamamoto and Harayama, 1995, 1998). Akhurst et al. (2004) utilized the *gyrB* gene to propose the split of the species *P. asymbiotica* into two sub-species, as well as confirm the presence of three species within the genus. Another gene that has shown promise in resolving relationships within the genus *Photorhabdus* is *glnA*, a gene that codes for

the glutamine synthetase enzyme (Tullius et al., 2003). Gerrard et al. (2006) conducted a neighbor joining analysis using concatenated *gyrB* and *glnA* datasets to successfully confirm the identity of a *Photorhabdus asymbiotica* strain isolated from a nematode, though support for the most basal node and many terminal nodes in their phylogeny were extremely low. Additionally, a similar tree was utilized by Waterfield et al. (2008) to illustrate the locations of *Photorhabdus* strains of interest in their study.

Analyses of single genes, while providing a good depiction of the gene tree, often do not accurately reflect the species tree for a given organism (for further discussion see (Maddison, 1997; Pamilo and Nei, 1988)). This is likely the case for the genus *Photorhabdus*. When looking at the relationships recovered using single gene phylogenies for *Photorhabdus*, it is evident that the phylogenies are discordant (Fig. 1). With regards to species relationships within the genus *Photorhabdus* the 16S gene tree shows *P. asymbiotica* and *P. luminescens* as polyphyletic with *P. temperata* as sister group to a *P. asymbiotica asymbiotica* + *P. luminescens luminescens* clade, while the *gyrB* and *glnA* gene trees show *P. asymbiotica* and *P. luminescens* as separate monophyletic groups and *P. luminescens* as sister group to *P. asymbiotica*. Furthermore, the *gyrB* gene tree shows *P. temperata* (which includes *P. temperata* and *P. temperata temperata*) as polyphyletic, while *glnA* shows *P. temperata* as monophyletic.

Horizontal gene transfer has been hypothesized as a factor causing incongruence in the 16S gene tree of *Photorhabdus* in relation to other *Photorhabdus* gene trees (Talliez et al., 2009). Additionally, disconcerted evolution among multiple copies of the 16S gene in *Photorhabdus* may account for some of the observed incidences of incongruence. An examination of the genomes of *Photorhabdus luminescens* strain TTO1 (GenBank accession # BX470251.1) and *Photorhabdus asymbiotica* strain ATCC94399 (GenBank accession # FM162591.1) show the

presence of multiple copies of the 16S gene. As such, when analyzing 16S *Photorhabdus* data we may be looking at a mixture of paralogous genes rather than a single group of orthologous genes, though further work examining the presence of 16S paralogs in other *Photorhabdus* species and subspecies needs to be undertaken. While 16S alone is inadequate for evaluating specific and subspecific designations, when combined with other loci 16S may retain some phylogenetic utility.

To date, most single gene and all total evidence phylogenetic analyses of molecular data for the genus *Photorhabdus* have utilized distance based methods of phylogenetic reconstruction (i.e. neighbor joining). Neighbor joining and other distance based methods are useful in that they can build a phylogenetic tree very rapidly, though their use of overall similarity (phenetics) to build phylogenetic trees wastes potentially informative character data (Farris, 1981) while lacking the ability to distinguish between homology and homoplasy (Siebert, 1992). Furthermore, observed distances between sequences do not accurately reflect the evolutionary distances between them, and as such sequences may appear more closely related than they actually are (Holder and Lewis, 2003). To thoroughly investigate the phylogenetic relationships within the genus *Photorhabdus,* we advocate an approach whereby a combined simultaneous analysis of multiple molecular datasets is conducted utilizing more rigorous methods of phylogenetic reconstruction including parsimony, Bayesian, and likelihood analyses. Using these non-phenetic based methods to conduct simultaneous analyses of multiple molecular datasets should provide more power in resolving issues of species delineation within the genus *Photorhabdus* by allowing phylogenetic signal to emphasize itself over phylogenetic noise (De Queiroz, 1993). Additionally, when conducting model based analyses of concatenated datasets, we also advocate the use of mixed models, as a single model is usually inadequate to account for

the differing histories of multiple genes (Bull et al., 1993; Huelsenbeck et al., 1996). The aforementioned phylogenetic methods have the ability to provide a more thorough analysis of the data and as such should produce a more robust phylogenetic hypothesis than has been produced in previous studies, which can then be used to test fundamental hypotheses regarding the evolutionary history of *Photorhabdus*. With this in mind, the aims of this study were: (1) to construct a more robust phylogenetic hypothesis by exploring combined Bayesian, likelihood, and parsimony analyses of multiple molecular datasets; (2) to evaluate specific and sub-specific taxonomic statements within the genus *Photorhabdus* by testing the monophyly of previously proposed groups against a more rigorous compilation of molecular data and tree reconstruction methods; (3) to evaluate the utility of 16S rRNA, *gyrB,* and *glnA* in resolving relationships within the genus *Photorhabdus*; and (4) to trace the evolution of bioluminescent intensity through the genus *Photorhabdus,* to determine if an increasing or decreasing trend of bioluminescent activity exists across the evolution of basal to the more derived clades.

Materials and Methods

DNA Extraction

Genomic DNA was purified using the DNeasy tissue kit (QIAgen, Valencia, CA) as per manufacturer's protocol. DNA was eluted from the column into 20µl of TE buffer and stored at -20°C.

PCR

Polymerase chain reaction amplifications of portions of the *gyrB* and *glnA* genes were carried out using a standard PCR reaction mixture that included 5X Go Taq buffer, 1.25mM of

Mgcl2, 0.25mM dNTPs, 1mM of each primer and 1µl of Go Taq polymerase 100 units (Promega, Madison, WI). All amplifications were performed using a Peltier thermal cycler (PTC) with an initial denaturation at 95˚C for two minutes, 30 cycles of 95˚C for 45 seconds, 50˚C for 45 seconds, and 72˚C for one minute, and a final extension at 72˚C for 10 minutes. PCR products were purified using Montage PCR centrifugal filter devices (Millipore, Billerica, MA) in order to remove the salts, primers and unincorporated dNTPs. Each 100µl PCR reaction was mixed with 300 µl TE buffer, applied on the DNA capture columns and centrifuged at 1000x for 15 min. The DNA fragments were eluted with 20µl TE and stored at -20°C.

Sequencing and Sequence Editing

Cycle sequencing was performed using BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA). DNA sequences were analysed and assembled using the SeqMan program of the DNAstar Lasergene software. Sequence data generated for *gyrB* (accession numbers GU731081 - GU731130) and *glnA* (GU731131 - GU731180) have been deposited in GenBank.

Taxon Sampling

Taxa used in this study include 55 representative strains of the three characterized species of *Photorhabdus*, four subspecies of *P. luminescens*, and two (possibly three) subspecies of *P. asymbiotica*, three subspecies of *P. temperata*, and two uncharacterized *Photorhabdus* spp. for a total of 57 ingroup taxa. Three outgroup taxa used in the analyses include *Xenorhabdus beddingii*, the sister group to *Photorhabdus* (Francino et al., 2006; Koppenhofer, 2007), as well as two more distantly related bacteria within the family Enterobacteriaceae, *Salmonella enterica*

and *Yersinia pestis*. Table 1 shows all of the taxa and strains used in the current analyses as well as the Genbank accession numbers for all available data.

Alignment

Three molecular markers, 16S rRNA, *gyrB*, and *glnA* were used in the present study. The 16S rRNA dataset (1502 characters) was aligned using Muscle (Edgar, 2004), under the default parameters. Additionally, the software Gblocks (Castresana, 2000; Talavera and Castresana, 2007) was utilized on the 16S dataset to eliminate poorly aligned positions that may be a result of paralogous sequences. Alignments for the *gyrB* (573 characters) and *glnA* (493 characters) genes were conducted using their amino acid sequences. Longest open reading frames for each protein coding dataset was determined using BioEdit (Hall, 1999). AlignmentHelper 1.2 (http://inbio.byu.edu/faculty/dam83/cdm) was used to convert nucleotide sequences into amino acids prior to alignment using MUSCLE 3.3 (Edgar, 2004). AlignmentHelper tracks the fate of each amino acid during the alignment process, allowing codon conformations to remain intact following conversion back to nucleotide data. Following alignment, sequences were back-translated by AlignmentHelper to convert the amino acid sequences back into nucleotide sequences.

Phylogenetic Analyses

MacClade 4.05 (Maddison and Maddison, 2002) was utilized to remove redundant sequences from the single gene alignments and to concatenate the three datasets into a single combined molecular dataset of 2,635bp and 49 taxa. Parsimony analyses were conducted in TNT (Goloboff et al., 2008) under the new technology search with drift, ratchet, and pruning and

1000 random addition sequences. A strict consensus tree was assembled for each analysis using the multiple most parsimonious (MP) trees recovered from the heuristic searches. The program TreeRot v3 (Sorenson and Franzosa, 2007) was used to calculate partitioned Bremer support values which were mapped onto the strict consensus tree to assess the contribution each dataset made to the overall topology. To evaluate the effect of missing data on the robustness of the resulting phylogeny, one additional combined analysis was conducted using all three datasets (*gyrB*, 16S, and *glnA*) with any taxa missing one or more of the genes removed (2,635 characters and 37 taxa). Analyses were also conducted for both the 49 taxa and 37 taxa datasets including only the gyrB and glnA datasets.

For model based phylogenetic analyses, best fit models of evolution were calculated using ModelTest 3.1 (Posada and Crandall, 1998). Mixed model Bayesian analyses were conducted in MrBayes 3.08 (Ronquist and Huelsenbeck, 2003). For the 49 taxa analysis, a mixed models Bayesian analysis was employed using partitioning by gene under the TrN+I+G model for the 16S partition, the TIM+I+G model for the *glnA* partition and the SYM+G model for the *gyrB* partition. For the 37 taxa analysis, a mixed models Bayesian analysis was employed using partitioning by gene under the HKY+I+G model for the 16S partition, the SYM+G model for the *glnA* partition and the TrN+G model for the *gyrB* partition. Two runs were conducted for each dataset using 20,000,000 generations sampled every 1000 generations. Stationarity was estimated using Tracer v1.4 (Rambaut and Drummond, 2007), with the combined analysis with missing data (49 taxa) having a burn-in value of 50,000 and the combined analysis with no missing data (37 taxa) having a burn-in of 30,000. Bremer support values were calculated by constructing a 50% majority rule consensus tree of the remaining trees (19,950 for the combined with missing data and 19,970 for the combined with no missing data) using PAUP*4.0b10

(Swofford, 2002). Log likelihood values for each run were compared to ensure that each Bayesian run converged on similar log likelihood mean values for each of the two independent runs for each gene.

Maximum likelihood analyses were conducted on the single gene and combined molecular datasets in RAxML (Stamatakis, 2006) using the GTRGAMMA model with partitioning by gene. Bootstrap values for the likelihood tree were calculated for both the 37 taxon and 49 taxon datasets in RAxML using 1000 bootstrap replicates.

Results

Alignment

Following alignment of the 16S dataset, a region at the 3' end of the alignment was discovered with large insertion/deletion events for a number of taxa followed by the presence of five to six nucleotides and then another large insertion/deletion event. This region was aligned manually and analyses were run with and without this region. While inclusion of this region of the 16S did not cause any difference in terms of the relationships recovered, it did cause a decrease in the robustness of the recovered phylogeny. As such, all phylogenies depicted in the present paper were constructed with this 130 bp region removed. Additionally, analysis of the 16S dataset in Gblocks resulted in the removal of an additional 36 nucleotide positions. Phylogenetic analyses were run with and without this additional 36 nucleotide deletion.

Phylogenetic Analyses

Partitioned homogeneity test results indicated a lack of incongruence ($p < 0.05$) between all pairwise comparisons. Bayesian (Fig. 2) and likelihood (Fig. 3) simultaneous analyses of all

three datasets including missing data suggests that *P. luminescens* is the sister group to *P. asymbiotica*. This sister relationship is supported in all analyses with a likelihood bootstrap support of 94 and a Bayesian posterior probability value of 93. Within this clade, the two *P. asymbiotica* subspecies form a monophyletic group with two unidentified *Photorhabdus* spp. (HIT and JUN) with a likelihood bootstrap value of 100 and a posterior probability value of 100. *Photorhabdus luminescens akhurstii , P. luminescens luminescens, P. luminescens kayaii,* and *P. luminescens laumondii* each form separate well supported monophyletic groupings within the monophyletic *P. luminescens* clade. *Photorhabdus temperata thracensis* forms a monophyletic group with two undescribed *Photorhabdus* spp. (MOL and AZ29) strains. *Photorhabdus temperata* is monophyletic, with *P. temperata thracensis* as part of a group that forms a sister group to *P. temperata temperata* in both the Bayesian (PP = 100) and likelihood (Bootstrap <50) analyses, while *P. temperata thracensis, P. temperata*, and *P. temperata temperata* form an unresolved polytomy in the parsimony analysis. *Photorhabdus temperata* and *Photorhabdus temperata temperata* each form separate strongly supported monophyletic groups, though as mentioned earlier, both likelihood and Bayesian analyses indicate that *P. temperata temperata* is more closely related to *P. temperata thracensis* than it is to any *P. temperata* strains.

Bayesian (Fig. 4), likelihood (Fig. 5), and parsimony (Fig. 6) simultaneous analysis of the three molecular datasets, with taxa missing data for one or more genes removed, again suggests that *P. luminescens* is the sister group to *P. asymbiotica*, though support for this grouping is higher (likelihood bootstrap = 100; Posterior Probability = 100) than in the 49 taxa combined analysis. This increase in support is likely a result of a decreased amount of missing data in the 37 taxon tree relative to the 49 taxon tree. *Photorhabdus asymbiotica* again forms a monophyletic group with two unidentified *Photorhabdus* spp. (strains HIT and JUN), while the

three *Photorhabdus luminescens* subspecies included in this analysis form separate well supported monophyletic groupings within the larger *P. luminescens* clade. As in the 49 taxon analyses, both *P. temperata* and *P. temperata temperata* strains form separate, strongly supported monophyletic groups. Bayesian analysis (Fig. 4) shows *P. temperata temperata* and the group containing *P. temperata thracensis* and two unidentified *Photorhabdus* spp. (MOL and AZ29) form a monophyletic group (PP = 95). Conversely, likelihood analysis (Fig. 5) shows *P. temperata* + the group containing *P. temperata thracensis* and two unidentified *Photorhabdus* spp. (MOL and AZ29) as monophyletic (bootstrap = 64), while the parsimony analysis (Fig. 6) shows *P. temperata* + *P. temperata temperata* forming a monophyletic group (bootstrap = 51). Very little difference was observed in regards to tree topology and support values when comparing the 37 taxon dataset to the 37 taxon dataset with 36 nucleotide positions removed from the 16S gene following Gblock analysis.

Bayesian analysis of the 37 taxon dataset with the 16S gene removed resulted in the *P. asymbiotica asymbiotic* clade becoming unresolved and forming a polytomy with *P. asymbiotica* strains HIT and JUN. Additionally, *P. luminescens akhurstii* was the sister taxon to *P. luminescens luminescens* while in the analysis of all three genes, *P. luminescens akhurstii* was sister to *P. luminescens laumondii*. In the 49 taxa Bayesian analysis with the 16S gene removed, a loss of resolution was observed at the branch supporting the monophyly of *P. luminescens akhustii* with *P. luminescens kayaii* and *P. luminescens laumondii*, and a significant decrease in support values were observed on the branch supporting the grouping of *P. asymbiotica* with *P. luminescens* (from 93 to 87), the branch supporting the grouping of *P. luminescens kayaii* and *P. luminescens laumondii* (from 100 to 65), and the branch supporting the *P. temperata* clade (from 100 to 51).

Partitioned Bremer Support (PBS)

Partitioned Bremer support (PBS) values for the 37 taxon simultaneous analysis tree are shown in Figure 6. In total, PBS values for the *gyrB*, 16S, and *glnA* datasets were 88.54, 173.20, and 109.26 respectively. Negative values for the 16S gene were observed at the branch supporting the grouping of *P. temperata* strains MEG1 and OH as well as the branch which grouped *P. temperata* + *P. temperata temperata* + the group that contains the strains KOH, AZ29, and MOL (all of which formed a monophyletic group with *P. temperata thracensis* in the 49 taxon analysis). Negative PBS values were observed for *gyrB* at the branch supporting the monophyly of strains KOH + AZ29 + MOL as well as the branch supporting the monophyly of *P. luminescens laumondii* + *P. luminescens akhurstii*. Negative PBS values for *glnA* were found on all branches within the *P. luminescens akhurstii* clade, the branch supporting the monophyly of undescribed strains AZ29 + MOL, the branch supporting the monophyly of *P. temperata* strains Heliothidis + NC19 + Wx12 + Wx11, the branch supporting the monophyly of *P. temperata* strains Heliothidis + NC19 + Wx12 + Wx11 + Wx10 + MEG1 + OH1 + Wx9, and the branch supporting the monophyly of *P. temperata* + *P. temperata temperata*.

GyraseB provided more support than the other two genes at the branch supporting the monophyly of all *P. temperata* strains, the branch supporting the monophyly of *P. asymbiotica asymbiotica* + uncharacterized *Photorhabdus* spp. HIT and JUN, the branch supporting the monophyly of *P. luminescens akhurstii* strains EG2 + IND + LN2 +W14, and the branch supporting the monophyly of uncharacterized strains AZ29 + MOL. Partitioned Bremer support for the *glnA* gene was higher than *gyrB* and 16S at the branch supporting the monophyly of *P. luminescens laumondii*, the branch supporting the monophyly of *P. luminescens akhurstii*, the

branch supporting the monophyly of *P. luminescens laumondii* + *P. luminescens akhurstii*, + *P. luminescens luminescens*, the branch supporting the monophyly of *P. luminescens laumondii* + *P. luminescens akhurstii* + *P. luminescens luminescens*, the branch supporting the monophyly of *P. asymbiotica* + *P. luminescens*, and the branch supporting the monophyly of *P. temperata temperata* + *P. temperata* + the clade that contains the strains KOH, AZ29, and MOL (all of which form a monophyletic group with *P. temperata thracensis* in the 49 taxon analysis). Finally, PBS for 16S was higher than *gyrB* and *glnA* at the branch supporting the monophyly of *P. asymbiotica asymbiotica*, the branch supporting the monophyly of the strains HIT + JUN, the branch supporting the monophyly of *P. temperata temperata*, the branch supporting the grouping of KOH + AZ29 + MOL, the branch supporting the monophyly of *P. luminescens luminescens*, and numerous branches supporting relationships within the species *P. temperata* and the sub-species *P. temperata temperata, P. luminescens laumondii,* and *P. luminescens akhurstii*.

Discussion

Photorhabdus taxonomy and previously unidentified species

In the 49 taxon analyses *P. asymbiotica, P. luminescens*, and *P. temperata* all formed consistent, well supported, monophyletic groups in the Bayesian and likelihood analyses. The consistent placement of *P. temperata thracensis* with a group of unidentified *Photorhabdus* spp. (AZ29, MOL, and KOH), that form a sister group to the *P. temperata temperata* (Bayesian), or the *P. temperata* (likelihood) clade indicates that this strain was indeed inaccurately designated as *P. luminescens thracensis* by Hazir et al. (2004) and, as previously suggested by Tailliez et al.(2009), should be renamed as *P. temperata thracensis*. Additionally, based on the current

analysis, the previously unidentified strains AZ29, MOL, and KOH should also be included in the new subspecies *P. temperata thracensis*.

The previously unclassified European strains HIT and JUN consistently formed a well supported group within the *P. asymbiotica* clade, though separate from *P. asymbiotica australis* (Australian strains) and *P. asymbiotica asymbiotica* (United States strains). The uncorrected sequence divergence across all three genes between *P. asymbiotica asymbiotica* and *P. asymbiotica australis* is 4.2% (*gyrB* = 6.6%; 16S = 2.9%; *glnA* = 5.1%) while the sequence divergence across all three genes between HIT/JUN and *P. asymbiotica asymbiotica* is 4.8% (*gyrB* = 5.7%; 16S = 3.7%; *glnA* = 6.8%), and the sequence divergence between HIT/JUN and *P. asymbiotica australis* is 4.7% (*gyrB* = 7.1%; 16S = 3.0%; *glnA* = 6.5%). As such, it logically follows that *P. asymbiotica* strains HIT and JUN form a novel group within *P. asymbiotica* and accordingly should likely be formally renamed in the future. Unidentified strain MX4A consistently formed a monophyletic group with *P. luminescens luminescens* in all of the 49 taxon analyses and as such should be included in this taxonomic grouping. Finally, unidentified strains X4 and Wx13, while consistently falling into the *P. temperata, P. temperata temperata, P. temperata thracensis* clade, did not consistently form a monophyletic group with either of these groups and as such further analyses need to be done to confirm the taxonomic affinity of these *Photorhabdus* strains.

Usefulness of 16S, gyrB, and glnA in the present study

Evaluated separately, 16S, *gyrB*, and *glnA* each tell a different story in regards to the relationships within the genus *Photorhabdus* (Fig. 1). This has proven to be problematic in previous studies that have utilized only 16S rRNA to name new species (Hazir et al., 2004).

Furthermore the use of only a single gene has led to poorly supported/resolved phylogenetic hypotheses for the genus *Photorhabdus,* which are inadequate for use in naming new species as well as conducting evolutionary based studies on this genus.

From partitioned Bremer support values, one can see that each of these three genes, when combined and analyzed with more rigorous phylogenetic methods can be useful in resolving a more robust phylogeny of *Photorhabdus*. More specifically, 16S can be useful at resolving sub-specific and inter-sub-specific relationships, as illustrated by the increased amount of Bremer support, relative to *glnA* and *gyrB,* given to the numerous sub-specific and inter-sub-specific relationships that were noted in the partitioned Bremer support results section above. Similarly, PBS values show that *gyrB* is useful in resolving some specific and subspecific relationships. Conversely, PBS values discussed in the PBS results section above indicate that *glnA* is extremely useful in resolving many of the specific and inter-specific relationships that, in previous studies (Tailliez et al., 2009; Toth and Lakatos, 2008), had been poorly resolved/supported. Additionally, while the utility of the 16S gene in resolving relationships within the bacterial genus *Photorhabdus* has been questioned (Koppenhofer, 2007; Tailliez et al., 2009), we have shown that in combination with *gyrB* and *glnA*, 16S can be useful in supporting terminal clades without confounding intra-specific and intra-sub-specific relationships. While the multiple copies of the 16S that have been observed in the genomes of multiple *Photorhabdus* taxa are disconcerting, the utilization of Gblocks to select the most conserved regions from an alignment of 16S data may aid in reducing, though likely not eliminating, potential paralogy problems with this dataset. Gblocks analysis in the current study removed few nucleotide positions (36) from the 16S alignment and little difference was observed in the resulting phylogenetic hypotheses. Additionally, the removal of the 16S from the combined dataset

resulted in loss of resolution and decrease in support at multiple branches in both the 37 taxon and 49 taxon trees, illustrating that the information content of the 16S may not be totally useless or unreliable.

While most of the relationships within the genus *Photorhabdus* were consistent and strongly supported across analyses (37 taxa and 49 taxa) and methods of tree reconstruction (parsimony, likelihood, and Bayesian), the relationship between *P. temperata*, *P. temperata temperata*, and the *P. temperata thracensis* (strains AZ29, MOL, FR32, and KOH) clade varied. The 37 taxon likelihood analysis (Fig. 5), shows *P. temperata* and *P. temperata thracensis* forming a poorly supported monophyletic group, the 37 taxon Bayesian analysis (Fig. 4) shows *P. temperata temperata* and *P. temperata thracensis* forming a monophyletic group, and the 37 taxon parsimony analysis (Fig. 6) shows *P. temperata* and *P. temperata temperata* forming a poorly supported monophyletic group. The 49 taxon Bayesian (Fig. 2) and likelihood (Fig. 3) analyses both show *P. temperata thracensis* and *P. temperata temperata* forming a monophyletic group, though the parsimony analysis failed to resolve these three clades. Much of this discrepancy is likely due to disagreement between the 16S and *glnA* genes, as the PBS analysis of the 16S supports the monophyly of *P. temperata temperata* and *P. temperata* while *glnA* does not. The mixed models Bayesian analysis for the *P. temperata temperata*/*P. temperata thracensis* clade is well supported, though further analyses with more genes needs to be conducted to definitively determine the relationships between these three clades.

Effect of missing data

While including missing data in phylogenetic analyses can lead to the decreased resolution and decreased phylogenetic accuracy (Huelsenbeck, 1991; Wiens, 2003, 2006), it should be

noted that including missing data in the present analyses did not cause incongruence. Removing taxa with missing data provided a more robust solution (Figs. 4 and 5), yet left intact the sister relationships recovered in the larger analysis (Figs. 2 and 3). So while *P. temperata thracensis, P. luminescens kayaii, and P. asymbiotica australis* were removed from the latter analysis, the taxa that they routinely formed clades with in the 49 taxon analyses appear in the same place in the 37 taxon tree. As such, it appears passable to leave missing data in phylogenetic analyses of *Photorhabdus* when using a combined analysis of 16S, *gyrB*, and *glnA* data.

Simultaneous analysis of 16S rRNA, *gyrB,* and *glnA* strongly support (likelihood bootstraps of 94 and 100 and Bayesian posterior probabilities of 93 and 100 for the 49 and 37 taxon analyses, respectively) the monophyly of *Photorhabdus asymbiotica* and *P. luminescens*. This is in contrast to the finding of Tailliez et al. (2009), whose four gene (*recA, gyrB, dnaN, gltX*) analysis supported the monophyly of *P. temperata* and *P. luminescens*, though support for this clade was quite low (NJ bootstrap = 63). As such, our analyses provide the strongest support for this clade to date. Regarding relationships within the *P. luminescens* clade, *P. luminescens laumondii* is the sister taxon to *P. luminescens kayaii* (49 taxon Bayesian posterior probability = 100) and *P. luminescens akhurstii* is more closely related to *P. luminescens laumondii* and *P. luminescens kayaii* than it is to *P. luminescens luminescens* (49 taxon Bayesian posterior probability = 99). Again, this is in contrast to the findings of Tailliez et al. (2009) who found that *P. luminescens akhurstii* is more closely related to *P. luminescens luminescens* than it is to *P. luminescens laumondii* or *P. luminescens kayaii*, though with much lower support (NJ bootstrap = 68). The present three gene analyses provide the most robust estimate of the evolutionary relationships within the genus *Photorhabdus* to date, and as such can be used to

facilitate studies on the evolution of a variety of traits that are present in *Photorhabdus* spp. We provide an example below.

Evolution of Bioluminescent Intensity

One of the most unique characteristics of *Photorhabdus* is its bioluminescent capabilities. *Photorhabdus* spp. are the only terrestrial natural bioluminescent bacteria, though other aquatic bioluminescent bacteria do exist (i.e. *Vibrio harveyi*, *Vibrio fischeri*, *Vibrio cholerae*, *Photobacterium leiognathi*, *Photobacterium phosphoreum*, *Shewanella hanedai*). In *Photorhabdus* and other bioluminescent bacteria, *lux* genes are responsible for the production of light (Baldwin et al., 1989; Friedland and Hastings, 1967; Kuwabara et al., 1965), and these *lux* genes are organized into an operon that varies in the organization of genes from one bacterium to the next (Kasai et al., 2007; Meighen and Szittner, 1992; O'Kane and Prasher, 1992). Much speculation exists as to why *Photorhabdus* biolumineses. Some hypotheses that have been proposed as to the functional significance of light production in *Photorhabdus* include a distraction mechanism, a molecular oxygen sink, an attractant, and a signal that synchronizes symbiosis (Waterfield et al., 2009). Alternatively, as suggested by Peat and Adams (2008), production of light in *Photorhabdus* may not have any real function and instead represents a trait present in an aquatic ancestor that is now being lost upon colonization of a terrestrial environment via the nematode vector or is a remnant of a horizontal gene transfer event that has not had sufficient evolutionary time to disappear. To examine these claims a bit more closely, we used the phylogenetic hypothesis generated in the present paper along with bioluminescent intensity data from stationary phase *Photorhabdus* from Hyrsl et al. (2004), to investigate the evolution of bioluminescent intensity across the genus *Photorhabdus*. Accordingly, we mapped

bioluminescence (mV) measurements (from Hyrsl et al., 2004) of five type strains of *Photorhabdus* onto the *Photorhabdus* phylogeny to explore the evolution of bioluminescent intensity. Mapping bioluminescence onto the *Photorhabdus* phylogeny (Fig. 7) shows that the basal *P. temperata* clade exhibits the highest bioluminescent intensity (1698 mV), while the intermediate *P. asymbiotica* clade exhibits the second highest intensity (406 mV), and the terminal *P. luminescens* clade exhibits the lowest intensities (22 – 188 mV). Because these luminescence measurements came from luminescence readings of stationary phase *Photorhabdus* (the phase with the greatest luminescence output), we believe that the luminescence data accurately reflects luminescence output of primary *Photorhabdus* phase variants, and that the discrepancy between luminescence values between species and subspecies is biologically relevant and not a function of a mixed analysis of primary and secondary phase variants. As such, our analysis indicates that bioluminescence in *Photorhabdus* was high early in the evolution of *Photorhabdus*, followed by a gradual decline to the more recently derived *P. luminescens* clade.

From these data, it remains plausible that *Photorhabdus* is indeed gradually evolving decreased bioluminescent intensity through subsequent speciation events. As such, Peat and Adams' (2008) idea that bioluminescence is just a non-functional evolutionary remnant that has not had sufficient time to disappear may be plausible for some of the more derived *Photorhabdus* clades, though the fact remains that some basal *Photorhabdus* spp. (i.e. *P. temperata*) have retained relatively high bioluminescent intensities. As such, if one subscribes to the prokaryotic mantra "use it or lose it" (Savageau, 1974), one would conclude that there is selection for the maintenance of bioluminescence in *Photorhabdus*, though an explanation for *Photorhabdus* spp. bioluminescence remains somewhat unsatisfying. Many of the functional hypotheses (i.e.

synchronization of symbiosis, redox sink, attractant, etc.) seem plausible and make logical sense until one examines the similar *Xenorhabdus*/*Steinernema* system. *Xenorhabdus* is a bacterial endosymbiont of the entomopathogenic nematode *Steinernema*, and similar to the *Photorhabdus/Heterorhabditis* system, both work together to kill larval insects, though *Xenorhabdus* does not bioluminesce. So why do bacteria that share an almost identical niche and are sister taxa, not have bioluminescence in common? It is possible that a compensatory mechanism exists in *Xenorhabdus* that carries out the same function that the production of light accomplishes in *Photorhabdus* (Peat and Adams, 2008), though that mechanism has yet to be discovered. Additionally, it has been suggested that the reduced operon of *P. luminescens* relative to other bioluminescent bacteria (i.e. *Vibrio fischeri, Photobacterium phosphoreum,* etc.), may indicate that the *lux* operon in *Photorhabdus* is being utilized for an alternative function than what it originally evolved for (Peat and Adams, 2008). The utilization of comparative and functional genomics to identify genes not in common between these two organisms may aid in identifying the aforementioned compensatory mechanisms and potentially solve the conundrum of why *Photorhabdus* bioluminesce. As such, additional analyses looking at the *lux* operons of the different *Photorhabdus* species and evaluating sequence evolution of these operons across the genus may shed a bit more light on the reasons why *Photorhabdus* bioluminesce.

Acknowledgements

This work was supported by a Brigham Young University Mentored Environment Grant. We would like to thank K. Crandall, M. Whiting, A. Harker, and G. Poinar for advice and guidance on this project.

References

Akhurst, R., Dunphy, G.B., 1993. Tripartite interactions between symbiotically associated entomopathogenic bacteria, nematodes, and their insect hosts. In: Beckage, N., Thompson, S., Federici, B. (Eds.), Parasites and Pathogens of Insects. Academic, New York, pp. 1-23.

Akhurst, R.J., Boemare, N.E., Janssen, P.H., Peel, M.M., Alfredson, D.A., Beard, C.E., 2004. Taxonomy of Australian clinical isolates of the genus *Photorhabdus* and proposal of *Photorhabdus asymbiotica* subsp *asymbiotica* subsp nov and *P. asymbiotica* subsp *australis* subsp nov. Int. J. Syst. Evol. Microbiol. 54, 1301-1310.

Akhurst, R.J., Mourant, R.G., Baud, L., Boemare, N.E., 1996. Phenotypic and DNA relatedness between nematode symbionts and clinical strains of the genus *Photorhabdus* (Enterobacteriaceae). Int. J. Syst. Bacteriol. 46, 1034-1041.

Baldwin, T.O., Devine, J.H., Heckel, R.C., Lin, J.-W., Shadel, G.S., 1989. The complete nucleotide sequence of the regulon of *Vibrio fischeri* and the luxABN region of *Photobacterium leiognathi* and the mechanism of control of bacterial bioluminescence. Journal of Bioluminescence and Chemiluminescence 4, 326-341.

Boemare, N., 2002. Interactions between the partners of the entomopathogenic bacterium nematode complexes, *Steinernema-Xenorhabdus* and *Heterorhabditis-Photorhabdus*. Brill Academic Publishers, pp. 601-603.

Boemare, N.E., Akhurst, R.J., Mourant, R.G., 1993. DNA relatedness between *Xenorhabdus* spp (Enterobacteriacae), symbiotic bacteria of entomopathogenic nematodes, and a proposal to transfer *Xenorhabdus luminescens* to a new genus, *Photorhabdus* Gen-Nov Int. J. Syst. Bacteriol. 43, 249-255.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and Combining Data in Phylogenetic Analysis. Systematic Biology 42, 384-397.

Castresana, J., 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Molecular Biology and Evolution 17, 540-552.

Ciche, T.A., Ensign, J.C., 2003. For the insect pathogen *Photorhabdus luminescens*, which end of a nematode is out? Appl. Environ. Microbiol. 69, 1890-1897.

De Queiroz, A., 1993. For consensus (sometimes) Systematic Biology 42, 368-372.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792-1797.

Farmer, J.J., Jorgensen, J.H., Grimont, P.A.D., Akhurst, R.J., Poinar, G.O., Ageron, E., Pierce, G.V., Smith, J.A., Carter, G.P., Wilson, K.L., Hickmanbrenner, F.W., 1989. *Xenorhabdus luminescens* (DNA hybridization group-5) from human clinical specimens J. Clin. Microbiol. 27, 1594-1600.

Farris, J.S., 1981. Distance data in phylogenetic analysis. In: Funk, F.A., Brooks, D.R. (Eds.), Advances in Cladistics. NY Botanical Garden, New York, pp. 3-23.

Fischer-Le Saux, M., Viallard, V., Brunel, B., Normand, P., Boemare, N.E., 1999. Polyphasic classification of the genus *Photorhabdus* and proposal of new taxa: *P. luminescens* subsp *luminescens* subsp nov., *P. luminescens* subsp *akhurstii* subsp nov., *P. luminescens* subsp *laumondii* subsp nov., *P. temperata* sp nov., *P. temperata* subsp *temperata* subsp nov and *P. asymbiotica* sp nov. Int. J. Syst. Bacteriol. 49, 1645-1656.

Forst, S., Dowds, B., Boemare, N., Stackebrandt, E., 1997. *Xenorhabdus* and *Photorhabdus* spp.: Bugs that kill bugs. Annu. Rev. Microbiol. 51, 47-72.

Forst, S., Nealson, K., 1996. Molecular biology of the symbiotic pathogenic bacteria *Xenorhabdus* spp and *Photorhabdus* spp. Microbiol. Rev. 60, 21-&.

Francino, M., Santos, S., Ochman, H., 2006. Phylogenetic relationships of bacteria with special reference to endosymbionts and enteric species. pp. 41-59.

Friedland, J., Hastings, J.W., 1967. Nonidentical subunits of bacterial luciferase: Their isolation and recombination to form active enzyme. Proceedings of the National Academy of Sciences of the USA 58, 2336–2342.

Fukushima, M., Kakinuma, K., Kawaguchi, R., 2002. Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the gyrB gene sequence. J. Clin. Microbiol. 40, 2779-2785.

Gerrard, J.G., Joyce, S.A., Clarke, D.J., Ffrench-Constant, R.H., Nimmo, G.R., Looke, D.F.M., Feil, E.J., Pearce, L., Waterfield, N.R., 2006. Nematode symbiont for *Photorhabdus asymbiotica*. Emerg. Infect. Dis 12, 1562-1564.

Gerrard, J.G., McNevin, S., Alfredson, D., Forgan-Smith, R., Fraser, N., 2003. *Photorhabdus* species: Bioluminescent bacteria as emerging human pathogens? Emerg. Infect. Dis 9, 251-254.

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774-786.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95-98.

Hazir, S., Stackebrandtz, E., Lang, E., Schumann, P., Ehlers, R.U., Keskin, N., 2004. Two new subspecies of *Photorhabdus luminescens*, isolated from *Heterorhabditis bacteriophora*

(Nematoda : Heterorhabditidae): *Photorhabdus luminescens* subsp *kayaii* subsp nov and *Photorhabdus luminescens* subsp *thracensis* subsp nov. Syst. Appl. Microbiol. 27, 36-42.

Holder, M., Lewis, P., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nature Reviews Genetics 4, 275-284.

Huelsenbeck, J.P., 1991. When are Fossils better than Extant Taxa in Phylogenetic Analysis? Systematic Zoology 40, 458-469

Huelsenbeck, J.P., Bull, J.J., Cunningham, C.W., 1996. Combining data in phylogenetic analysis. Trends in Ecology & Evolution 11, 152-158.

Hyrsl, P., Ciz, M., Lojek, A., 2004. Comparison of the bioluminescence of *Photorhabdus* species and subspecies type strains. Folia Microbiol. 49, 539-542.

Kasai, S., Okada, K., Hoshino, A., Iida, T., Honda, T., 2007. Lateral Transfer of the lux Gene Cluster. J Biochem 141, 231-237.

Koppenhofer, H.S., 2007. Bacterial symbionts of *Steinernema* and *Heterorhabditis*. In: Nguyen, K.B., Hunt, D.J. (Eds.), Entomopathogenic Nematodes: Systematics, Phylogeny and Bacterial Symbionts. Brill, Leiden, pp. 735-808.

Kuwabara, S., Cormier, M.J., Dure, L.S., Kreiss, L.S., Pfuderer, P., 1965. Crystalline bacterial luciferase from *Photobacterium fischeri.* Proceedings of the National Academy of Sciences of the USA 53, 822–828.

Liu, J., Berry, R., Poinar, G., Moldenke, A., 1997. Phylogeny of *Photorhabdus* and *Xenohabdus* species and strains as determined by comparison of partial 16S rRNA gene sequences. Int. J. Syst. Bacteriol. 47, 948-951.

Maddison, W., 1997. Gene trees in species trees. Systematic Biology 46, 523-536.

Maddison, W.P., Maddison, D.R., 2002. MacClade version 4.0.5. Sinauer, Sunderland, Massachusetts.

Meighen, E.A., Szittner, R.B., 1992. Multiple repetitive elements and organization of the lux operons of luminescent terrestrial bacteria. J. Bacteriol. 174, 5371-5381.

O'Kane, D.J., Prasher, D.C., 1992. Evolutionary origins of bacterial bioluminescence. Mol. Microbiol. 6, 443-449.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol Biol Evol 5, 568-583.

Peat, S.M., Adams, B.J., 2008. Natural selection on the *luxA* gene of bioluminescent bacteria. Symbiosis 46, 101-108.

Peel, M.M., Alfredson, D.A., Gerrard, J.G., Davis, J.M., Robson, J.M., McDougall, R.J., Scullie, B.L., Akhurst, R.J., 1999. Isolation, identification, and molecular characterization of strains of *Photorhabdus luminescens* from infected humans in Australia. J. Clin. Microbiol. 37, 3647-3653.

Poinar, G.O.J., 1990. Biology and taxonomy of Steinernematidae and Heterorhabditidae. In: Gaugler, R., Kaya, H.K. (Eds.), Entomopathogenic Nematodes in Biological Control. CRC Press, Boca Raton, FL, pp. 23-62.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Rainey, F.A., Ehlers, R.U., Stackebrandt, E., 1995. Inability of the polyphasic approach to systematics to determine the relatedness of the genera *Xenorhabdus* and *Photorhabdus* Int. J. Syst. Bacteriol. 45, 379-381.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Savageau, M.A., 1974. Genetic regulatory mechanisms and the ecological niche of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 71, 2453-2455.

Siebert, D.J., 1992. Tree statistics; trees and 'confidence'; consensus trees; alternatives to parsimony; character weighting; character conflict and its resolution. In: Forey, P.L., Humphries, C.J., Kitching, I.J., Scotland, R.W., Siebert, D.J., Williams, D.M. (Eds.), Cladistics: A Practical Course in Systematics. Oxford University Press Inc., New York, pp. 72-88.

Sorenson, M.D., Franzosa, E.A., 2007. TreeRot version 3. Boston University, Boston, MA.

Stamatakis, A., 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688-2690.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods. In: Associates, S. (Ed.), Sunderland, Massachusetts.

Tailliez, P., Laroui, C., Ginibre, N., Paule, A., Pages, S., Boemare, N., 2009. Phylogeny of *Photorhabdus* and *Xenorhabdus* based on universally conserved protein-coding sequences and implications for the taxonomy of these two genera. Proposal of new taxa: *X. vietnamensis* sp. nov., *P. luminescens* subsp. *caribbeanensis* subsp. nov., *P. luminescens* subsp. *hainanensis* subsp. nov., *P. temperata* subsp. *khanii* subsp. nov., *P. temperata* subsp. *tasmaniensis* subsp. nov., and the reclassification of *P. luminescens*

subsp. *thracensis* as *P. temperata* subsp. *thracensis*. International Journal of Systematic and Evolutionary Microbiology, ijs.0.014308-014300.

Talavera, G., Castresana, J., 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Systematic Biology 56, 564-577.

Toth, T., Lakatos, T., 2008. *Photorhabdus temperata* subsp *cinerea* subsp nov., isolated from *Heterorhabditis* nematodes. Int. J. Syst. Evol. Microbiol. 58, 2579-2581.

Tullius, M.V., Harth, G., Horwitz, M.A., 2003. Glutamine synthetase GlnA1 is essential for growth of *Mycobacterium tuberculosis* in human THP-1 macrophages and guinea pigs. Infection and Immunity 71, 3927-3936.

Wang, R.F., Cao, W.W., Franklin, W., Campbell, W., Cerniglia, C.E., 1994. A 16S rDNA-based PCR method for rapid and specific detection of *Clostridium perfringens* in food Molecular and Cellular Probes 8, 131-138.

Waterfield, N.R., Ciche, T., Clarke, D., 2009. *Photorhabdus* and a host of hosts. Annu. Rev. Microbiol. 63, 557-574.

Waterfield, N.R., Sanchez-Contreras, M., Eleftherianos, I., Dowling, A., Yang, G., Wilkinson, P., Parkhill, J., Thomson, N., Reynolds, S.E., Bode, H.B., Dorus, S., ffrench-Constant, R.H., 2008. Rapid Virulence Annotation (RVA): Identification of virulence factors using a bacterial genome library and multiple invertebrate hosts. Proceedings of the National Academy of Sciences 105, 15967-15972.

Wiens, J.J., 2003. Missing Data, Incomplete Taxa, and Phylogenetic Accuracy. Systematic Biology 52, 528-538.

Wiens, J.J., 2006. Missing data and the design of phylogenetic analyses. Journal of Biomedical Informatics 39, 34-42.

Yamamoto, S., Harayama, S., 1995. PCR amplification and direct seqeuncing of *gyr*B genes with universal primer and their application to the detection and taxonomic analysis of *Pseudomonas putrida* strains Appl. Environ. Microbiol. 61, 1104-1109.

Yamamoto, S., Harayama, S., 1998. Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *ropD* and 16S rRNA genes. Int. J. Syst. Bacteriol. 48, 813-819.

Table Legend

Table 1

List of species/subspecies and strains of *Photorhabdus* used in the analyses and their GenBank accession numbers.

Figure Legends

Figure 1

Photorhabdus 16S, *gyrB*, and *glnA* phylogenies illustrating the discordance between gene trees. Gene trees were constructed in RAxML using the GTRGAMMAI model of nucleotide evolution for 16S and *glnA* and the GTRGAMMA model for *gyrB*.

Figure 2

Mixed models Bayesian tree for the 49 taxon dataset (including missing data) with posterior probability values indicated above branches. Analyses were run for 20,000,000 generations, sampling every 1000 generation, and partitioning by gene with the 16S partition run under the TrN+I+G model, the *glnA* partition run under the TIM+I+G model, and the *gyrB* partition run under the SYM+G model.

Figure 3

Maximum likelihood tree for the 49 taxon dataset constructed in RAxML using the GTRGAMMA model of nucleotide evolution with partitioning by gene. Likelihood bootstrap (1000 replicates) values are indicated above branches.

Figure 4

Mixed models Bayesian analysis of the 37 taxon dataset (no missing data). Posterior probability values indicated above branches. Analyses were run for 20,000,000 generations, sampling every 1000 generation, and partitioning by gene with the 16S partition run under the HKY+I+G model, the *glnA* partition run under the SYM+G model, and the *gyrB* partition run under the TrN+G model.

Figure 5

Maximum likelihood analysis for the 37 taxon dataset (no missing data) constructed in RAxML using the GTRGAMMA model of nucleotide evolution with partitioning by gene. Likelihood bootstrap (1000 replicates) values are indicated above branches.

Figure 6

Single most parsimonious tree for the 37 taxon dataset (no missing data) constructed using the new technology search in TNT with ratcheting, fusing, and drifting and 1000 random addition sequences. Partitioned Bremer support values are indicated above branches (*gyrB*/16S/*glnA*) and parsimony bootstrap values are indicated below branches.

Figure 7

Bioluminescent intensity of five species/subspecies of *Photorhabdus* mapped onto the combined Bayesian tree. Bioluminescence values (mV) of 120 X 10^6 bacterial cells at 37ºC (from Hyrsl et. al., 2004) values are indicated on corresponding clades. The basal *P. temperata* clade exhibits

the highest bioluminescent intensity (1698 mV), while the intermediate *P. asymbiotica* clade exhibits the second highest intensity (406 mV), and the terminal *P. luminescens* clade exhibits the lowest intensities (22 – 188 mV).

Table 1: List of species/subspecies and strains of *Photorhabdus* used in the analyses and their GenBank accession numbers.

Species	Strain	*gyrB*	*glnA*	16S
Photorhabdus asymbiotica asymbiotica	a949(ATCC43949)	GU731082	GU731132	Z76752
P. asymbiotica asymbiotica	a948(ATCC43948)	GU731081	GU731131	Not Available
P. asymbiotica asymbiotica	a950(ATCC43953)	GU731083	GU731133	Not Available
P. asymbiotica asymbiotica	a951(ATCC43951)	GU731084	GU731134	*Z76754*
P. asymbiotica asymbiotica	a952(ATCC43952)	GU731085	GU731135	Z76753
P. asymbiotica australis	9802892	Ay278496	Not Available	AY280572
P. asymbiotica australis	H1Gladstone	GU731092	GU731142	Not Available
P. asymbiotica australis	H2BeauDesert	GU731093	GU731143	Not Available
P. asymbiotica australis	H4Melbourne	GU731095	GU731145	Not Available
P. asymbiotica australis	H5Wangaratta	GU731096	GU731146	Not Available
P. asymbiotica australis	H6Murwillumbah	GU731097	GU731147	Not Available
P. asymbiotica australis	GCH001	AY278500	Not Available	AY280574
P. asymbiotica australis	MB	AY278511	Not Available	AY280573
P. asymbiotica s ubsp.	HIT	GU731102	GU731152	AY278671
P. asymbiotica subsp.	JUN	GU731110	GU731160	AY278670
P. luminescens kayaii	FR33	EU930349	Not Available	EU930334
P. luminescens kayaii	ITH-LA3	EU930350	Not Available	EU930333
P. luminescens luminescens	Hb	GU731098	GU731148	AY278640
P. luminescens luminescens	Hm	GU731105	GU731155	AY278641
P. luminescens luminescens	MX4A	GU731116	GU731166	Not Available
P. luminescens ssp. akhurstii	EG1	GU731090	GU731140	Not Available
P. luminescens ssp. akhurstii	EG2	GU731091	GU731141	AY278644
P. luminescens ssp. akhurstii	IND	GU731108	GU731158	AY278643
P. luminescens ssp. akhurstii	IS5	GU731109	GU731159	AY278645
P. luminescens ssp. akhurstii	LN2	GU731113	GU731163	AB355866
P. luminescens ssp. akhurstii	W14	GU731121	GU731171	AY278642
P. luminescens ssp. laumondii	ARG	GU731086	GU731136	*AY278650*
P. luminescens ssp. laumondii	AZ36	GU731088	GU731138	AY278649
P. luminescens ssp. laumondii	Brecon	GU731089	GU731139	*AY278647*
P. luminescens ssp. laumondii	HK86	GU731103	GU731153	Not Available
P. luminescens ssp. laumondii	HP88	GU731106	GU731156	*AY278648*
P. luminescens ssp. laumondii	TT01	GU731120	GU731170	*NC_005126*
P. luminescens thracensis	DSM15199T	Not Available	Not Available	AJ560634
P. luminescens thracensis	FR32	EU930352	Not Available	EU930335
P. sp.	Wx13	GU731125	GU731175	Not Available
P. sp.	X4	GU731130	GU731180	Not Available
P. temperata	He86	GU731099	GU731149	Not Available
P. temperata	Heliothidis	GU731100	GU731150	AY278658
P. temperata	Hepialius	GU731101	GU731151	Not Available
P. temperata	MEG1	GU731114	GU731164	AY278655
P. temperata	NC19	GU731117	GU731167	AY278657
P. temperata	OH1	GU731118	GU731168	AY278656
P. temperata	Pmeg	GU731119	GU731169	Not Available
P. temperata	Wx10	GU731122	GU731172	AY278663
P. temperata	Wx11	GU731123	GU731173	AY278664
P. temperata	Wx12	GU731124	GU731174	AY278665
P. temperata	Wx6	GU731126	GU731176	AY278659
P. temperata	Wx8	GU731127	GU731177	AY278660
P. temperata	Wx9	GU731128	GU731178	AY278661
P. temperata	Wx9Hyper	GU731129	GU731179	AY278662
P. temperata temperata	H4	GU731094	GU731144	AY278654
P. temperata temperata	HL81	GU731104	GU731154	AY278653
P. temperata temperata	HSH2	GU731107	GU731157	AY278652
P. temperata temperata	K122	GU731111	GU731161	AY278651
P. temperata thracensis	AZ29	GU731087	GU731137	AY278668
P. temperata thracensis	KOH	GU731112	GU731162	AY278667
P. temperata thracensis	MOL	GU731115	GU731165	AY278669
Salmonella enterica	SC-B67	NC_006905	NC_006906	NC_006907
Xenorhabdus nemtophila	AN6	AY322431	Not Available	Not Available
Xenorhabdus nemtophila	DSM3370	Not Available	Not Available	AY278674
Yersinia pestis	Pestoides F	NC_009381	NC_009381	NC_009381

Figure 1

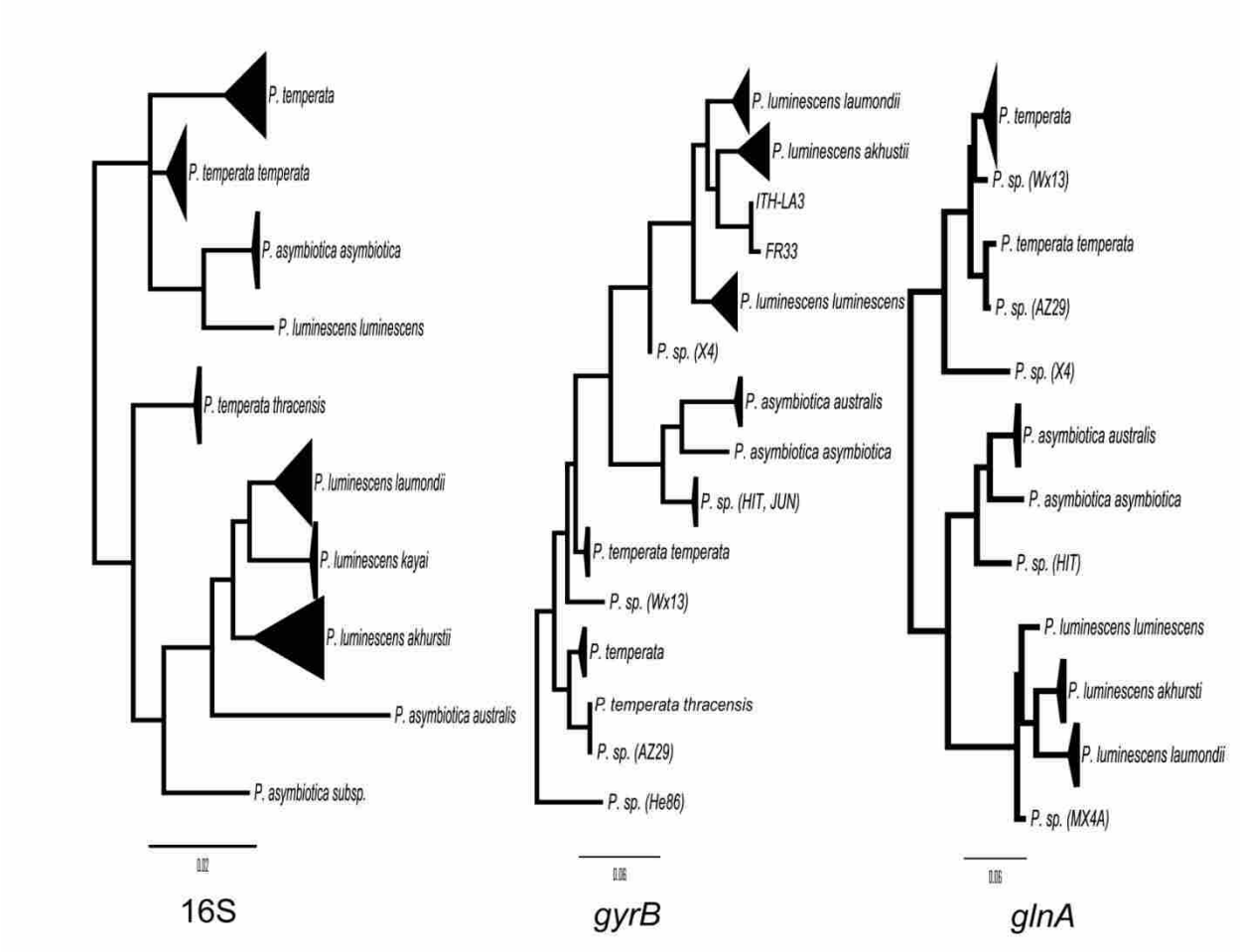

P. temperata
P. temperata temperata
P. asymbiotica asymbiotica
P. luminescens luminescens
P. temperata thracensis
P. luminescens laumondii
P. luminescens kayai
P. luminescens akhurstii
P. asymbiotica australis
P. asymbiotica subsp.
0.02
16S
P. luminescens laumondii
P. luminescens akhustii
ITH-LA3
FR33
P. luminescens luminescens
P. sp. (X4)
P. asymbiotica australis
P. asymbiotica asymbiotica
P. sp. (HIT, JUN)
P. temperata temperata
P. sp. (Wx13)
P. temperata
P. temperata thracensis
P. sp. (AZ29)
P. sp. (He86)
0.05
gyrB
P. temperata
P. sp. (Wx13)
P. temperata temperata
P. sp. (AZ29)
P. sp. (X4)
P. asymbiotica australis
P. asymbiotica asymbiotica
P. sp. (HIT)
P. luminescens luminescens
P. luminescens akhursti
P. luminescens laumondii
P. sp. (MX4A)
0.05
glnA

Figure 2

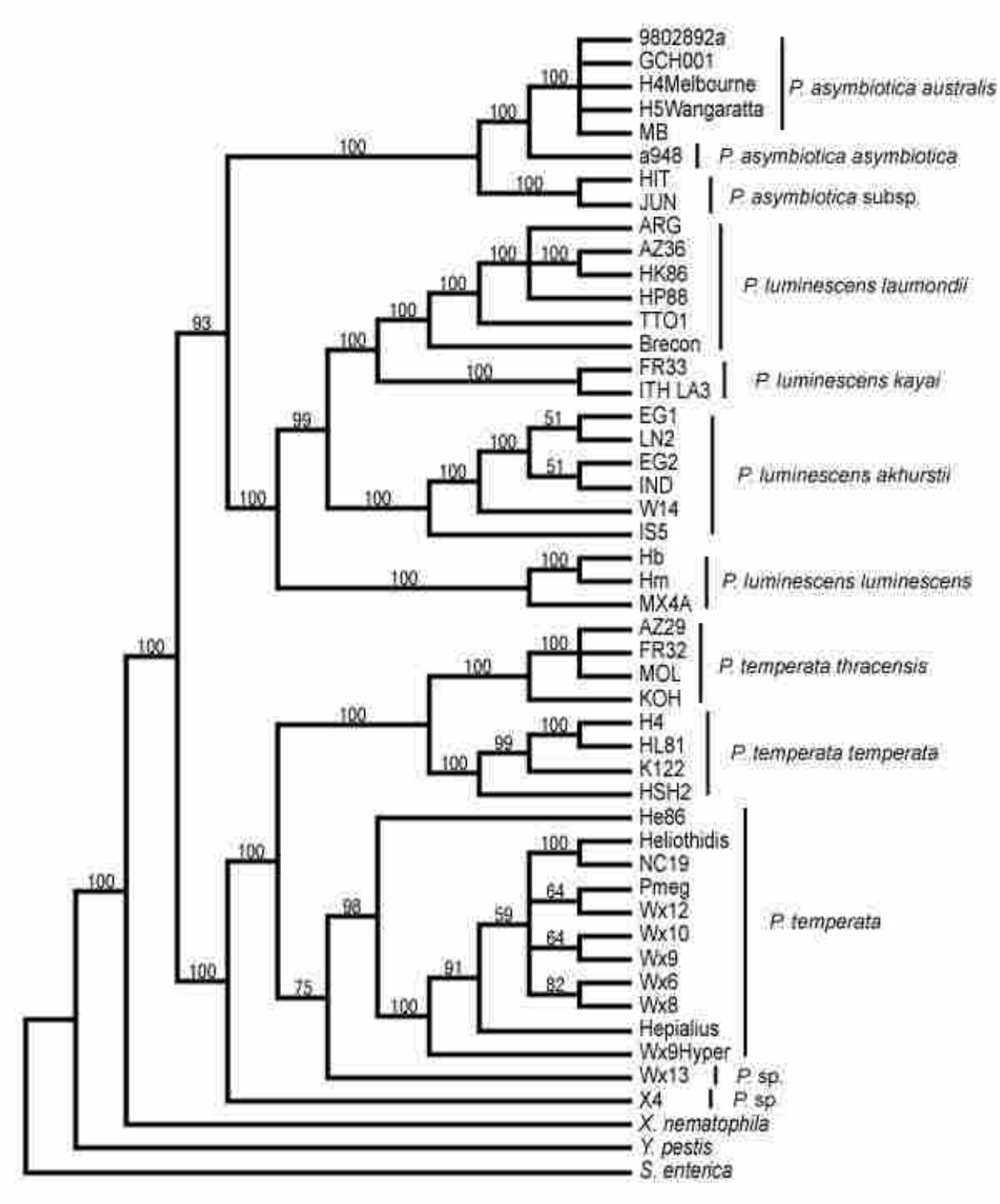

9802892a
GCH001
H4Melbourne
H5Wangaratta
MB
P. asymbiotica australis
a948
P. asymbiotica asymbiotica
HIT
JUN
P. asymbiotica subsp.
ARG
AZ36
HK86
HP88
TTO1
Brecon
P. luminescens laumondii
FR33
ITH LA3
P. luminescens kayai
EG1
LN2
EG2
IND
W14
IS5
P. luminescens akhurstii
Hb
Hm
MX4A
P. luminescens luminescens
AZ29
FR32
MOL
KOH
P. temperata thracensis
H4
HL81
K122
HSH2
P. temperata temperata
He86
Heliothidis
NC19
Pmeg
Wx12
Wx10
Wx9
Wx6
Wx8
Hepialius
Wx9Hyper
P. temperata
Wx13
P. sp.
X4
P. sp.
X. nematophila
Y. pestis
S. enterica

Figure 3

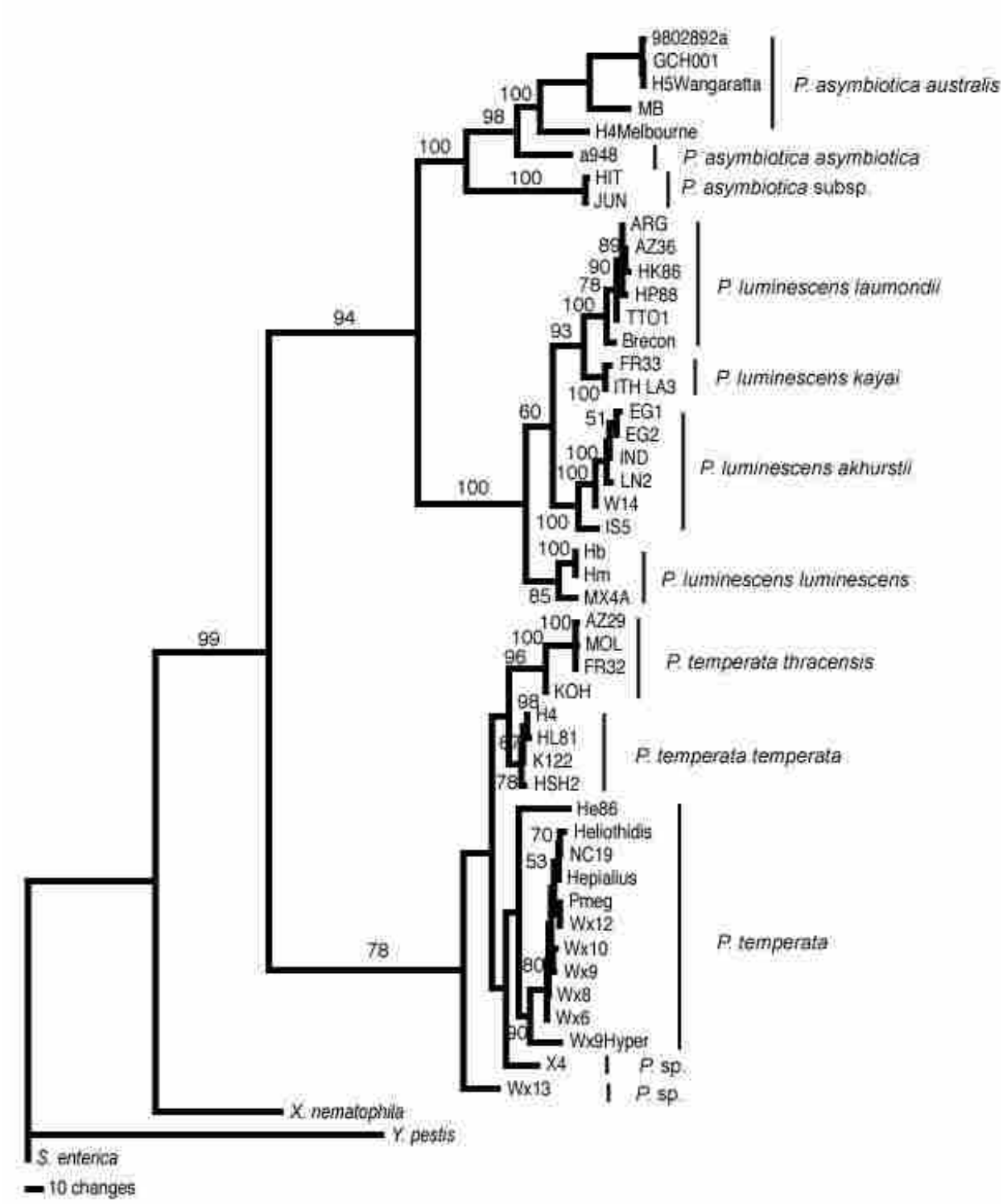

9802892a
GCH001
H5Wangaratta
MB
H4Melbourne
a948
HIT
JUN
P. asymbiotica australis
P. asymbiotica asymbiotica
P. asymbiotica subsp.
ARG
AZ35
HK86
HP88
TTO1
Brecon
FR33
ITH LA3
P. luminescens laumondii
P. luminescens kayai
EG1
EG2
IND
LN2
W14
IS5
P. luminescens akhurstii
Hb
Hm
MX4A
P. luminescens luminescens
AZ29
MOL
FR32
KOH
P. temperata thracensis
H4
HL81
K122
HSH2
P. temperata temperata
He86
Heliothidis
NC19
Hepialius
Pmeg
Wx12
Wx10
Wx9
Wx8
Wx6
Wx9Hyper
P. temperata
X4
Wx13
P. sp.
P. sp.
X. nematophila
Y. pestis
S. enterica
10 changes

Figure 4

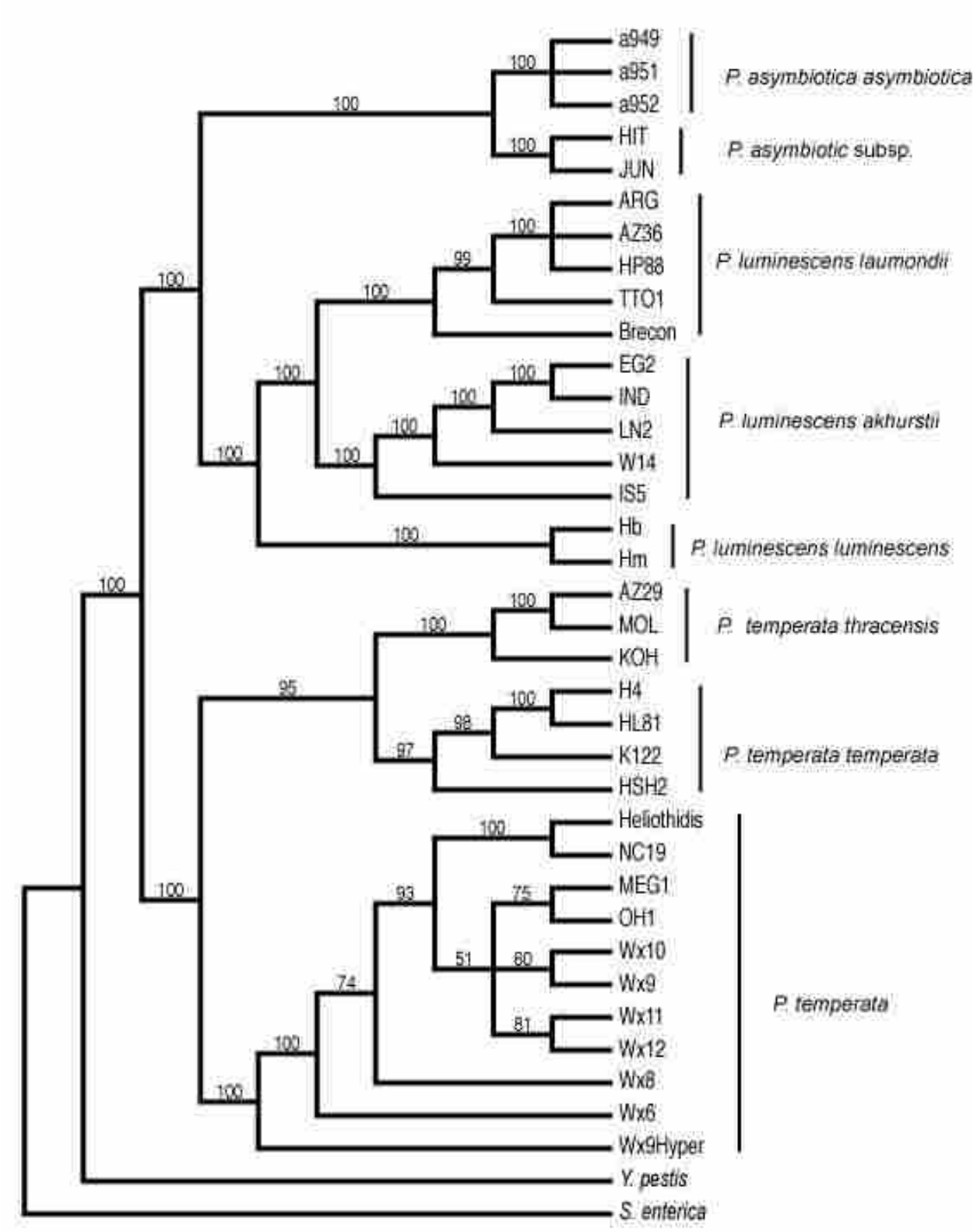

a949
a951
a952
HIT
JUN
ARG
AZ36
HP88
TTO1
Brecon
EG2
IND
LN2
W14
IS5
Hb
Hm
AZ29
MOL
KOH
H4
HL81
K122
HSH2
Heliothidis
NC19
MEG1
OH1
Wx10
Wx9
Wx11
Wx12
Wx8
Wx6
Wx9Hyper
Y. pestis
S. enterica
P. asymbiotica asymbiotica
P. asymbiotic subsp.
P. luminescens laumondii
P. luminescens akhurstii
P. luminescens luminescens
P. temperata thracensis
P. temperata temperata
P. temperata

Figure 5

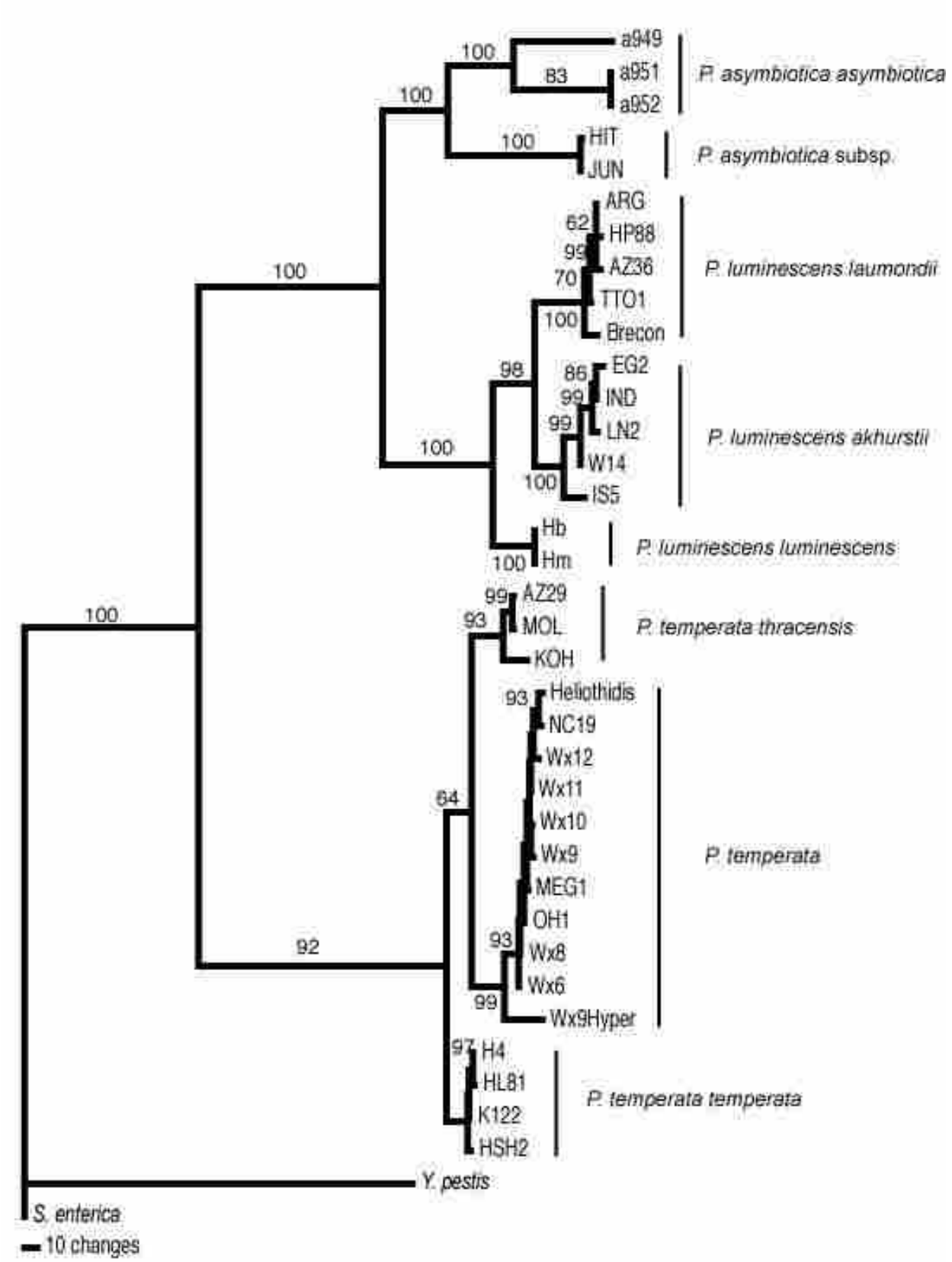

a949
a951
a952
HiT
JUN
ARG
HP88
AZ36
TTO1
Brecon
EG2
IND
LN2
W14
IS5
Hb
Hm
AZ29
MOL
KOH
Heliothidis
NC19
Wx12
Wx11
Wx10
Wx9
MEG1
OH1
Wx8
Wx6
Wx9Hyper
H4
HL81
K122
HSH2
P. asymbiotica asymbiotica
P. asymbiotica subsp.
P. luminescens laumondii
P. luminescens akhurstii
P. luminescens luminescens
P. temperata thracensis
P. temperata
P. temperata temperata
Y. pestis
S. enterica
10 changes

Figure 6

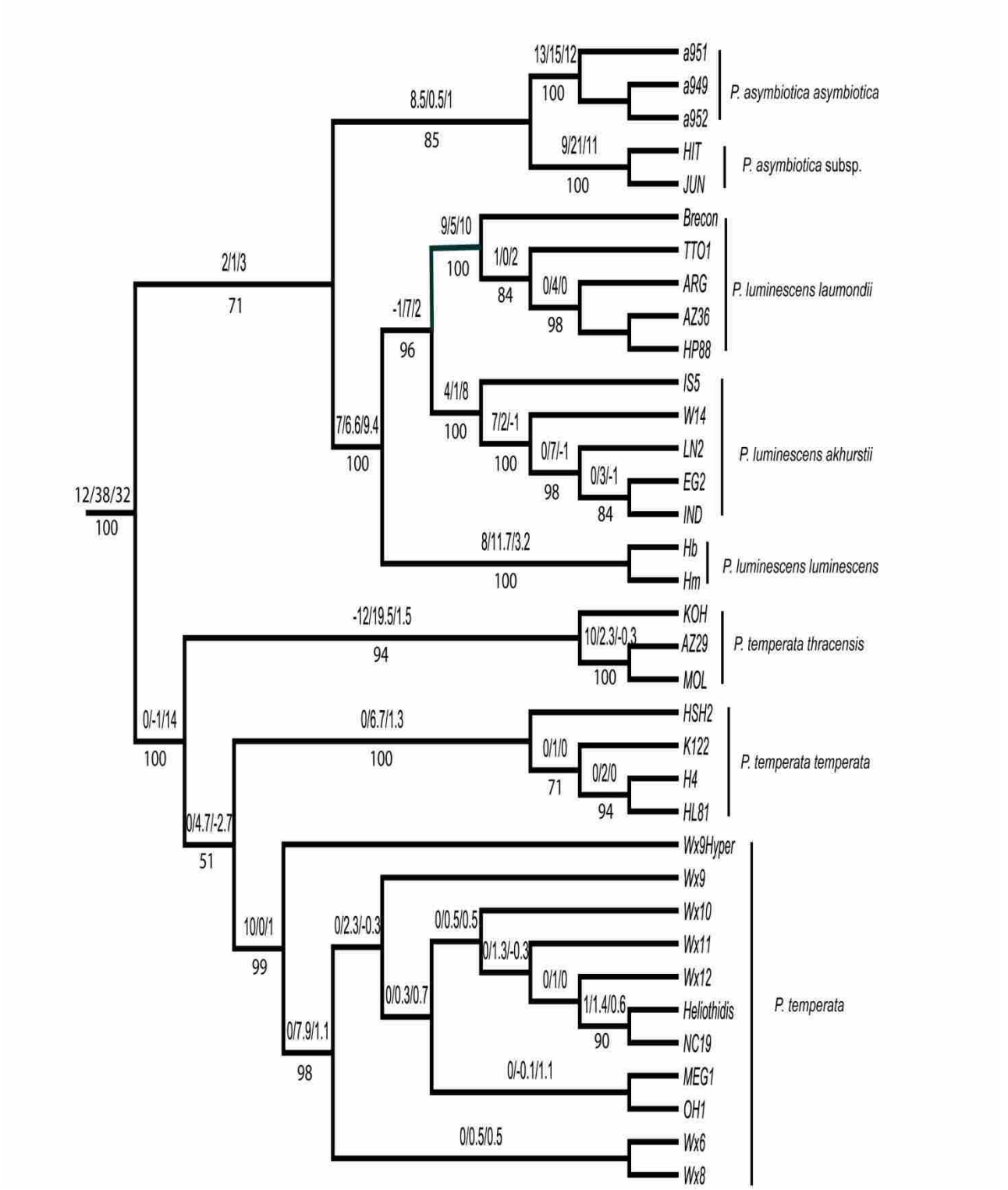
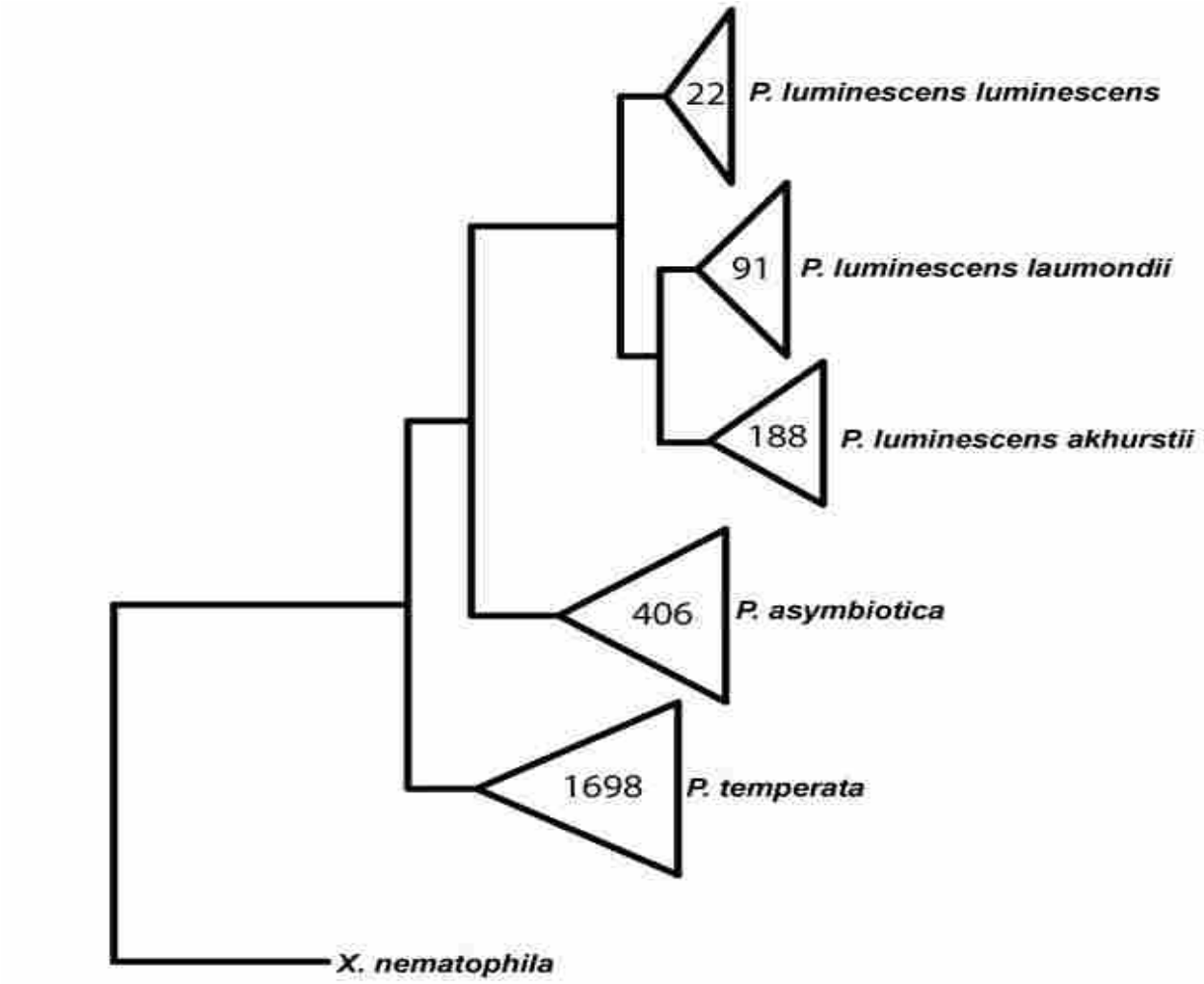

13/15/12
100
a951
a949
a952
P. asymbiotica asymbiotica
8.5/0.5/1
85
9/21/11
100
HIT
JUN
P. asymbiotica subsp.
9/5/10
100
1/0/2
84
0/4/0
98
Brecon
TTO1
ARG
AZ36
HP88
P. luminescens laumondii
2/1/3
71
-1/7/2
96
4/1/8
100
7/2/-1
100
0/7/-1
98
0/3/-1
84
IS5
W14
LN2
EG2
IND
P. luminescens akhurstii
7/6.6/9.4
100
12/38/32
100
8/11.7/3.2
100
Hb
Hm
P. luminescens luminescens
-12/19.5/1.5
94
10/2.3/-0.3
100
KOH
AZ29
MOL
P. temperata thracensis
0/-1/14
100
0/6.7/1.3
100
0/1/0
71
0/2/0
94
HSH2
K122
H4
HL81
P. temperata temperata
0/4.7/-2.7
51
10/0/1
99
0/2.3/-0.3
0/0.5/0.5
0/1.3/-0.3
0/1/0
1/1.4/0.6
90
0/0.3/0.7
0/7.9/1.1
98
0/-0.1/1.1
0/0.5/0.5
Wx9Hyper
Wx9
Wx10
Wx11
Wx12
Heliothidis
NC19
MEG1
OH1
Wx6
Wx8
P. temperata

Figure 7


22
P. luminescens luminescens
91
P. luminescens laumondii
188
P. luminescens akhurstii
406
P. asymbiotica
1698
P. temperata
X. nematophila

Chapter 4

Fine Scale Morphological Characterization of Mycetophagous and Entomophagous *Deladenus* (=*Beddingia*) *siricidicola* (Tylenchomorpha: Sphaerularoidea) Females Using Scanning Electron Microscopy

Peat, S. M.

Department of Biology, Brigham Young University, Provo, Utah, 84602

The author would like to acknowledge D. Williams for providing initial cultures of *Deladenus siricidicola* and *Amylosterium areolatum.* Additionally, the author would like to thank Michael Standing and John Gardner for helpful feedback on SEM methodology, and Byron Adams, George Poinar, Erik Ragsdale, Keith Crandall, Michael Whiting, and Alan Harker for their helpful comments on the manuscript.

E-mail: speat@byu.net

Abstract

The present study presents the first use of scanning electron microscopy (SEM) to investigate fine scale morphological differences between mycetophagous and entomophagous females of the sphaerularoid nematode *Deladenus siricidicola* in order to better understand how ultrastructural differences may facilitate interactions at two different trophic levels by the same species. Dramatic differences in head, face, and stylet morphology were observed between the two *D. siricidicola* female morphs that were not detected in previous studies using only light microscopy. Scanning electron microscopy enabled visualization of features that included a laterally compressed head, a sharply pointed stylet tip, and a stylet orifice located on the sub-terminal beveled surface of the stylet cone, all features that likely aid in the utilization of an insect host by entomophagous *D. siricidicola* females. Moreover, the squared head, more rounded stylet tip, and short stylet orifice of the mycetophagous females allow utilization of a fungal food source. Furthermore, an extracellular matrix extruding from the oral aperture was observed in only mycetophagous females and may be analagous to the feeding plug or peg found in numerous plant parasitic nematodes. Additional studies of the function of the entomophagous stylet following entry into the insect host as well as studies of stylet cone morphology in other insect parasitic tylenchids will facilitate further understanding of the functional significance of this unique stylet cone shape and stylet orifice position.

Introduction

Deladenus siricidicola, a nematode in the infarorder Tylencomorpha, superfamily Sphaerularoidea is unique in that it has two autonomous and trophically diverse life stages: a mycetophagous (fungal feeding) life stage and an entomophagous (insect parasitic) life stage (Bedding, 1967, 1968, 1972; Bedding and Iede, 2005; Poinar Jr., 1979). In its mycetophagous life stage, *D. siricidicola* uses its stylet, a needle-like feeding apparatus, to feed on the fungus *Amylostereum areolatum*, which grows in the wood of *Pinus* trees. Male and female *Deladenus* mate and the female lays her eggs in the tracheids and resin canals of the tree (Bedding, 1972; Bedding, and Akhurst, 1974). Upon hatching, nematode juveniles will either continue their mycetophagous life cycle or switch to an insect parasitic life cycle. The insect parasitic stage of *Deladenus* is induced by high carbon dioxide (CO_2) concentrations and low pH (Bedding, 1993), conditions that are usually present in the microenvironment around *Sirex* larvae (Bedding, and Akhurst, 1974; Bedding, and Iede, 2005). If these conditions are present, larval nematodes will develop into a pre-parasitic type of female, which after reproducing will use a large spear-like stylet to penetrate into *Sirex* woodwasp larvae.

The females of the two life stages of *Deladenus siricidicola* are morphologically very different. In fact, they are morphologically so different that each on its own would have been placed in a separate nematode family had not the life cycle of the nematode been carefully observed(Bedding, and Iede, 2005). One of the more pronounced of these differences is the stylet. The stylet of the mycetophagous female is approximately 10 to 11 microns long, tapers finely at the anterior end, and has well developed basal knobs. Conversely, the stylet of the infective female is much longer (14 to 30 microns), lacks conspicuous basal knobs, and is wider and stouter than the stylet of the mycetophagous female (Bedding, 1968). Differences in trophic

interactions faced by these two nematodes (fungal feeding vs. insect parasitic)can explain selection for the differing stylet size and shapes. Further fine scale differences in stylet morphology and head morphology, which facilitate interactions at different trophic levels, would also be expected between these two female morphs, though light microscopy studies have not detailed any significant differences to date.

Previous studies (Bedding, 1968) have used light microscopy to describe the morphology of these two female morphs, though no morphological studies have been conducted using scanning electron microscopy (SEM). As such, we used SEM to investigate fine scale morphological differences between mycetophagous and entomophagous females of *Deladenus siricidicola* in order to better understand how ultrastructural differences may facilitate interactions at two different trophic levels by the same species. The use of SEM for this project provides a much more detailed study of the head and stylet morphology of *D. siricidicola* than has been conducted previously.

Materials and Methods

Culturing

The mycetophagous stage of *Deladenus siricidicola* was cultured in the lab with the fungus *Amylostereum areolatum* on half strength potato dextrose agar (PDA)(9.75g PDA + 2.25g purified agar + 500ml distilled water) plates at room temperature (approximately 18˚C to 20˚C). Nematodes were washed from 25 to 30 day old plates with distilled water and approximately 50 mycetophagous females were picked into 10µl of distilled water in a petri dish using an eyelash pick (Ted Pella Inc., Redding, CA). Three different methods were evaluated to determine the best method of inducing entomophagous *D. siricidicola* females. The first method

was a mass culturing method described by Bedding and Akhurst (1974) that utilized a 500ml flask with 100g of winter red wheat and 150mls of distilled water and sealed with a cotton bung. Following inoculation, flasks were stored at room temperature for 40 to 50 days until nematodes were harvested. The second method utilized a modified growth chamber made using a FoodSaver® bag and two 1/8 inch Swedgelok fittings with a ½ hole cylindrical septa inserted into each fitting. Fittings were inserted into the top layer of the FoodSaver® bag, one at each end of the bag, to facilitate introduction of CO_2into the sealed chamber. Nematodes were cultured with *Amylosterium areolatum* on hard 0.5% lactic acid half strength PDA plates (9.75g PDA + 8.75g purified agar + 500ml distilled water). Following subculturing onto the lactic acid half strength hard PDA, plates were placed into the modified growth chamber, all of the air was removed, and the chamber was sealed using a vacuum sealer. Pure air and CO_2 were mixed using a gas mixer to a CO_2 concentration of 10% and the growth chamber was flushed withCO_2 for five minutes through the Swedgelok ports. Air samples were taken from the sealed chambers using a syringe inserted into the 1/8 inch Swedgelok fitting with a ½ hole cylindrical septa. Samples were checked using gas chromatography to ensure proper CO_2 concentrations within growth chambers. Plates were incubated for 25 to 30 days prior to harvest. The third method was carried out by culturing nematodes on hard 0.5% lactic acid half strength PDA plates that were sealed with Parafilm®. Plates were stored in sealed paper bags at room temperature for 30 to 35 days prior to harvest. While all three methods produced infective females, the Parafilm® method produced the greatest number of entomophagous females per plate with the least amount of effort, and as such the Parafilm® method was the primary method utilized to produce entomophagous females for the current study. Nematodes were washed from plates and

entomophagous females were picked into a 10 µl drop of distilled water in a petri dish using an eyelash pick (Ted Pella Inc., Redding, CA).

SEM preparation

Sequential fixation of *Deladenus* was carried out by first chilling the nematodes in a 10µl drop of tap water at 4˚C for 30 minutes. After 30 minutes, one drop of cold 3% glutaraldehyde was added every 10 to 15 minutes for two hours. Nematodes were then transferred to approximately 500 µl of fresh 3% glutaraldehyde and left at 4˚C for 24 to 48 hours (Eisenback, 1985; Eisenback, 1986). Following fixation, nematodes and glutaraldehyde were transferred to a modified BEEM capsule (the bottom of a BEEM capsule was removed and an additional lid was placed on the open bottom. Holes were punched into the BEEM capsule body and top and bottom lids using a fine probe lined with filter paper. The BEEM capsule was submerged in 0.03M sodium cacodylate buffer in a 10 ml beaker and left at room temperature for 10 to 15 minutes. After 15 minutes, all buffer was removed and replaced with fresh buffer. This process was carried out four more times for a total of 5 washes. After the final wash, all sodium cacodylate buffer was removed and 1% osmium tetroxide (OsO_4; diluted with 0.06 sodium cacodylate buffer) was added so that it completely covered the BEEM capsule. The beaker was covered with Parafilm® and left overnight. All OsO_4 was removed and residual OsO_4 was rinsed by filling a beaker with distilled water and submerging the BEEM capsule for approximately 10 minutes. The water was removed and the washing process was repeated five additional times for a total of six washes.

Following removal of all excess OsO_4, nematodes were dehydrated using an eight-step graded series of acetone, with gradual increases of 10, 30, 50, 70, 95, 100, 100, and 100%

acetone and each exchange lasting 10 to 15 minutes in duration. Following dehydration, the nematodes were critical point dried. Once dried, the nematodes were mounted onto a stub with double stick carbon tape. In an attempt to gain a better view of the mouth and lips, some nematodes were propped up onto a piece of human hair with the anterior portion of the nematode slightly projecting above the hair and at a 45˚ angle to the surface of the stub (Eisenback, 1985). Once mounted to the stub, the nematodes were sputter coated with 200 Å of gold.

Stylet extraction and preparation

Because the stylet of *Deladenus* is housed within its body, the stylet must be removed before it can be observed using SEM. The stylet was removed by placing the female nematode into a drop of 45% lactic acid on a coverslip. The head of the nematode was cut off and the stylet was pushed out through the cut opening using an eyelash pick. The stylet was cleaned by swishing it in 45%lactic acid; it was afterwards attached to the coverslip by applying pressure to the stylet. Following attachment to the coverslip, the stylet was washed by placing one drop of 2% formalin onto the lactic acid every minute for 10 minutes. The lactic acid/formalin was drained from the cover slip and fresh 2% formalin was placed back onto the extracted stylet and left for 5 to 10 minutes. The formalin was again drained and the stylet was allowed to air dry in a desiccator overnight (Eisenback, 1991; Eisenback and Rammah, 1987). Once dry, the position of the stylet was marked with a fine tipped marker (one dot on each side of the stylet) and the coverslip was attached to an SEM stub with carbon tape. Ends of the carbon tape were wrapped around two of the coverslip edges to aid in grounding the specimen. Finally, stylets were sputter coated with 200 Å of gold.

All images were collected using an environmental SEM (Philips XL30 ESEM FEG) in the Brigham Young University Microscopy Lab. Most images were taken under the high vacuum mode, though some specimens were viewed under low vacuum conditions (0.7 to 1.5 Torr) to alleviate charging issues.

Results

Head and face morphology

Scanning electron microscopy (SEM) of the head and face of both mycetophagous and entomophagous females show dramatic differences in morphology. The head of the mycetophagous female (Figs. 1A,C) is more square and even on all sides, while the head of the entomophagous female (Figs. 1B,D) is laterally compressed, forming a face in the shape of a bow tie. The oral aperture of the mycetophagous female is rounded and surrounded by six oral papillae (three on either side; Fig. 2A), while the oral aperture of the infective female (Fig. 2B) is dorso-ventrally elongated, with an oral pouch at each end of the aperture. A structure believed to be a single, subventral oral papilla was observed hidden under the ventral oral pouch of the entomophagous female (Fig. 2C); other possible oral papillae under the opposite oral pouch or in the other subventral side of the same pouch were not observed in the present study. Amphids are located approximately 500 nm from either side of the entrance to the oral aperture in the mycetophagous female while the amphids are slightly dorsal on each lateral side of the elongated oral aperture (Figs. 1,2).

The presence of a matrix, possibly a feeding plug or peg or gland secretions, extruding from the oral aperture of the mycetophagous form (Fig. 3) was observed in numerous

mycetophagous samples, though such a matrix was not observed in any of the entomophagous specimens.

Stylet morphology

SEM of both stylets revealed that a pronounced difference exists in the structure of the stylet cone. The mycetophagous stylet cone (Figs. 4A,C) is approximately 2.5 um long, 0.25 um wide, has a rounded tip, and a stylet orifice approximately 0.25 um long located approximately 0.25 um from the tip of the stylet cone on the ventral surface. Conversely, the stylet cone of the entomophagous female (Figs. 4B,D) is approximately 4 um long, 1.5 um wide, ends in a pointed tip, and is slightly beveled. The stylet orifice is located on the sub-terminal beveled ventral surface. Stylet knobs are prominent and well developed in the mycetophagous female, in contrast to its being less prominent in the entomophagous female (Figs. 4C,D).

Vulva

Mycetophagous female *D. siricidicola* are easily recognized by the presence of a protuberant vulva (Figs. 7A,C) while the less conspicuous vulva of the entomophagous female is flush with the surface of her body (Figs. 7B,D).

Discussion

While much of the internal and external anatomy of *Deladenus siricidicola* has been detailed previously (Bedding, 1968), the present fine scale study of *D. siricidicola* morphology using SEM provides further insight into the mechanisms this nematode utilizes to occupy two different niches. The laterally compressed head of the entomophagous female likely facilitates

entrance into the *Sirex* larvae via the small opening made with the stylet. To our knowledge this is the first description of such compressed head morphology in an insect parasite from the nematode order Tylenchida. This is in contrast to the more squared head of the mycetophagous female, which likely is optimal for feeding on fungal hyphae. A similarly square head has been detailed in the fungal feeding form of the insect parasitic/fungal feeding tylenchid nematode, *Hexatylus viviparus*, noting that a squarer head allows for the presence of greater amounts of tissue and musculature than is found in plant parasitic tylenchids such as *Ditylenchus dipsaci* (Shepherd et al., 1983).

In the initial description of *D. siricidicola* by Bedding (1968), mycetophagous *D. siricidicola* is noted as possessing four lips each with a single papillae for a total of four oral papillae, though later publications note the presence of six oral papillae in the genus *Deladenus* (Siddiqi, 2001). The present study clearly confirms the presence of six oral papillae surrounding the mouth of mycetophagous *D. siricidicola* females (figure 3). Prior to the present study, the presence of oral papillae in *D. siricidicola* entomophagous females has never been reported. As such, the discovery of what appears to be a single papilla in the oral pouch adjacent to the oral aperture is the first description of oral papillae in entomophagous *D. siricidicola*. Furthermore, since this papilla was observed in one subventral corner of the pouch, a second papilla is likely found in the opposite corner of the same pouch, and two more papillae are likely present in the opposite oral pouch. As such, at least four oral papillae may be present in entomophagous *D. siricidicola*, though these papillae are only visible upon lifting of the oral pouches, which only seems to occur upon opening of the oral aperture.

One feature that was observed in the mycetophagous females and not in the entomophagous females was the presence of some type of extrusion from the oral aperture. We

hypothesize that this extrusion may function analogously to the feeding pegs and plugs that have been noted in numerous plant parasites (Kisiel et al., 1971; Rebois, 1980; Sobczak et al., 1999), though no similar structures have been previously reported from *Deladenus* or other fungal feeding nematodes. Feeding plugs are composed primarily of nematode secretions (Endo, 1978; Rebois, 1980) while feeding pegs are composed of plant derived materials (Razak, and Evans, 1976; Rebois, 1980). Feeding plugs and pegs in plant parasitic nematodes are both thought to function in sealing perforated cell walls of plant roots (Sobczak et al., 1999) to prevent cytoplasmic leakage out of the feeding cell (Rebois, 1980), while feeding plugs are also believed to aid in maintaining a plant parasitic nematode's feeding position for long periods of time (Kisiel et al., 1971). As such, a structure similar to a feeding plug/peg would likely prove useful for preventing cytoplasmic leakage from the fungal hyphae during mycetophagous *D. siricidicola* feeding. Additionally, since the entomophagous female uses its stylet to aid in penetration into its insect host rather than feeding sedentarily while attached to the outside of its host, it likely does not require such a structure. Further analyses utilizing molecular analyses of the exudates and/or transmission electron microscopy of mycetophagous *D. siricidicola* feeding on *Amylosterium areolatum* are needed to further understand the functional significance of this extruded material.

Previous studies of nematode stylets have shown that while the position (ventral) and shape (slit-like) of the stylet orifice is relatively similar, the length of the stylet orifice has gone through a reduction throughout the evolutionary history of the Tylenchomorpha. The stylet orifice in the fungal feeding nematode *Aphelenchus avenae* is a wide ventral slit that appears to extend from near the tip to a point about half-way down the stylet cone (Ragsdale et al., 2008). This is in contrast to the relatively small ventral stylet orifice possessed by members of the plant

parasitic nematode genus *Meloidogyne* (Eisenback, and Hirschmann, 1982). The present study shows that the stylet orifice of mycetophagous *D. siricidicola* females is a small slit-like opening on the ventral side of the stylet cone. As such, the stylet orifice of mycetophagous *D. siricidicola* females is similar in size, location, and shape to *Meloidogyne* spp. (and likely other plant parasitic nematodes). Conversely, the stylet cone of entomophagous *D. siricidicola* females appears to be unlike any stylet previously characterized, with the stylet orifice located on the sub-terminal ventral surface. A previous study has indicated that following entrance into the insect hemocoel, *D. siricidicola* uses microvilli on the surface of their body to acquire nutrients from their host (Riding, 1970). As such, entomophagous *D. siricidicola* are likely not using their stylet opening as a primary feeding mechanism, though it is possible that initial feeding takes place through the stylet until the cuticle is shed and the microvilli begin to uptake nutrients. Alternatively, the wide sub-terminal orifice might provide a port for the expulsion of secretions that aid in penetrating into host cuticle or evasion of host defenses. Furthermore, the tip of the stylet cone terminates in a sharp point rather than the rounded tip observed in mycetophagous and plant parasitic stylets. This pointed tip undoubtedly facilitates perforation of the *Sirex* larval cuticle allowing entrance of the nematode into the insect hemocoel. While we expect that other insect parasitic tylenchids will likely possess a similar stylet to the one observed in entomophagous *D. siricidicola* females, further studies are needed to evaluate if subtle variations exist in stylet morphology across insect parasitic taxa and to assess what factors (insect host, point of entrance into host, etc.) are correlated with the evolution of stylet morphology in insect parasites. Further studies of the function of the entomophagous stylet following entry into the insect host as well as studies of the stylet cones of other insect parasitic tylenchids will yield

further understanding of the functional significance of this unique stylet cone shape and stylet orifice position.

The shaft of all *D. siricidicola* stylets degraded during the extraction process, likely due to the lactic acid that was used to extract the stylets. A variety of concentrations of sodium hypochlorite (0.5% to 2%) were used as an alternative media for stylet dissections, though the lack of viscosity and rapid evaporation of these solutions limited our ability to cleanly extract intact stylets. While these solutions might limit the degradation of the stylet shaft, we were unable to extract an intact stylet using sodium hypochlorite.

References

Bedding, R. A. 1967. Parasitic and free-living cycles in entomogenous nematodes of the genus *Deladenus*. Nature 214: 174-75.

Bedding, R. A. 1968. *Deladenus wilsoni* n. sp. and *D. siricidicola* n. sp. (Neotylenchidae), entomophagous-mycetophagous nematodes parasitic in siricid woodwasps. Nematologica 14: 515-25.

Bedding, R. A. 1972. Biology of *Deladenus siricidicola* (Neotylenchidae), an entomophagous-mycetophagous nematode parasitic on siricid woodwasps. Nematologica 18: 482-93.

Bedding, R. A. 1993. Biological control of *Sirex noctilio* using the nematode *Deladenus siricidicola*. Pp. 11-20 *In* R. A. Bedding, R. J. Akhurst, and H. K. Kaya, ed. Nematodes and the Biological Control of Insect Pests. East Melbourne, Australia: CSIRO.

Bedding, R. A. and R. J. Akhurst. 1974. Use of the nematode *Deladenus siricidicola* in the biological control of *Sirex noctilio* in Australia Journal of the Australian Entomological Society 13: 129-35.

Bedding, R. A. and E. T. Iede. 2005. Application of *Beddingia siricidicola* for Sirex Woodwasp Control. Pp. 389-99 *In* P. S. Grewal, ed. Nematodes As Biocontrol Agents. Oxfordshire, UK: CABI Publishing.

Eisenback, J. D. 1985. Techniques for preparing nematodes for scanning electron microscopy. Pp. 79-105 *In* K. R. Barker, C. C. Carter, and J. N. Sasser, ed. An Advanced Treatise on *Meloidogyne*. Raleigh, North Carolina: North Carolina State University Graphics.

Eisenback, J. D. 1986. A comparison of techniques useful for preparing nematodes for scanning electron-microscopy Journal of Nematology 18: 479-87.

Eisenback, J. D. 1991. Preparation of nematodes for scanning electron microscopy: Part 2. Methods for collection and preparation of nematodes. Pp. 87–96 *In* W. R. Nickle, ed. Manual of Agricultural Nematology. New York, USA: Marcel Dekker, Inc.

Eisenback, J. D. and H. Hirschmann. 1982. Morphological comparison of stylets of male root-knot nematodes (*Meloidogyne* spp.). Scanning Electron Microscopy 2: 837-43.

Eisenback, J. D. and A. Rammah. 1987. Evaluation of the utility of a stylet extraction technique for understanding morphological diversity of several genera of plant-parasitic nematodes Journal of Nematology 19: 384-86.

Endo, B. Y. 1978. Feeding plug formation in soybean roots infected with the soybean cyst nematode. . Phytopathology 68: 1022-31.

Heckmann, R. A., O. M. Amin, and M. D. Standing. 2007. Chemical Analysis of Metals in Acanthocephalans Using Energy Dispersive X-Ray Analysis (EDXA) in Conjunction with a Scanning Electron Microscope (SEM). Comparative Parasitology 74: 388-91.

Kisiel, M., J. Castillo, and B. M. Zuckerman. 1971. Adhesive plug associated with feeding of *Hemicycliphora similis* on cranberry Journal of Nematology 3: 296-97.

Poinar Jr., G. O. 1979. Nematodes for Biological Control of Insects. Boca Raton, Florida: CRC Press.

Ragsdale, E. J., J. Crum, Mark H. Ellisman, and J. G. Baldwin. 2008. Three-dimensional reconstruction of the stomatostylet and anterior epidermis in the nematode *Aphelenchus avenae* (Nematoda: Aphelenchidae) with implications for the evolution of plant parasitism. Journal of Morphology 269: 1181-96.

Razak, A. R. and A. A. F. Evans. 1976. Intracellular tube associated with feeding by *Rotylenchulus reniformis* on cowpea root. . Nematologica 22: 182-89.

Rebois, R. V. 1980. Ultrastucture of a feeding peg and tube associated with *Rotylenchulus reniformis* in cotton. Nematologica 26: 396-405.

Riding, I. L. 1970. Microvilli on the outside of a nematode. Nature 226: 179-80.

Shepherd, A. M., S. A. Clark, and D. J. Hooper. 1983. *Hexatylus viviparus* (Nematoda, Tylenchida, Hexatylina): head skeleton mophology and a comparison of head symmetry with that of *Ditylenchus dipsaci* (Tylenchina). Revue Nematol 6: 275-83.

Siddiqi, M. R. 2001. Tylenchida: parasites of plants and insects. Wallingford: CABI Publishing.

Sobczak, M., W. A. Golinowski, and F. M. W. Grundler. 1999. Ultrastructure of feeding plugs and feeding tubes formed by *Heterodera schachtii*. Nematology 1: 363-74.

Legends for Figures

Fig. 1 Scanning electron micrographs of the head and face of mycetophagous and entomophagous *Deladenus siricidicola* females. A) side view of the head and face of a mycetophagous female showing the more squared head, and the position of the amphid apertures (aa) and circular oral aperture (oa). B) Enface view of an entomophagous female showing the laterally compressed head, and the position of the amphid apertures and elongated oral aperture. C) View of the anterior region of a mycetophagous female showing the squared head. D) View of the anterior region of an entomophagous female showing the ventrally compressed head and the bowtie-shaped face region.

Fig. 2 Scanning electron micrographs of the oral region of mycetophagous and entomophagous *Deladenus siricidicola* females. A) Enface view of a mycetophagous female showing the circular oral aperture (oa), one of the amphidal apertures (aa), and six oral papillae (op). B) Enface view of aentomophagous female showing the dorso-ventrally elongate oral aperture (oa), two oral pouches (opo), and two amphidal apertures. C) Enface view of a mycetophagous female with its oral aperture open and protruding stylet, revealing the presence of a putative subventral oral papilla (op) under an oral pouch (opo).

Fig. 3 Enface view of *Deladenus siricidicola* mycetophagous females showing the presence of an oral extrusion (oe) emerging from the oral aperture.

Fig. 4 Scanning electron micrographs of the two *Deladenus siricidicola* stylet morphs. A) Stylet cone of a mycetophagous female showing the rounded tip and the ventral position of the stylet orifice (so). B) Stylet cone of an entomophagous female showing the pointed tip, and the position of the stylet orifice (so) on the ventral sub-terminal beveled edge. C) Full stylet of a mycetophagous female showing the stylet shaft (ss), narrow stylet cone (sc), ventral position of the stylet orifice (so), and the well developed stylet knobs (sk). D) Full stylet of an entomophagous female showing the stylet shaft (ss), thick stylet cone (sc), sub-terminal location of the stylet orifice (so), and less prominent stylet knobs (sk).

Fig. 5 Scanning electron micrographs of the anterior region of *Deladenus siricidicola* females showing the tip of the mycetophagous stylet cone (sc) protruding from the oral aperture.

Fig. 6 Scanning electron micrographs of the face and head of entomophagous *Deladenus siricidicola* females. A, B) Enface view showing the tip of the stylet cone (sc) protruding from the oral aperture and the sub-terminal location of the style orifice (so). C) View of the anterior region of an entomophagous female showing the laterally compressed head and the tip of the stylet cone (sc) protruding from the oral aperture.

Fig. 7 Scanning electron micrographs of the vulva (v) of mycetophagous and entomophagous *Deladenus siricidicola* females. A, C) Image showing the characteristic protuberant vulva of a mycetophagous female.B, D) Image showing the flatter vulva of an entomophagous female.

Figures

Figure 1

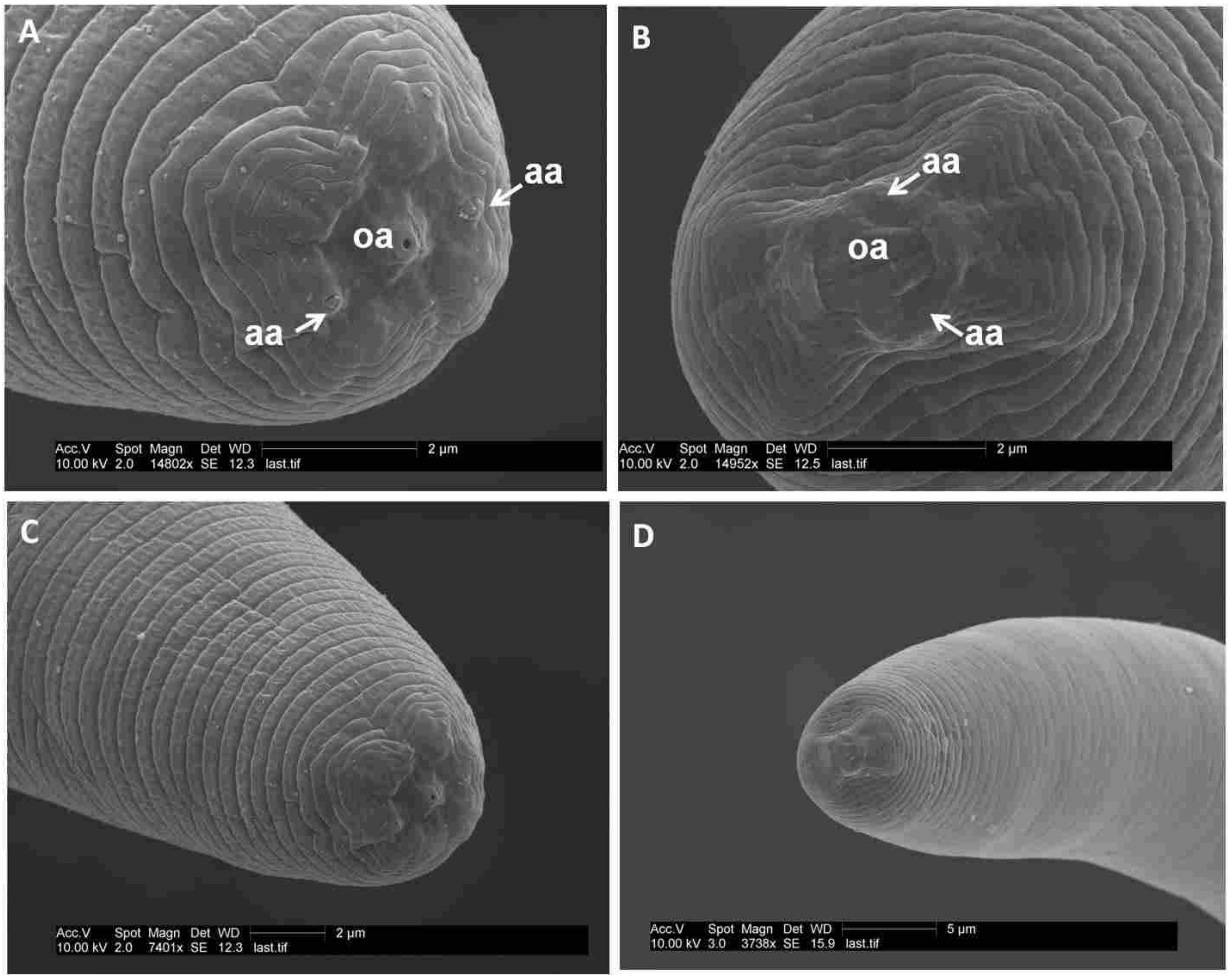


Figure 2

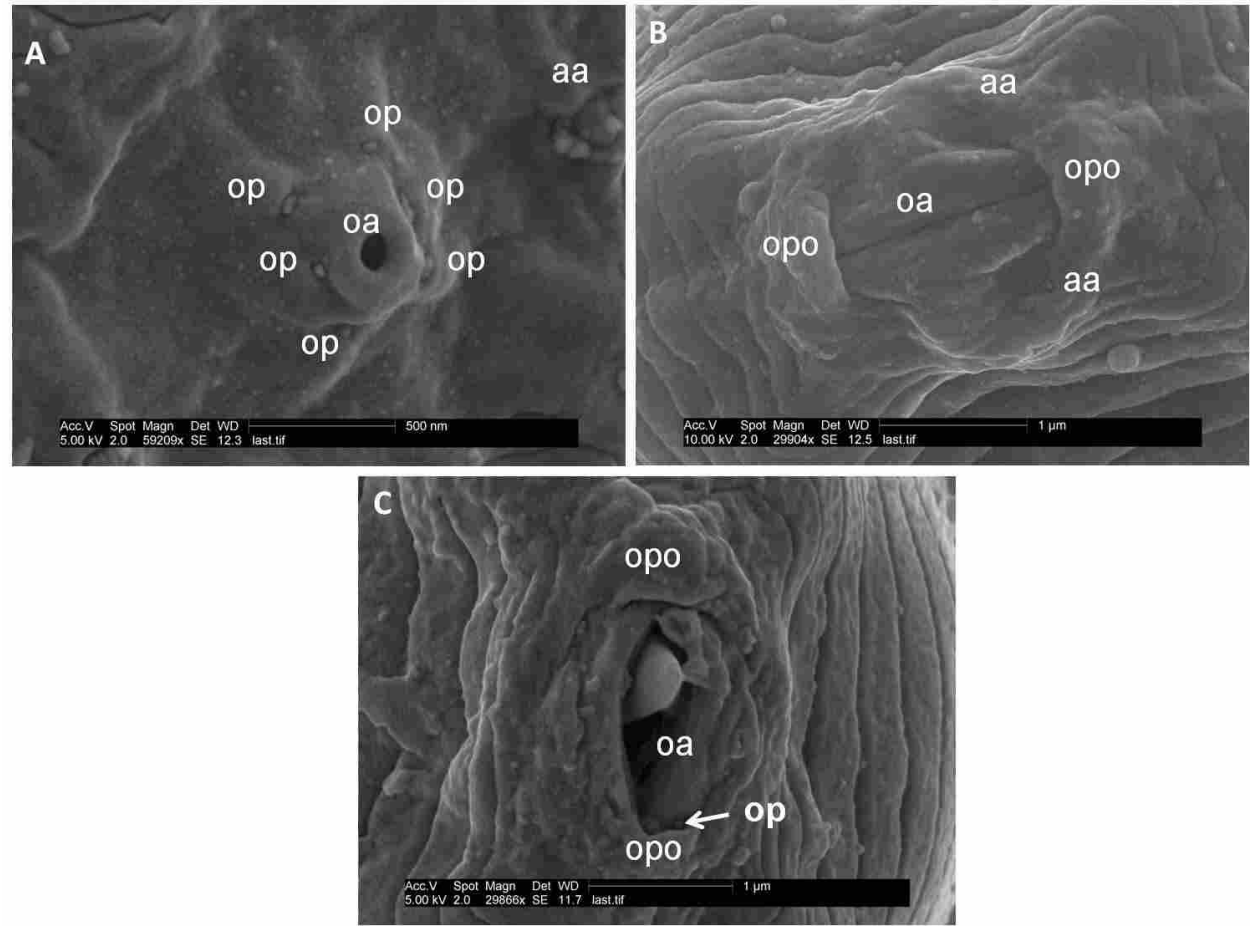

A
aa
op
op
op
oa
op
op
op
Acc.V Spot Magn Det WD
5.00 kV 2.0 59209x SE 12.3 last.tif
500 nm
B
aa
opo
oa
opo
aa
Acc.V Spot Magn Det WD
10.00 kV 2.0 29904x SE 12.5 last.tif
1 µm
C
opo
oa
op
opo
Acc.V Spot Magn Det WD
5.00 kV 2.0 29866x SE 11.7 last.tif
1 µm

Figure 3

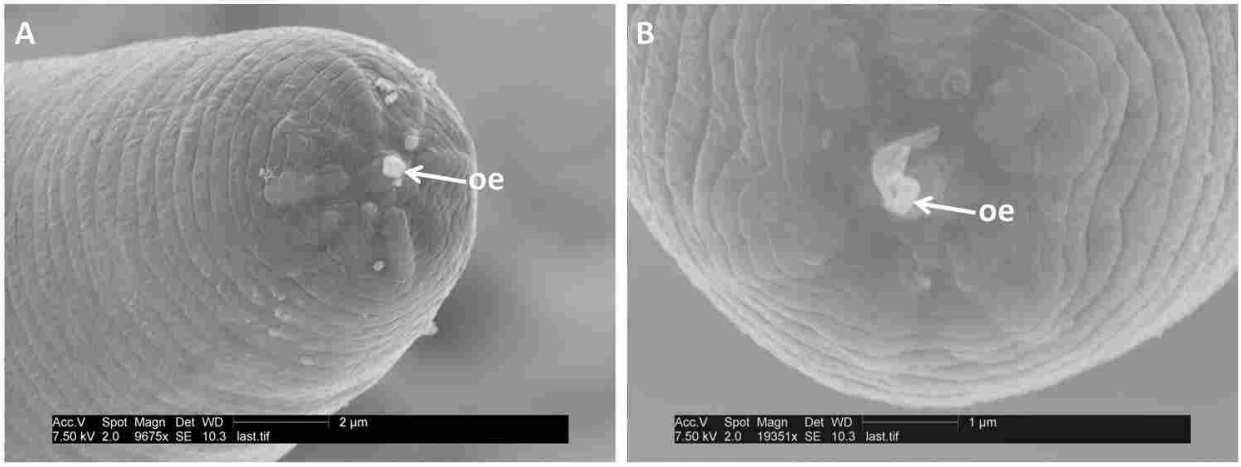

A
oe
Acc.V Spot Magn Det WD 2 µm
7.50 kV 2.0 9675x SE 10.3 last.tif
B
oe
Acc.V Spot Magn Det WD 1 µm
7.50 kV 2.0 19351x SE 10.3 last.tif

Figure 4

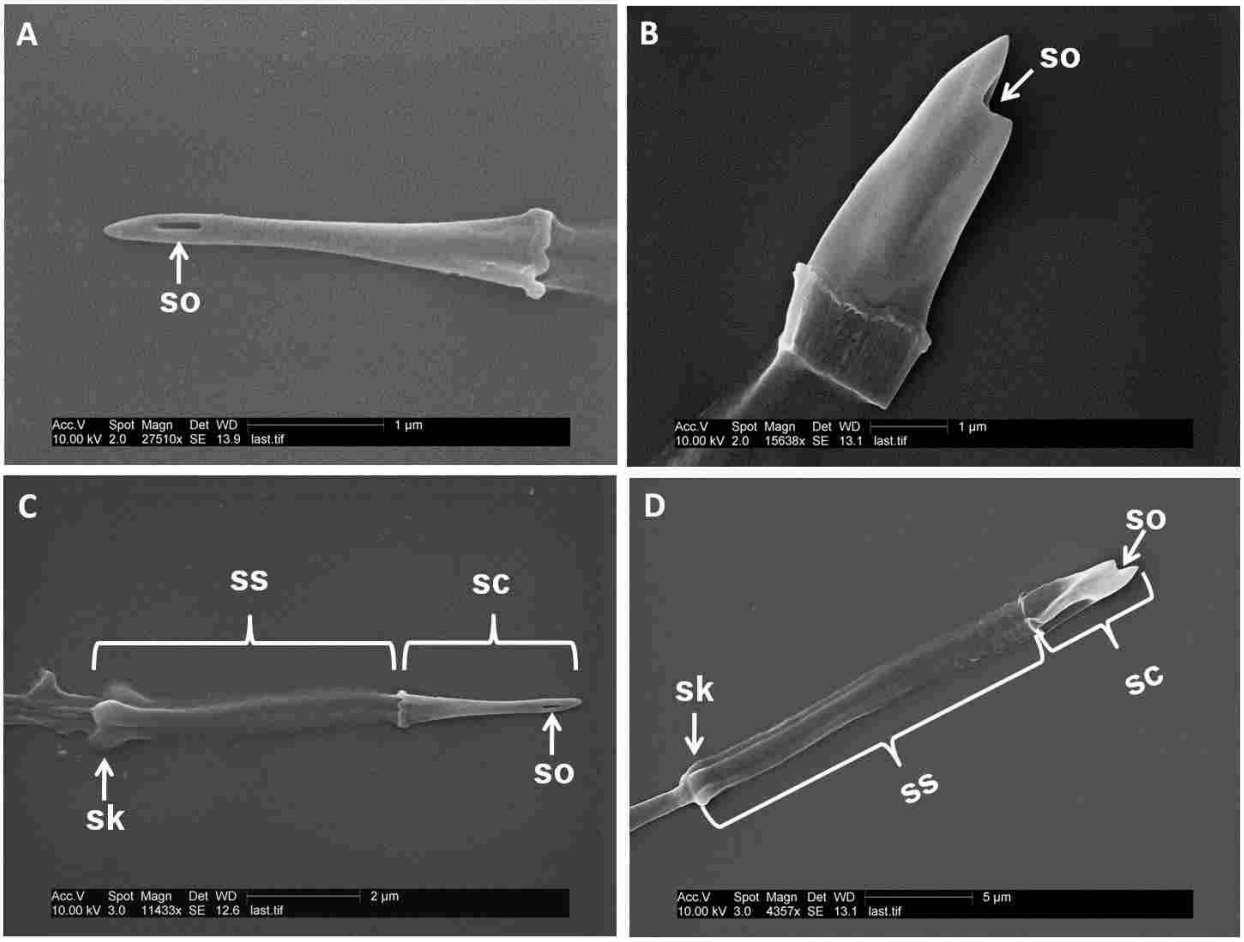

A
so
Acc.V Spot Magn Det WD 1 µm
10.00 kV 2.0 27510x SE 13.9 last.tif
B
so
Acc.V Spot Magn Det WD 1 µm
10.00 kV 2.0 15638x SE 13.1 last.tif
C
ss
sc
so
sk
Acc.V Spot Magn Det WD 2 µm
10.00 kV 3.0 11433x SE 12.6 last.tif
D
so
sk
sc
ss
Acc.V Spot Magn Det WD 5 µm
10.00 kV 3.0 4357x SE 13.1 last.tif

Figure 5

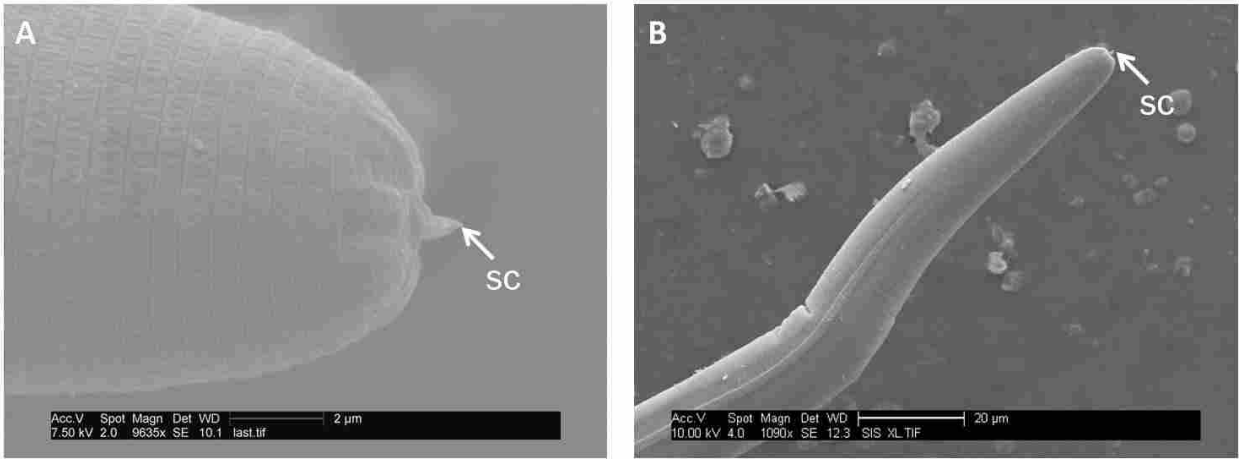

A
SC
Acc.V Spot Magn Det WD 2 µm
7.50 kV 2.0 9635x SE 10.1 last.tif
B
SC
Acc.V Spot Magn Det WD 20 µm
10.00 kV 4.0 1090x SE 12.3 SIS XL.TIF

Figure 6

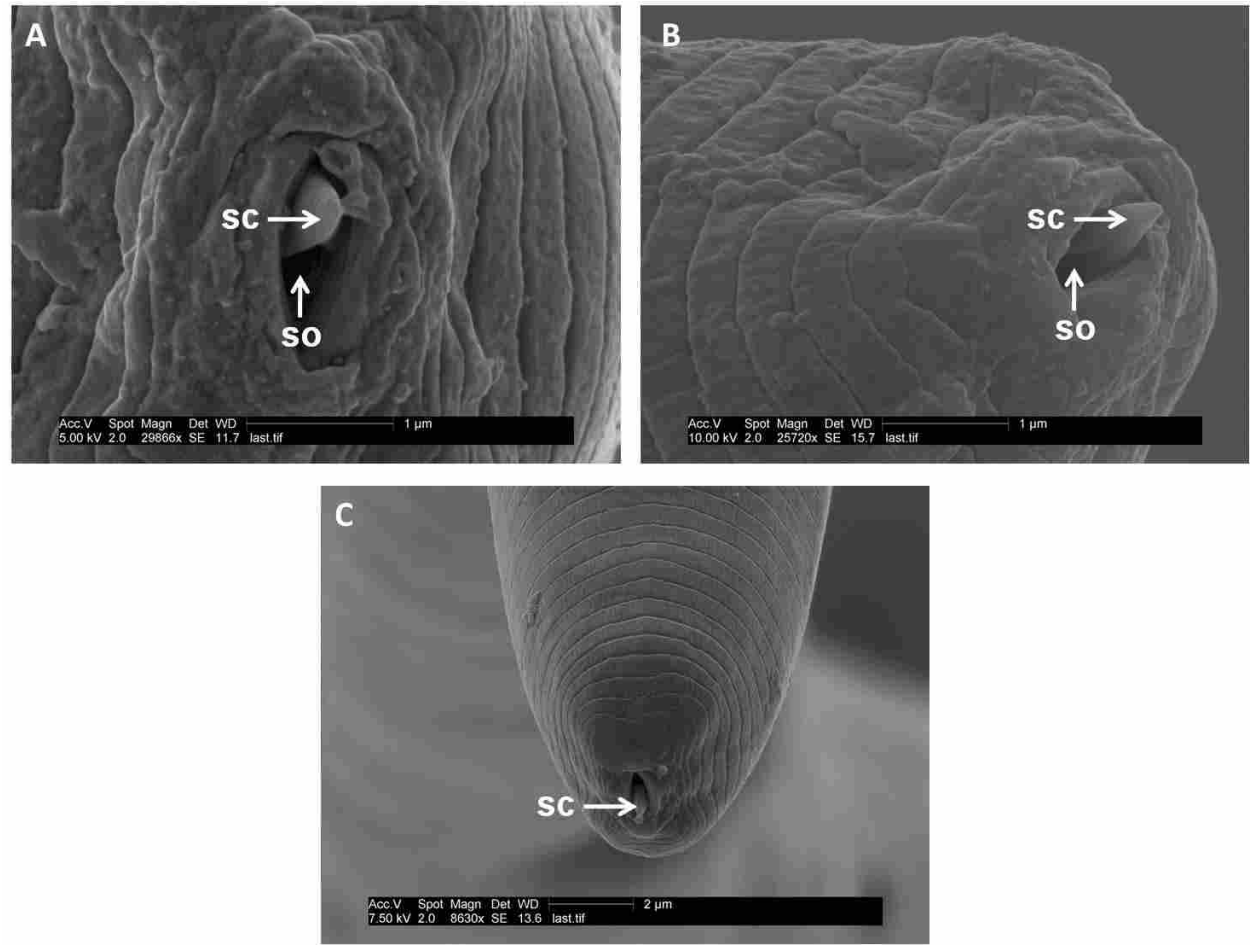
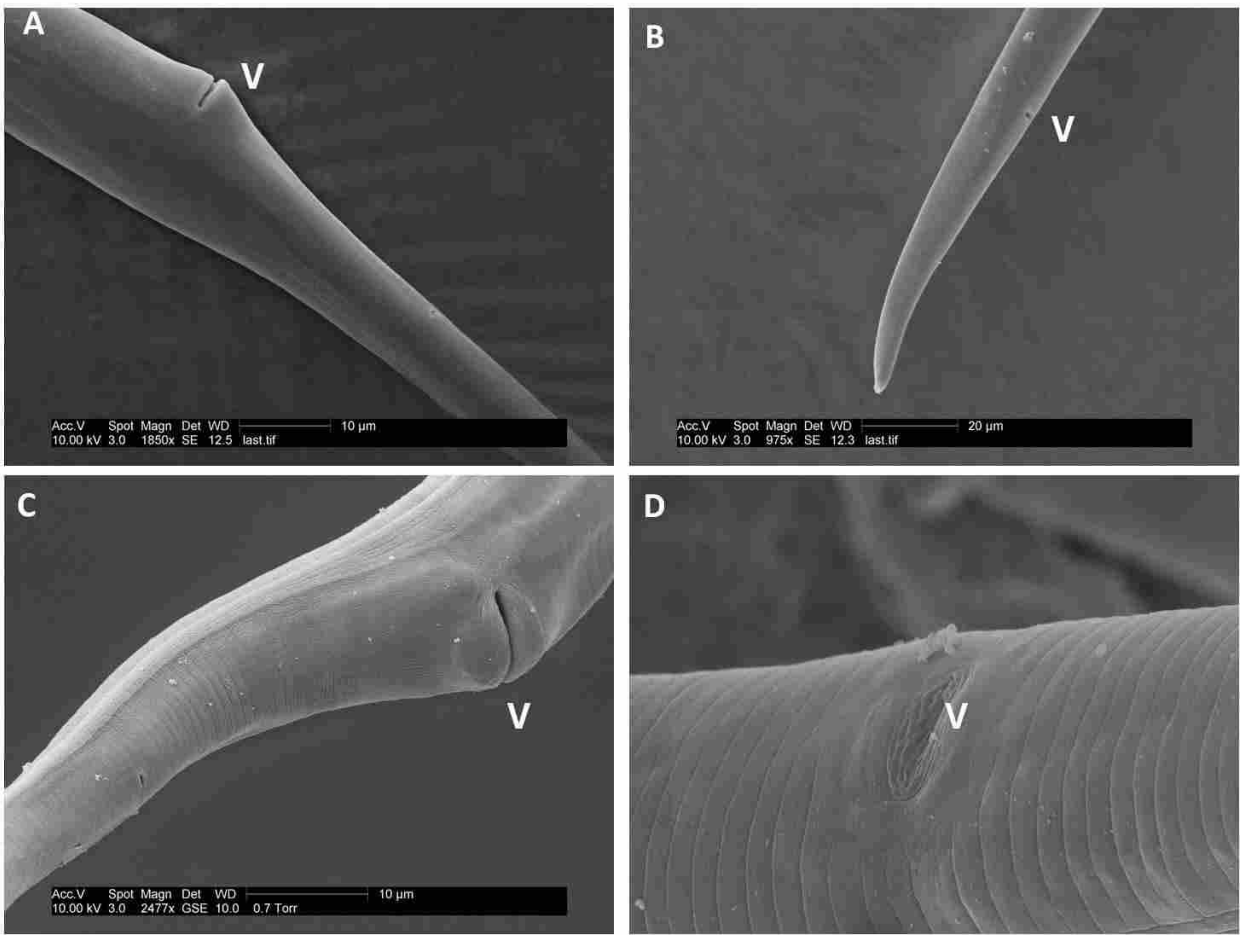

A
SC
SO
Acc.V Spot Magn Det WD 1 µm
5.00 kV 2.0 29866x SE 11.7 last.tif
B
SC
SO
Acc.V Spot Magn Det WD 1 µm
10.00 kV 2.0 25720x SE 15.7 last.tif
C
SC
Acc.V Spot Magn Det WD 2 µm
7.50 kV 2.0 8630x SE 13.6 last.tif

Figure 7


A
V
Acc.V Spot Magn Det WD 10 µm
10.00 kV 3.0 1850x SE 12.5 last.tif
B
V
Acc.V Spot Magn Det WD 20 µm
10.00 kV 3.0 975x SE 12.3 last.tif
C
V
Acc.V Spot Magn Det WD 10 µm
10.00 kV 3.0 2477x GSE 10.0 0.7 Torr
D
V

Manuscript submitted to *BMC Genomics*

Chapter 5

Comparative Transcriptomics Reveals Putative Plant and Insect Parasitism Genes from Mycetophagous and Entomophagous Life Stages of the Nematode *Deladenus* (=*Beddingia*) *siricidicola*

Scott M. Peat*, Byron J. Adams

Department of Biology, Brigham Young University, Provo, Utah, 84602

* Corresponding author

Email addresses

SMP: speat@byu.net

BJA: byron_adams@byu.edu

Abstract

Introduction

The insect parasitic nematode *Deladenus* (=*Beddingia*) *siricidicola* is unique in that it has two autonomous and trophically diverse life history stages; a mycetophagous (fungal feeding) life stage and an entomophagous (insect parasitic) life stage. The dramatically different ecological niches that are inhabited by this single species of nematode as well as the placement of *D. siricidicola* within a group of insect parasites that forms an intermediate clade between fungal feeding and parasitic nematodes in the infraorder Tylenchomorpha establishes *D. siricidicola* as an excellent model system for comparative transcriptomic analyses that explore the presence or absence of genes with a potential role in both plant and insect parasitism.

Results

We present the first analysis of the transcriptomes of both the mycetophagous and entomophagous forms of the nematode *D. siricidicola* using next generation (454) sequencing techniques. In total, 501,507 reads and 24,745 contigs and singletons were generated from mycetophagous *D. siricidicola* females with identity being assigned to 8,624 of the mycetophagous sequences, while 439,228 reads and 7,207contigs and singletons were assembled from entomophagous *D. siricidicola* females with identity being assigned to 3,216 of the entomophagous sequences. Gene ontology (GO) mapping for the identified unigenes from each library showed statistically significant differences between the two libraries for three terms within the biological process GO category, two terms within the cellular component GO category, and one term within the molecular function GO category. A number of genes were found that were expressed in the entomophagous library and not in the mycetophagous library, and as such will make interesting targets for further studies on genes that may play a role in the parasitism of insects by nematodes. Finally, a number of genes were discovered in both libraries

that are homologues of previously identified plant parasitic nematode secretory and plant parasitism genes, including genes that code for a ubiquitin extension protein, two beta-1,4-endoglucanases, annexin, cathepsin l protease, chitinase, and peroxiredoxin.

Conclusion

Our data provides the first transcriptomic characterization for the nematode *Deladenus siricidicola* and for an insect parasitic member of the nematode infraorder Tylenchomorpha. Furthermore, we have identified a number of genes that could potentially play a role in the parasitism of insects by nematodes, though more focused research on these specific genes are required to determine their functional significance in *D. siricidicola.* Finally, the numerous plant parasitism gene homologues found in both *D. siricidicola* libraries suggests that this nematode has co-opted these plant parasitism genes for other functions, possibly related to fungal feeding and insect parasitism.

Introduction

Deladenus (=*Beddingia*) *siricidicola* is a nematode parasite of the woodwasp *Sirex noctilio*, a major insect pest in *Pinus* plantations where it has been introduced [1-2]. Since the initial descriptions of the utility of *Deladenus siricidicola* as a biological control agent [3], *D. siricidicola* has been used to effectively control woodwasp populations in New Zealand, Australia, and South America [3-5]. *Deladenus siricidicola* is somewhat unique in that it has two autonomous and trophically diverse life history stages; a mycetophagous (fungal feeding) life stage and an entomophagous (insect parasitic) life stage [4, 6-8]. During the mycetophagous life stage, *D. siricidicola* juveniles feed on the fungus *Amylosterium areolatum* as it grows in *Pinus* trees. Once mature, females mate and lay their eggs in the tracheids and resin canals of the tree [3, 6]. Eggs hatch and the mycetophagous cycle resumes. When in close proximity to larval *Sirex* galleries (where increased carbon dioxide and decreased pH levels are common), *D. siricidicola* larvae are stimulated to develop into a pre-parasitic entomophagous morph [3-4, 9]. Once mated, the entomophagous female penetrates into a larval *Sirex* using its stylet (a needle like feeding apparatus), feeds within the hemocoel using microvilli on the surface of its body [6, 10], and produces and releases thousands of juveniles into the insect [4]. Juveniles enter the *Sirex* host's eggs and are deposited back into trees upon host oviposition.

What makes this system even more intriguing and unique is the fact that the mycetophagous female and entomophagous female are morphologically so distinct that when characterized based on morphological characters alone, each would be placed into a different taxonomic Family [4]. Selection for these differing morphologies has likely been driven by differing trophic interactions that each morph encounters (i.e. fugal feeder vs. insect parasite) [11], though the genetic mechanisms that are responsible for these shifts are unknown. The

differing female morphs and the dramatically different ecological niches that are inhabited by this single species of nematode provides an excellent model system to conduct comparative transcriptomic analyses for use in the identification and study of genes with a potential role in insect parasitism.

The placement of *Deladenus siricidicola* within a group of insect parasites that forms an intermediate clade between the fungal feeding and plant parasitic nematodes within the infraorder Tylenchomorpha [12] (Figure 1) makes this nematode an excellent model organism for studies on the origin and maintenance of plant parasitism genes. Several genes in plant parasitic nematodes have been suggested to enable plant parasitism [13-21], though few studies have identified genes which might play a role in the parasitism of insects by nematodes [22]. Furthermore, many of the putative plant parasitism genes have been hypothesized to have originated via horizontal gene transfer events from bacteria [23-24], though no studies have examined the presence, absence or expression of these plant parasitism genes in insect parasitic Tylenchomorpha.

Ever developing molecular techniques have enabled the sequencing of whole nematode genomes, though whole genome amplification is costly, and assembling the genome is extremely time intensive and is less tractable for non-model organisms. Alternatively, expressed sequence tag (EST) data readily allows for investigations into the diversity of gene expression and the biological functions of genes [25] without the excess cost and turnaround time which is standard on de novo genome sequencing.

Expressed sequence tag analyses have been conducted on numerous nematode taxa [26-32] and these data have proven useful in studies of gene discovery [22, 33], horizontal gene transfer [24], and evolution [34-35]. Currently, an abundance of expressed sequence tag (EST)

data are available (1,087,669 sequences on GenBank) for numerous groups within the phylum Nematoda, including plant parasitic nematodes, free living nematodes, and vertebrate parasitic nematodes, though few data are available for entomopathogenic/insect parasitic nematodes (Table 1; Figure 2). Historically, nematode EST datasets have been generated using the conventional EST generation method of amplifying cDNA, cloning the cDNA, and sequencing cDNA from clones. While this method generates an abundance of extremely useful data, advances in next generation sequencing technologies (i.e. 454 sequencing) allow for a more rapid and cost effective method of EST sequencing that produces greater numbers of ESTs and for the amplification of rare transcripts that might otherwise be missed via conventional methods [36-37].

With this in mind, we utilized 454 pyrosequencing to characterize the transcriptomes of mycetophagous and entomophagous female morphs of the nematode *D. siricidicola*. In total, we generated and analyzed 501,507 reads from mycetophagous *D. siricidicola* females and 439,228 reads from entomophagous *D. siricidicola* females. From these reads, a total of 24,745 and 7,207contigs and singletons were assembled for the mycetophagous and entomophagous libraries, respectively. We use these data to conduct numerous comparative analyses to facilitate the identification of potential insect parasitism genes as well as to identify putative plant parasitism genes in a non-plant parasite. This work provides the first characterization of the transcriptome of an insect parasitic member of the nematode infraorder Tylenchomorpha and provides a wealth of data that can inform future studies investigating potential insect parasitism genes and the origin and maintenance of plant parasitism genes.

Results and Discussion

454 sequence from female *Deladenus siricidicola* cDNA libraries

A total of 501,507 high quality reads were obtained from the mycetophagous library, while a total of 439,228 reads were obtained from the entomophagous library. Files containing the sequences for each library have been deposited in NCBI's Short Read Archive (SRA) [accession numbers SRS009919 and SRS009923]. The average read length for the mycetophagous library was 344 bp while the average read length for the entomophagous library was 329 bp. Of the 501,507 reads from mycetophagous *D. siricidicola*, 403,527 reads were assembled into 16,444 contigs and 8,301 singletons, while 374,296 reads were assembled into 3,828 contigs and 3,379 singletons from the original 433,206 entomophagous *D. siricidicola* 454 reads (Table 2). Files for each assembly have been deposited in NCBI's Transcriptome Shotgun Assembly (TSA) archive [accession numbers: mycetophagous EZ424488 - EZ449232; entomophagous EZ449233 - EZ456439].

The difference in the number of reads and the number of contigs formed are likely the result of a lower quality entomophagous cDNA library preparation and less efficient normalization reactions. Difficulties with RNA stability encountered during entomophagous RNA extractions may have contributed to lower quality total RNA template and as such produced lower quality cDNA libraries. Additionally, transcripts in the entomophagous library were represented by a much larger number of ESTs than transcripts in the mycetophagous library. For example, a transcript for a senescence-associated protein was assembled from 3,440 ESTs in the entomophagous library while the same transcript was assembled from 825 ESTs in the mycetophagous library. As such, it appears that the normalization reaction for the entomophagous library was less efficient and as such we recovered many copies of fewer genes

while from the mycetophagous library we recovered fewer copies of many genes. While the number of genes in each library varies, the use of 454 technology facilitated the amplification of two relatively robust EST datasets, informing important questions about the expressed genome of a neglected group of insect parasitic nematodes.

BLAST analyses and functional annotation of mycetophagous and entomophagous *D. siricidicola*

BlastX analysis of both the contigs and singletons were conducted against GenBank's non-redundant (nr) database using an e-value cutoff of 10^{-4}. Of the 24,745 mycetophagous contigs and singletons, we were able to assign an identity to 8,624 sequences following BlastX analysis against GenBank's nr database. Of these 8,624 well-identified sequences (e-value $\leq 10^{-4}$), 76% had best blast hits to a nematode, with 38 % to *Caenorhabditis* spp. and 36% to *Brugia malayi*. Of the 7,207entomophagous contigs and singletons, we were able to assign an identity to 2,258 sequences following BlastX analysis against GenBank's nr database. Of these 2,258 well-identified sequences (e-value $\leq 10^{-4}$), 71% had best blast hits to a nematode, with 34 % to *Caenorhabditis* spp and 29% to *Brugia malayi*.

An additional tBLASTx analysis was conducted against all nematode EST sequences from GenBank using an e-value cutoff of 10^{-4}. Of the 24,745 entomophagous contigs and singletons, 8,818 had a best blast hit to at least one nematode EST sequence. Of the 8,818 best blast hits, 26% were to free living nematodes, 29% were to vertebrate parasitic nematodes, 39% were to plant parasitic nematodes, and 6% were to entomopathogenic nematodes (Figure 3A). Of the 7,207entomophagous contigs and singletons, 2,071 had a best blast hit to at least one nematode EST sequence. Of the 2,071 best blast hits, 21% were to free living nematodes, 26%

were to vertebrate parasitic nematodes, 47% were to plant parasitic nematodes, and 6 % were to entomopathogenic nematodes (Figure 3B). While plant parasitic nematodes are represented in the nematode EST database with significantly fewer sequences (167,718) than free living (487,156) or vertebrate parasitic nematodes (376,880) (Figure 2), the percent of best blast hits to plant parasitic nematodes was larger than either of the greater represented categories. We attribute much of the greater representation of plant parasitic best blast hits to more recent, shared common ancestry, as *D. siricidicola* is in the same infraorder as the most of the representative plant parasitic nematodes in GenBank. However, one of the plant parasitic nematode EST libraries belongs to the non-Tylenchomorpha nematode *Xiphinema index*. The plant parasitic lifestyle is believed to have evolved at least three times within the phylum Nematoda [38], and as such, it is likely that some of the genes in common between *Xiphinema* and *D. siricidicola* are due to convergent evolution rather than shared ancestry. Further work examining the presence of these shared genes in other groups of nematodes will aid in determining their origin and maintenance within the phylum Nematoda.

To predict possible gene function and classification for genes identified from the BLAST analyses, Gene Ontology (GO) terms were assigned using BLAST2GO [39]. Gene ontology mapping for non-redundant BLAST analyses produced functional annotations for 6,531 genes (76% of the identified unigenes) in the mycetophagous library and 1,761 genes (78% of the identified unigenes) in the entomophagous library. Gene ontology analysis showed that cellular processes, metabolic processes, developmental processes, and multicellular organismal processes were the most represented biological process categories for both the mycetophagous and entomophagous libraries (Figure 4). When comparing the representation of biological process GO terms between the mycetophagous and entomophagous libraries, a statistically significant

difference ($p < 0.05$) was detected in metabolic process, localization and establishment of localization categories. The most represented cellular component GO terms for both libraries included intracellular part, intracellular, and intracellular organelle (Figure 5). Comparison of cellular component GO term representation between the two libraries showed a statistically significant difference in the membrane and ribonucleoprotein complex categories. The most abundant molecular function GO terms for both libraries were protein binding, hydrolase activity, nucleotide binding, and transferase activity (Figure 6). Comparison of the proportion of molecular function GO terms between the two libraries revealed a statistically significant difference in the structural constituent of ribosome category. The differential representation detected between select GO categories is likely a function of the differential expression of genes responding to the different trophic interactions (insect vs. fungi) and living environments (insect hemocoel vs. canals within tree trunks) encountered by the two *D. siricidicola* female morphs.

Most abundant transcripts from mycetophagous and entomophagous libraries

In EST studies, transcript abundance is generally a correlate for the expression level of a gene, where high abundance transcripts are interpreted as having higher expression levels than low abundance transcripts. Some of the most abundant transcripts for the mycetophagous library include a senescence-associated protein (*Brugia malayi*), phosphatidylcholine:ceramide cholinephosphotransferase (*Brugia malayi*), cytochrome c oxidase subunit iii (*Cooperia oncophora*), and lin-45 raf (*Meloidogyne artiellia*) (Table 3). Some of the most abundant transcripts for the entomophagous library include an rRNA promoter binding protein (*B. malayi*), senescence-associated protein (*B. malayi*), elegans protein (*B. malayi*), histone deacetylase3 (*B. malayi*), piwi domain containing protein (*B. malayi*), hypothetical tyrosinase-like protein

(*Caenorhabditis briggsae*), and endothelin-converting enzyme 1 (*Ancylostoma caninum*) (Table 4). Only one transcript, a senescence-associated protein, is found in the top 20 most abundantly expressed transcripts in both libraries. A homologue of this senescence-associated protein exists in the animal parasitic nematode *B. malayi*, though no putative homologues of this gene have been found in either *C. elegans* or *C. briggsae* [40].

Comparison of mycetophagous and entomophagous *D. siricidicola* transcriptomes

We hypothesize that many of the genes expressed in the entomophagous female and not in the mycetophagous female are genes that play a role in the parasitism of insects, including genes which allow inhabitation of the insect hemocoel, genes which produce a larger, more robust stylet, and genes which allow resistance to host defenses. Following a BLAST analysis of the entomophagous library against the mycetophagous library, numerous transcripts were identified that were present in the entomophagous transcriptome and absent in the mycetophagous transcriptome. These transcripts were subject to BLAST searches against both the GenBank nr database as well as the GenBank nematode EST database. Genes with best blast hits to a nematode sequence and with an e-value equal to or greater than 10^{-10} are shown in Table 5.

To further classify entomophagous transcripts not found in the mycetophagous library, transcripts with BLAST hits to the nematode EST database (65 total transcripts) were re-BLAST against subsets of the nematode EST database that were partitioned by trophic associations (i.e. plant parasitic nematodes (PPN), vertebrate parasitic nematodes (VPN), entomopathogenic nematodes (EPN), and free living nematodes (FLN)). From this, we were able to identify genes in the entomophagous library that were specific to certain groups, for example, genes only

known from insect associated nematodes or genes found only in nematodes with a parasitic lifestyle. The distribution of blast hits, expressed in Venn diagram form (Figure 7), for 65 entomophagous *D. siricidicola* genes that were not present in the mycetophagous library BLAST against the partitioned nematode EST databases shows that one entomophagous *D. siricidicola* transcript (Table 5; sequence ID F0QM4P002JUV4U) was found in only entomopathogenic nematodes. As such, this unknown protein is potentially insect parasitic/pathogenic nematode specific and should be investigated further to determine its potential role in insect parasitism. A single transcript (Table 5; sequence ID F0QM4P002I6Q44) was found in common with plant parasitic, entomopathogenic, and vertebrate parasitic nematodes only. This Acetyl-CoAcarboxylase homolog appears to be specific to parasitic nematodes in the present analysis of GenBank EST data, though an entry for a similar gene from *C. elegans* can be found in GenBank's protein database, suggesting that is likely not restricted only to nematodes with parasitic lifestyles.

A single transcript (Table 5; sequence ID F0QM4P002IO3AS) from the entomophagous library was found in common with entomopathogenic and vertebrate parasitic nematode ESTs only. This is a homologue of a neurotransmitter-gated ion-channel subunit family member gene in *C. elegans*. Further characterization and sequence analysis of this gene may aid in determining if this gene plays a unique role in animal parasitism. Nine transcripts matched only Tylenchomopha EST sequences. From our analyses, we consider these genes to be specific to the Tylenchomorpha, and as such may play a role in the evolutionary divergence of the Tylenchomorpha from other closely related nematode groups.

Many of the transcripts found only in the entomophagous library were singletons, and as such the possibility exists that transcripts of these genes may be present, but in low copy number,

in both life stages and may have been missed (i.e. not amplified/sequenced) in the mycetophagous library. However, we believe this is unlikely, as the mycetophagous library is almost three times larger than the entomophagous library. There were a number of genes that were found in abundance in the entomophagous library but were not found in the mycetophagous library. One interesting gene is an egf-like domain-containing protein whose biological process GO term suggests that it plays a role in the positive regulation of growth. While entomophagous *D. siricidicola* females are typically smaller than their mycetophagous counterparts, the stylets of the entomophagous females are considerably larger [7]. As such, further work investigating the expression levels of this egf-like domain containing protein in the stylets of these two female morphs, may uncover the genetic mechanism for the presence of a larger stylet in the entomophagous form.

Other transcripts found in the entomophagous library but not in the mycetophagous library which might provide fruitful targets for future parasitism gene research include an unknown protein similar to *D. melanogaster* GenBank sequence AAF54060.1 (sequence ID F0QM4P002JN7FN), a solute carrier family 33 (acetyl-transporter) member 1 (sequence ID contig00821),which is a multiple transmembrane protein that serves as a substrate of acetyltransferases that modify the sialyl residues of gangliosides and glycoproteins [41], and aKETtin (*Drosophila* actin binding) homolog family member (ketn-1) (Sequence ID contig02753) which is believed to function in regulating the organization of actin and the stability of myofibril in nematode muscle cells [42].

Another protein that was found only in the entomophagous library is the iron regulatory protein-1 like protein. This protein controls the synthesis of the iron storage protein ferritin and controls iron uptake and storage within mammalian cells [43-45]. In *Manduca sexta* larvae,

IRP1 binding activity declines while the production of ferritin increases in response to increased iron [45-46]. Furthermore, iron has been shown to be vital in the removal of microbes from the hemolymph of *Galleria mellonella* [47]. When entomophagous *D. siricidicola* females enter into their larval insect host, their smooth cuticle is shed, leaving their body surface covered with microvilli [10] which are used to rapidly uptake nutrients from the surrounding hemolymph, allowing for a 1000-fold increase in female nematode volume over the course of just a few weeks [4]. As such, it might be necessary for the nematode to regulate the iron content in its body due to the uptake of excess iron from the hemolymph of the insect. Decreasing the amount of iron in the hemolymph decreases the *Sirex* larvae's cellular and humoral factors [47], thus facilitating *D. siricidicola* survival, growth, and reproduction within the larval host.

Identification of putative parasitism genes in *D. siricidicola* libraries

A gene similar to a cystatin-type cystein protease inhibitor (cpi-2), which regulates proteolytic processes [22] was detected in both the mycetophagous and entomophagous libraries. Cystatins have been found in a number of nematodes including *Onchocerca volvulus[48], Haemonchus contortus[49], Brugia malayi, and Caenorhabditis elegans*[50]. Cystein protease inhibitors are considered to play a large role in interfering with host immune factors, and as such likely are a major pathogenicity factor of parasitic nematodes [22]. Homologous cystein protease inhibitor genes have been identified as being upregulated in response to insect hemolymph in the entomopathogenic nematode *Steinernema carpocapsae* [22]. An alignment of the mycetophagous and entomophagous cystatin-type cystein protease inhibitor showed that these genes are identical among 89 amino acids (but note the entomophagous EST appears to be only a partial sequence, while the mycetophagous EST is a complete sequence). The CPI-2 gene

was highly abundant in both libraries, suggesting that this gene could play a role in evading fungal and insect immune defenses. Furthermore, the cystein protease inhibitor gene in *D. siricidicola* appears to be a novel form of this gene, as an amino acid alignment to numerous other nematode cystein protease inhibitor genes showed tremendous sequence variation.

A number of secretory genes, many of which have been identified in previous studies as playing a role in the parasitism of plants, were identified (Table 6 and 7) in both libraries using Blast searches against a secretory/parasitism gene database that was constructed from sequences obtained from GenBank. Homologues to two *Meloidogyne incognita* beta-1,4 endoglucanase genes (eng-1 and eng-4) were identified in the mycetophagous library, while one form, eng-4, was detected in the entomophagous library. Both of these genes encode glycosyl hydrolase family 5 (GHF5) cellulases which function in degrading cellulose and xyloglucan, and are believed to play a role in the degradation of plant cell walls by plant parasitic nematodes [51]. Nematode plant cell wall degrading enzymes are believed to have arisen via horizontal gene transfer [23-24, 52]. Additionally, GHF5 cellulases, which have been found in plant parasitic (*Meloidogyne*, *Heterodera*, *Pratylenchus*, *Globodera*, and *Ditylenchus*) and mycetophagous (*Aphelenchus*) Tylenchomorpha genera, are suspected to have a bacterial origin [18, 52-53] while GHF45 cellulases found in *Bursaphelenchus xylophilus* likely have a fungal origin [53]. Based on the discovery of GHF 5 homologs in *Deladenus*, the intermediate phylogenetic position of *Deladenus* relative to other GHF5 possessing members of the Tylenchomorpha (Figure 1), and the absence of GHF5 cellulases in *Bursaphelenchus xylophilus*, we predict (assuming PPN GHF5 originated via an HGT event) that a single HGT event likely occurred on or before the *Aphelenchus avenae* lineage, though after the divergence of the *Bursaphelenchus xylophilus* lineage.

It has been suggested that multiple gene duplication events following the initial acquisition via HGT likely played a significant role in the diversity of GHF5 cellulases that are observed in plant parasitic Tylenchomorpha [23, 52, 54]. Additionally, an evolutionary model for the PPN GHF5 gene family put forth by Kyndt et al. [54] suggests that the ancestral GHF5 in plant parasitic nematodes was likely intronless, and that introns were gained following multiple gene duplication events. As such, it is plausible that ancestral nematode GHF5, or one or more early GHF5 paralogs, were initially utilized in some form by fungal feeding Tylenchomorpha to aid in the fungal feeding lifestyle and potentially used by insect parasitic Tylenchomorpha (i.e. *D. siricidicola*). Furthermore, at some point between the transition from a fungal feeding to a plant parasitic lifestyle, one or more paralogs of the ancestral GHF5 cellulases were likely co-opted for plant parasitism. Future analyses investigating the exon/intron structure and presence of CBD and peptide linkers in *Deladenus siricidicola* GHF5 cellulases in relation to the evolutionary model for the PPN GHF5 gene family put forth by Kyndt et al. [54] should provide a better understanding of the evolutionary history of GHF5 cellulases in Tylenchomorpha nematodes. Additionally, molecular evolutionary and selection based studies evaluating GHF5 cellulases across trophically diverse members of the Tylenchomorpha should aid in our understanding of the functional significance of beta-1,4 endoglucanases in fungal feeding/insect parasitic Tylenchomorpha nematodes.

While the function of cellulases in a fungal feeding/insect parasitic nematode are not entirely obvious based on their trophic interactions, the presence of cellulases in *D. siricidicola* might make more sense when viewed in light of their habitat. Once *D. siricidicola* eggs hatch, juvenile nematodes migrate through the tracheids of a tree in search of a *Sirex* host, all the while feeding on *Amylosterium* fungus. While it has been suggested that nematodes can only move

from tracheid to tracheid via infrequent holes in the tracheids [4], we hypothesize that the GHF5 cellulases secreted by *D. siricidicola* could potentially aid in digesting cell walls within the tracheids and facilitate movement between tracheids. Additional studies investigating movement of *D. siricidicola* from tracheid to tracheid as well as looking at the expression of cellulases from *D. siricidicola* under different conditions (i.e. while feeding on fungus, while moving though tracheids with no fungus present, etc.) will aid in determining if in fact this nematode could be utilizing cellulases to facilitate movement through tree tracheids and canals.

Homologues of *Heterodera schachtii* and *H. glycines* ubiquitin extension protein were identified in both libraries. These proteins are believed to play a role in feeding cell formation in cyst nematodes [55] as well as regulate host cell protein degradation which could work to suppress a host's defenses [14]. While it is unlikely that *D. siricidicola* uses the ubiquitin extension protein to parasitize plants, the presence of ubiquitin extension protein in both the mycetophagous and entomophagous libraries suggests that both of these morphs might utilize this protein to evade host responses when feeding on fungi or parasitizing larval insects.

Chitinase is an enzyme that can break down the beta-1,4-glycosidic bonds within chitin, a major component of fungal cell walls [56]. Expression of chitinase in the plant parasitic nematode *Heterodera glycines* has been found to be localized to the subventral oesophageal gland cells, suggesting a potential role in parasitism [56-57]. A homolog of a gene that encodes the enzyme chitinase in *Heterodera glycines* was found in the *D. siricidicola* mycetophagous library but not in the *D. siricidicola* entomophagous library. The presence of a gene coding for chitinase was expected in mycetophagous *D. siricidicola* females, as they only feed on fungus. The absence of this gene in the entomophagous library suggests that due to a lack of need to digest chitin, genes that code for chitinase have been downregulated and/or turned off in

entomophagous *D. siricidicola* females, and as such were not amplified, though additional gene expression studies need to be conducted to confirm this hypothesis.

The entomophagous library contained homologues of a number of other plant parasitism/secretory genes including 14-3-3, calreticulin, annexin 2, cathepsin 1 protease, peroxiredoxin, and glutathione peroxidase. This is the first report of plant parasitism genes in an insect parasitic nematode. While the function of these genes in this insect parasite is unclear, their presence indicates that many of the genes currently thought to facilitate plant parasitism likely have additional functions that are not specific to plant parasitism. Further functional and molecular evolutionary based analyses will hopefully shed light on the potential function of these plant parasitism genes in an insect parasite.

Conclusion

The use of 454 pyrosequencing in the present study facilitated the production of an expressed sequence tag library of approximately 32,000 sequences. As this EST library is the first from an insect parasitic member of the primarily plant parasitic infraorder Tylenchomorpha, this dataset provides a wealth of information that can be utilized to better understand gene evolution within this unique nematode order as well as aid in the discovery of genes which can help resolve phylogenetic relationships within this order. The sequencing of two libraries from the two different *D. siricidicola* female morphs has allowed for the identification of genes that may play a role in insect parasitism, and as such these identified genes provide a starting point for future work exploring the potential functions of these genes in *D. siricidicola*. The discovery of a battery of plant parasitism genes including cellulases, chitinase, and ubiquitin within both libraries, indicates that these genes are modified from their plant parasitic homologues and are

likely playing a role in both fungal feeding and insect parasitism, though further analyses are needed to fully understand the reason these non-plant parasites express 'putative parasitism genes'. While the transcriptomic data presented in the present study will aid in furthering the understanding of gene maintenance and evolution within the order Tylenchida, more transcriptomic data is needed, particularly for the fungal feeding and insect parasitic tylenchid nematodes, to facilitate more complete studies of parasitism genes, horizontal gene transfer events, and their impact in driving the evolution of Tylenchomorpha nematodes.

Methods

Nematodes and culture conditions

Deladenus siricidicola used in the present study was the E7OAA isolate. Mycetophagous *D. siricidicola* was cultured on a North American isolate of the fungus *Amylosterium areolatum,* which was grown on half-strength potato dextrose agar (PDA) plates. Plates were stored in sealed paper bags at room temperature for 25 to 30 days. Mycetophagous nematodes were washed from plates and mycetophagous females were picked into a 4% hyamine hydroxide wash, washed again in ultra-pure water, and transferred to RNAlater (Ambion, Austin, TX). Nematodes were stored in RNAlater at -80C until RNA extractions were performed.

Induction of *D. siricidicola* entomophagous females was carried out by culturing nematodes on hard 0.5% lactic acid half-strength PDA plates that were sealed with parafilm. Plates were stored in sealed paper bags at room temperature for 30 to 35 days prior to harvest. Nematodes were washed from plates and infective females were picked into a 4% hyamine hydroxide wash, washed again in ultra-pure water and transferred to RNAlater. Nematodes were stored in RNAlater at -80˚C until RNA extractions were performed.

RNA Extraction

Two separate *D. siricidicola* cDNA libraries were constructed. The mycetophagous library was made from only mycetophagous females and the entomophagous library was made from only pre-parasitic females. Three separate RNA extractions were carried out on 500 to 750 fresh and RNAlater preserved nematodes for each library. Extractions were carried out by adding 500 to 750 nematodes to 500ul of Trizol (Molecular Research Center Inc., Cincinnati, OH) in an ice-cold mortar and pestle, freezing the nematode/trizol solution with liquid N_2, and crushing. Once sufficiently crushed, nematode lysate was homogenized using a 20-gauge needle, and RNA was purified from the homogenate using an RNeasy kit (Qiagen, Valencia, CA). Isolated RNA was treated with DNase using the Turbo DNase Kit (Ambion Inc., Austin, TX) to eliminate genomic DNA contamination. RNA integrity was checked with an Agilent Bioanalyzer before continuing on to first strand synthesis.

cDNA library construction

First strand synthesis was carried out with approximately 1 μg of total RNA using a modified SMART first strand cDNA synthesis protocol, which uses the Accuscript High Fidelity reverse transcriptase (Stratagene, La Jolla, CA). Second strand cDNA synthesis was carried out using an Advantage HF-2 Polymerase PCR kit (Clonetech, Mountain View, CA) with 2 μl of first strand cDNA per 100 μl reaction. Following second strand amplification, PCR product was purified using a QIAquick mini elute kit (QiagenValencia, CA). Clean cDNA was normalized using the Trimmer kit (Evrogen, Moscow, Russia). Following normalization, a second round of PCR was conducted to amplify the normalized cDNA. Normalized cDNA was digested with the

restriction enzyme MmeI to facilitate the removal of adaptors. Digestions were cleaned using the Qiaquick PCR purification kit (Qiagen, Valencia, CA). Adaptor fragments were subsequently removed using AMPure Beads (Agencourt, Beverly, MA).

Library Preparation and sequencing

Sample concentrations for both of the cDNA samples were measured fluorometrically using Quant-iT picogreen dye (Invitrogen, Carlsbad, CA). Each cDNA sample was sheared by nebulization. Following nebulization, the samples were electrophoresed in a 1.5% Metaphor agarose gel (Cambrex Bioscience, East Rutherford, NJ), run in 0.5X TAE at 40V for 8 h and visualized using ethidium bromide staining. A single gel slice, representing DNA fragments ranging from 550 to 700 bp, was excised from the gel for each library and the fragments were extracted using a Qiaquick gel extraction kit (Qiagen, Germantown, MD). Library preparation and emPCR were carried out according to Roche 454 FLX Titanium technical manuals. Sequencing was performed at the Brigham Young University DNASC (Provo, UT) using a Roche-454 GS FLX instrument and Titanium reagents (Brandford, CT). A single sequencing run was conducted with the mycetophagous library on one half of the plate and the entomophagous library on the other half of the plate, with the two libraries separated by a gasket.

Transcript assembly and analysis

Initial quality filtering of the 454 ESTs was performed at the machine level prior to base calling. Sequences were screened for adaptor sequences and subsequently trimmed. Short (<100 bp) sequences were filtered out prior to assembly. Contigs were assembled with a minimum

overlap of at least 40 bp and an overlap identity of at least 90%. Screened ESTs were assembled from the raw reads using the Newbler Assembler (454 Life Science, Branford, CT).

Sequence annotation was carried out using BLASTx searches against the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database with an e-value cutoff of 10^{-4}. BLAST results were imported into the program Blast2GO [39] where subsequent Gene ontology (GO) and KEGG pathway annotation was carried out. A Fisher's-exact test was utilized to test for significant differences between the mycetophagous and entomophagous libraries for each GO term. Identification of genes present in the entomophagous library that were not present in the mycetophagous library was carried out by constructing a mycetophagous blastable database. A tBLASTx analysis was then utilized to blast the entomophagous library against the mycetophagous library with an e-value cutoff of 10^{-10}. Custom perl scripts were used to extract the sequences with no hits, and those sequences were subsequently subjected to further rounds of BLASTx and tBLASTx analysis against GenBank's nr database and a database of nematode ESTs that were obtained from GenBank. Identification of secretory and putative parasitism genes was carried out for both libraries using tBLASTx analysis against a custom plant parasitic nematode secretory and putative parasitism gene database with an e-value cutoff of 10^{-4}.

Authors' Contributions

SMP & BJA conceived the study. SMP was responsible for the experimental design and carried out all fundamental components of data acquisition, analysis, interpretation, and drafting of the manuscript. BJA provided supervision of the work and critical reviews of the manuscript.

Acknowledgements

This work was supported by a Brigham Young University Mentored Environment Grant, and by a National Research Initiative of the USDA Cooperative State Research, Education and Extension Service grant (number 2002-01974) to BJA. We would like to acknowledge D. Williams for providing initial cultures of *Deladenus siricidicola* and *Amylosterium areolatum*, J. Udall and J. Maughan for invaluable suggestions on cDNA library construction, and E. Wilcox for assistance with RNA quantification and sequencing. We are grateful to M. Bendall, D. Standage and J. Cheeseman for bioinformatics and perl script assistance. Finally, we would like to thank K. Crandall, M. Whiting, A. Harker, and G. Poinar for advice and guidance on this project.

References

1. Carnegie AJ, Matsuki M, Haugen DA, Hurley BP, Ahumada R, Klasmer P, Sun JH, Iede ET: **Predicting the potential distribution of *Sirex noctilio* (Hymenoptera : Siricidae), a significant exotic pest of Pinus plantations.** *Annals of Forest Science* 2006, **63:**119-128.
2. Hoebeke ER, Haugen DA, Haack RA: ***Sirex noctillo*: Discovery of a palearctic siricid woodwasp in New York.** *Newsletter of the Michigan Entomological Society* 2005, **50:**24-25.
3. Bedding RA, Akhurst RJ: **Use of the nematode *Deladenus siricidicola* in the biological control of *Sirex noctilio* in Australia** *Journal of the Australian Entomological Society* 1974, **13:**129-135.
4. Bedding RA, Iede ET: **Application of *Beddingia siricidicola* for Sirex Woodwasp Control.** In *Nematodes As Biocontrol Agents.* Edited by Grewal PS. Oxfordshire, UK: CABI Publishing; 2005: 389-399
5. Bedding RA, Akhurst RJ: **Geographical distribution and host preferences of *Deladenus* species (Nematoda Neotylenchidae) parasitic in siricid woodwasps and associated hymenopterous parasitoids** *Nematologica* 1978, **24:**286-294.
6. Bedding RA: **Biology of *Deladenus siricidicola* (Neotylenchidae), an entomophagous-mycetophagous nematode parasitic on siricid woodwasps.** *Nematologica* 1972, **18:**482-493.
7. Bedding RA: ***Deladenus wilsoni* n. sp. and *D. siricidicola* n. sp. (Neotylenchidae), entomophagous-mycetophagous nematodes parasitic in siricid woodwasps.** *Nematologica* 1968, **14:**515-525.
8. Bedding RA: **Parasitic and free-living cycles in entomogenous nematodes of the genus *Deladenus*.** *Nature* 1967, **214:**174-175.
9. Bedding RA: **Biological control of *Sirex noctilio* using the nematode *Deladenus siricidicola*.** In *Nematodes and the Biological Control of Insect Pests.* Edited by Bedding RA, Akhurst RJ, Kaya HK. East Melbourne, Australia: CSIRO; 1993: 11-20
10. Riding IL: **Microvilli on the outside of a nematode.** *Nature* 1970, **226:**179-180.
11. Yeates GW, Bongers T, Degoede RGM, Freckman DW, Georgieva SS: **Feeding habits in soil nematode families and genera - an outline for soil ecologists.** *Journal of Nematology* 1993, **25:**315-331.
12. De Ley P, Blaxter M: **Systematic position and phylogeny.** In *Biology of Nematodes.* Edited by Lee DL. Florence, KY: Taylor & Francis; 2002: 1 - 30
13. Davis EL, Hussey RS, Baum TJ: **Getting to the roots of parasitism by nematodes.** 2004, **20:**134-141.
14. Davis EL, Hussey RS, Mitchum MG, Baum TJ: **Parasitism proteins in nematode-plant interactions.** *Current Opinion in Plant Biology* 2008, **11:**360-366.
15. Hussey RS, Davis EL, Baum TJ: **Secrets in secretions: Genes that control nematode parasitism of plants.** *Brazilian Journal of Plant Physiology* 2002, **14:**183-194.
16. Qin L, Kudla U, Roze EHA, Goverse A, Popeijus H, Nieuwland J, Overmars H, Jones JT, Schots A, Smant G, et al: **Plant degradation: A nematode expansin acting on plants.** *Nature* 2004, **427:**30-30.

17. Popeijus H, Overmars H, Jones J, Blok V, Goverse A, Helder J, Schots A, Bakker J, Smant G: **Enzymology: Degradation of plant cell walls by a nematode.** *Nature* 2000, **406:**36-37.
18. Davis EL, Hussey RS, Baum TJ, Bakker J, Schots A: **Nematode parasitism genes.** *Annual Review of Phytopathology* 2000, **38:**365-396.
19. Huang GZ, Allen R, Davis EL, Baum TJ, Hussey RS: **Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103:**14302-14306.
20. Huang GZ, Dong RH, Maier T, Allen R, Davis EL, Baum TJ, Hussey RS: **Use of solid-phase subtractive hybridization for the identification of parasitism gene candidates from the root-knot nematode *Meloidogyne incognita*.** *Molecular Plant Pathology* 2004, **5:**217-222.
21. Huang GZ, Gao BL, Maier T, Allen R, Davis EL, Baum TJ, Hussey RS: **A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode *Meloidogyne incognita*.** *Molecular Plant-Microbe Interactions* 2003, **16:**376-381.
22. Hao YJ, Montiel R, Nascimento G, Toubarro D, Simoes N: **Identification, characterization of functional candidate genes for host-parasite interactions in entomopathogenetic nematode *Steinernema carpocapsae* by suppressive subtractive hybridization.** *Parasitology Research* 2008, **103:**671-683.
23. Jones JT, Furlanetto C, Kikuchi T: **Horizontal gene transfer from bacteria and fungi as a driving force in the evolution of plant parasitism in nematodes.** *Nematology* 2005, **7:**641-646.
24. Scholl EH, Thorne JL, McCarter JP, Bird DM: **Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach.** *Genome Biology* 2003, **4**.
25. Hide WA, Ptitsyn A: **What is an EST? .** In *Encyclopedia of Genetics, Genomics, Proteomics, and Bioinformatics.* Edited by Jorde LB. Hoboken N.J.: John Wiley & Sons; 2005: 1501-1509.
26. Bai XD, Grewal PS, Hogenhout SA, Adams BJ, Ciche TA, Gaugler R, Sternberg PW: **Expressed sequence tag analysis of gene representation in insect parasitic nematode *Heterorhabditis bacteriophora*.** *Journal of Parasitology* 2007, **93:**1343-1349.
27. Bai XD, Saeb ATM, Michel A, Grewal PS: **Isolation and characterization of microsatellite loci in the entomopathogenic nematode *Heterorhabditis bacteriophora*.** *Molecular Ecology Resources* 2009, **9:**207-209.
28. Kikuchi T, Aikawa T, Kosaka H, Pritchard L, Ogura N, Jones JT: **Expressed sequence tag (EST) analysis of the pine wood nematode *Bursaphelenchus xylophilus* and *B. mucronatus*.** *Molecular and Biochemical Parasitology* 2007, **155:**9-17.
29. Yin Y, Martin J, McCarter JP, Clifton SW, Wilson RK, Mitreva M: **Identification and analysis of genes expressed in the adult filarial parasitic nematode *Dirofilaria immitis*.** *International Journal for Parasitology* 2006, **36:**829-839.
30. Furlanetto C, Cardle L, Brown DJF, Jones JT: **Analysis of expressed sequence tags from the ectoparasitic nematode *Xiphinema index*.** *Nematology* 2005, **7:**95-104.
31. Haegeman A, Jacob J, Vanholme B, Kyndt T, Mitreva M, Gheysen G: **Expressed sequence tags of the peanut pod nematode *Ditylenchus africanus*: The first**

transcriptome analysis of an Anguinid nematode. *Molecular and Biochemical Parasitology* 2009, **167:**32-40.
32. Ranganathan S, Nagaraj SH, Hu M, Strube C, Schnieder T, Gasser RB: **A transcriptomic analysis of the adult stage of the bovine lungworm, *Dictyocaulus viviparus*.** *Bmc Genomics* 2007, **8**.
33. Nagaraj SH, Gasser RB, Ranganathan S: **Needles in the EST haystack: large-scale identification and analysis of excretory-secretory (ES) proteins in parasitic nematodes using expressed sequence tags (ESTs).** *Plos Neglected Tropical Diseases* 2008, **2**.
34. Scholl EH, Bird DM: **Resolving tylenchid evolutionary relationships through multiple gene analysis derived from EST data.** *Molecular Phylogenetics and Evolution* 2005, **36:**536-545.
35. Wasmuth J, Schmid R, Hedley A, Blaxter M: **On the extent and origins of genic novelty in the phylum Nematoda.** *Plos Neglected Tropical Diseases* 2008, **2**.
36. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH: **Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing.** *Molecular Ecology* 2008, **17:**1636-1647.
37. Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB: **Sampling the arabidopsis transcriptome with massively parallel pyrosequencing.** *Plant Physiology* 2007, **144:**32-42.
38. Baldwin JG, Nadler SA, Adams BJ: **Evolution of plant parasitism among nematodes.** *Annual Review of Phytopathology* 2004, **42:**83-105.
39. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21:**3674-3676.
40. Whitton C, Daub J, Quail M, Hall N, Foster J, Ware J, Ganatra M, Slatko B, Barrell B, Blaxter M: **A genome sequence survey of the filarial nematode *Brugia malayi*: repeats, gene discovery, and comparative genomics.** *Molecular and Biochemical Parasitology* 2004, **137:**215-227.
41. Hirabayashi Y, Kanamori A, Nomura KH, Nomura K: **The acetyl-CoA transporter family SLC33.** *Pflugers Archiv-European Journal of Physiology* 2004, **447:**760-762.
42. Ono K, Yu R, Mohri K, Ono S: ***Caenorhabditis elegans* kettin, a large immunoglobulin-like repeat protein, binds to filamentous actin and provides mechanical stability to the contractile apparatuses in body wall muscle.** *Molecular Biology of the Cell* 2006, **17:**2722-2734.
43. Hentze MW, Rouault TA, Caughman SW, Dancis A, Harford JB, Klausner RD: **A cis-acting element is necessary and sufficient for translational regulation of human ferritin expression in response to iron** *Proceedings of the National Academy of Sciences of the United States of America* 1987, **84:**6730-6734.
44. Schalinske KL, Chen OS, Eisenstein RS: **Iron differentially stimulates translation of mitochondrial aconitase and ferritin mRNAs in mammalian cells - Implications for iron regulatory proteins as regulators of mitochondrial citrate utilization.** *Journal of Biological Chemistry* 1998, **273:**3740-3746.
45. Zhang D, Dimopoulos G, Wolf A, Minana B, Kafatos FC, Winzerling JJ: **Cloning and molecular characterization of two mosquito iron regulatory proteins.** *Insect Biochemistry and Molecular Biology* 2002, **32:**579-589.

46. Zhang DZ, Ferris C, Gailer J, Kohlhepp P, Winzerling JJ: ***Manduca sexta* IRP1: molecular characterization and in vivo response to iron.** *Insect Biochemistry and Molecular Biology* 2001, **32:**85-96.
47. Dunphy GB, Niven DF, Chadwick JS: **Iron contributes to the antibacterial functions of the haemolymph of *Galleria mellonella*.** *Journal of Insect Physiology* 2002, **48:**903-914.
48. Lustigman S, Brotman B, Huima T, Prince AM, McKerrow JH: **Molecular cloning and characterizatio of onchocystatin, a cysteine proteinase inhibitor of *Onchocera volvulus*.** *Journal of Biological Chemistry* 1992, **267:**17339-17346.
49. Newlands GFJ, Skuce PJ, Knox DP, Smith WD: **Cloning and expression of cystatin, a potent cysteine protease inhibitor from the gut of *Haemonchus contortus*.** *Parasitology* 2001, **122:**371-378.
50. Murray J, Manoury B, Balic A, Watts C, Maizels RM: **Bm-CPI-2, a cystatin from *Brugia malayi* nematode parasites, differs from *Caenorhabditis elegans* cystatins in a specific site mediating inhibition of the antigen-processing enzyme AEP.** *Molecular and Biochemical Parasitology* 2005, **139:**197-203.
51. Smant G, Stokkermans JPWG, Yan Y, de Boer JM, Baum TJ, Wang X, Hussey RS, Gommers FJ, Henrissat B, Davis EL, et al: **Endogenous cellulases in animals: Isolation of β-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95:**4906-4911.
52. Ledger TN, Jaubert S, Bosselut N, Abad P, Rosso MN: **Characterization of a new beta-1,4-endoglucanase gene from the root-knot nematode *Meloidogyne incognita* and evolutionary scheme for phytonematode family 5 glycosyl hydrolases.** *Gene* 2006, **382:**121-128.
53. Kikuchi T, Jones JT, Aikawa T, Kosaka H, Ogura N: **A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode *Bursaphelenchus xylophilus*.** *FEBS Letters* 2004, **572:**201-205.
54. Kyndt T, Haegeman A, Gheysen G: **Evolution of GHF5 endoglucanase gene structure in plant-parasitic nematodes: no evidence for an early domain shuffling event.** *BMC Evolutionary Biology* 2008, **8:**305.
55. Tytgat T, Vanholme B, De Meutter J, Claeys M, Couvreur M, Vanhoutte I, Gheysen G, Van Criekinge W, Borgonie G, Coomans A: **A new class of ubiquitin extension proteins secreted by the dorsal pharyngeal gland in plant parasitic cyst nematodes.** *Molecular Plant-Microbe Interactions* 2004, **17:**846-852.
56. Karim N, Jones JT, Okada H, Kikuchi T: **Analysis of expressed seqeunce tags and identifiction of genes encoding cell-wall-degrading enzymes from the fungivorous nematode *Aphelenchus avenae*.** *BMC Genomics* 2009, **10:**525.
57. Gao B, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS: **The parasitome of the phytonematode *Heterodera glycines*.** *Molecular Plant-Microbe Interactions* 2003, **16:**720-726.
58. Bert W, Leliaert F, Vierstraete AR, Vanfleteren JR, Borgonie G: **Molecular phylogeny of the Tylenchina and evolution of the female gonoduct (Nematoda : Rhabditida).** *Molecular Phylogenetics and Evolution* 2008, **48:**728-744.

Figures

Figure 1-Phylogenetic relationships of Tylenchomoporpha, and the evolution of different feeding types

The ancestral feeding type is inferred to be fungivorous while the more derived feeding type is plant parasitic. *Deladenus siricidicola* forms a clade with fungal feeding Tylenchomorpha that is an intermediary between the fungal feeding Aphelenchidae (*Aphelenchus* and *Paraphelenchus*) and a very large and diverse plant parasitic clade. Figure adapted from Bert et al. [58].

Figure 2-Pie chart showing the distribution of EST sequences available in GenBank (as of November 15, 2009) for each of the major nematode groups.

Figure 3-Pie charts showing the distribution of best blast hits against GenBank nematode EST's partitioned by nematode sub-group for the mycetophagous (A) and entomophagous (B) libraries.

Figure 4-Bar graph of the proportion of biological processes gene ontology terms for unigenes identified from mycetophagous (DSM) and entomophagous (DSI) *D. siricidicola* EST libraries. Terms marked with an * denote a statistically significant difference between the two libraries.

Figure 5- Bar graph of the proportion of cellular component gene ontology terms for unigenes identified from mycetophagous (DSM) and entomophagous (DSI) *D. siricidicola*

EST libraries. Terms marked with an * denote a statistically significant difference between the two libraries.

Figure 6- Bar graph of the proportion of molecular function gene ontology terms for unigenes identified from mycetophagous (DSM) and entomophagous (DSI) *D. siricidicola* EST libraries. Terms marked with an * denote a statistically significant difference ($p \leq 0.05$) between the two libraries.

Figure 7-Venn diagram depicting the number of entomophagous *D. siricidicola* transcripts not present in the mycetophagous library with similarity to GenBank nematode ESTs grouped by nematode trophic groups:

Figure 7A - Number of ESTs with matches to ESTs of free living nematodes (FLN), vertebrate parasitic nematodes (VPN), entomopathogenic nematodes (EPN), and plant parasitic nematodes (PPN)

Figure 7B - Number of ESTs with matches to ESTs of only parasitic nematodes (i.e. vertebrate parasitic nematodes (VPN), entomopathogenic nematodes (EPN), and plant parasitic nematodes (PPN))

Tables

Table 1 – List of nematode species with available EST data on GenBank
All nematode EST sequences from GenBank (1,087,669) were formatted into a database and used in blast searches against both *D. siricidicola* libraries.

	Species	# ESTs
Vertebrate Parasitic	*Ancylostoma caninum*	80905
	Ascaris suum	56118
	Strogyloides ratti	27366
	Brugia malayi	26215
	Trichinella spiralis	25268
	Haemonchus contortus	21975
	Trichinella pseudospiralis	17330
	Onchocerca volvulus	14974
	Nippostrongylus brasiliensis	14686
	Strogyloides stercoralis	11392
	Ancylostoma ceylanicum	10651
	Parastrongyloides trichosuri	7963
	Trichuris muris	7102
	Ostertagia ostertagi	7006
	Necator americanus	6694
	Teladorsagia circumcincta	6058
	Toxocara canis	4889
	Wuchereria bancrofti	4847
	Dictyocaulus viviparus	4465
	Loa loa	4173
	Dirofilaria immitis	4005
	Trichuris vulpis	3063
	Litomosoides sigmodontis	2699
	Onchocerca flexuosa	2124
	Ascaris lumbricoides	1822
	Angiostrongylus cantonensis	1302
	Anisakis simplex	475
	Toxaskaris leonina	439
	Trichostrongylus vitrinus	368
	Oesophagostomum dentatum	299
	Parelaphostrongylus tenuis	99
	Onchocerca ochengi	60
	Brugia pahangi	28
	Ancylostoma braziliense	20
Plant Parasitic	*Meloidogyne hapla*	24452
	Heterodera glycines	24444
	Meloidogyne incognita	20334
	Bursaphelenchus xylophilus	13340
	Meloidogyne chitwoodi	12218
	Globodera rostochiensis	11851
	Xiphinema index	9351
	Globodera pallida	9020
	Meloidogyne javanica	7587
	Radopholus similis	7380
	Pratylenchus vulnus	5812
	Meloidogyne arenaria	5042
	Ditylenchus africanus	4847
	Melooidogyne paranaensis	3710
	Bursaphelenchus mucronatus	3193
	Heterodera schachtii	2812
	Pratylenchus penetrans	1916
	Zeldia punctata	391
	Globodera mexicana	17
	Heterodera avenae	1
Entomopathogenic	*Heterorhabditis bacteriophora*	53614
	Steinernema carpocapsae	2218
	Steinernema feltia	83
Free Living	*Caenorhabditis elegans*	355217
	Pristionchus pacificus	37195
	Caenorhabditis japonica	33050
	Caenorhabditis brenneri	29929
	Caenorhabditis remanei	20292
	Caenorhabditis sp. 5 AC-2008	3868
	Plectus murrayi	2591
	Aphelenchus avenae	2586
	Caenorhabditis briggsae	2424
	Panagrolaimus superbus	3
	Panagrolaimus davidi	1

Table 2 – Summary of *Deladenus siricidicola* EST Data

	Mycetophagous Library	Entomophagous Library
Total Bases	173,871,662	142,210,918
High Quality Reads	506,799	439,228
Average Read Length	344	329
Number of Contigs	16,444	3,828
Average Contig Length	566	446
Range Contig Length	50 to 3,265	50 to 1,738
Number of Singletons	8,301	3,379
Sequences with BLAST matches vs. NR database ($e<10^{-4}$)	8,624	2,258

Table 3 – Most abundant *D. siricidicola* mycetophagus female transcripts

Sequence ID	Length	# ESTs	Sequence Description	*Organism Identity*	Hit ACC	e-Value
contig04201	3057	902	abnormal nuclear anchorage family member (anc-1)	*Brugia malayi*	XP_001894426	1.15E-43
contig01083	4970	825	senescence-associated protein	*Brugia malayi*	XP_001900327	8.83E-51
contig00469	483	532	cytochrome c oxidase subunit iii	*Cooperia oncophora*	NP_851329	3.39E-48
contig00354	1042	465	phosphatidylcholine:ceramide cholinephosphotransferase	*Brugia malayi*	XP_001900386	7.60E-64
contig01348	1109	461	lin-45 raf	*Meloidogyne artiellia*	CAD56892	1.18E-113
contig07428	503	413	122 kda protein tmem16	*Brugia malayi*	XP_001896067	2.21E-26
contig01023	396	409	membrane-associated tyrosine- and threonine-specific cdc2-inhibitory kinase	*Caenorhabditis briggsae*	XP_001678102	1.22E-05
contig04220	617	399	hypothetical protein [Brugia malayi]	*Brugia malayi*	XP_001898879	1.05E-10
contig07596	1176	387	dead box atp-dependent rna helicase	*Brugia malayi*	XP_001893740	1.03E-123
contig00565	1298	368	inosine-5 -monophosphate dehydrogenase family protein	*Brugia malayi*	XP_001897693	3.26E-156
contig02419	1291	364	viral a-type inclusion protein	*Brugia malayi*	XP_001898330	6.01E-11
contig02545	626	360	zinc finger protein	*Schistosoma mansoni*	XP_002570059	4.01E-05
contig02498	957	351	short chain dehydrogenase reductase family protein	*Caenorhabditis elegans*	NP_505941	8.50E-27
contig03415	957	351	t-complex protein alpha subunit	*Brugia malayi*	XP_001900198	5.24E-85
contig04557	957	351	elegans protein partially confirmed by transcript evidence	*Caenorhabditis elegans*	NP_501256	2.91E-77
contig04619	957	351	btb poz domain containing protein	*Brugia malayi*	XP_001893330	1.39E-53
contig00959	678	341	cytochrome c oxidase subunit ii	*Toxocara vitulorum*	ACM88524	4.57E-56
contig01484	678	341	ccg1-interacting factor b	*Brugia malayi*	XP_001899361	5.32E-09
contig06271	1428	331	pg1 protein	*Lactobacillus jensenii*	ZP_04645459	7.32E-45
contig04091	1076	304	sr protein kinase family member (spk-1)	*Brugia malayi*	XP_001900384	1.25E-93
contig06409	1076	304	hypothetical protein Bm1_01425	*Brugia malayi*	XP_001891801	2.43E-04
contig06720	1076	304	hypothetical protein BRAFLDRAFT_85664	*Branchiostoma floridae*	XP_002223423	1.17E-14
contig02192	1170	300	intestinal prolyl carboxypeptidase 2	*Haemonchus contortus*	CAM84574	2.51E-13
contig00776	1524	293	cytochrome c oxidase subunit i	*Caenorhabditis briggsae*	ACB06132	4.52E-154
contig07906	520	293	elegans protein partially confirmed by transcript evidence	*Caenorhabditis briggsae*	XP_001667035	3.57E-17
contig07910	520	293	elegans protein partially confirmed by transcript evidence	*Caenorhabditis briggsae*	XP_001667342	1.31E-27
contig08121	520	293	prpf39 protein	*Brugia malayi*	XP_001902019	1.11E-10
contig08214	520	293	protein kinase domain containing protein	*Caenorhabditis briggsae*	XP_001679797	4.84E-06
contig01480	1358	289	set domain and mariner transposase partial	*Hydra magnipapillata*	XP_002170964	2.35E-21
contig00570	667	283	hypothetical protein C44C1.1	*Caenorhabditis elegans*	NP_508189	9.28E-46
contig01742	1259	279	hypothetical protein BRAFLDRAFT_92247	*Branchiostoma floridae*	XP_002229621	8.94E-04
contig01808	1259	279	elegans protein partially confirmed by transcript evidence	*Caenorhabditis elegans*	NP_001076643	2.19E-26
contig03569	1161	268	leucine-rich repeat containing protein	*Acyrthosiphon pisum*	XP_001947101	3.96E-11
contig00894	1734	262	ar protein family member (nol-5)	*Caenorhabditis elegans*	NP_491134	1.50E-143
contig09543	482	262	elegans protein partially confirmed by transcript evidence	*Caenorhabditis elegans*	NP_499688	1.65E-39
contig09593	482	262	vacuolar protein sorting 8 homolog	*Brugia malayi*	XP_001900441	1.11E-11
contig00135	326	260	atp synthase f0 subunit 6	*Caenorhabditis briggsae*	ACB06367	6.37E-10

Table 4 – Most abundant *D. siricidicola* entomophagous female transcripts

Sequence ID	Length	# EST	Description	Organism	hit accession	e-value
contig01470	1738	5357	rrna promoter binding protein	*Brugia malayi*	XP_001891902	1.50E-34
contig03785	926	3440	senescence-associated protein	*Brugia malayi*	XP_001900327	3.56E-52
contig00043	412	1454	hypothetical tyrosinase-like protein in chromosome	*Caenorhabditis briggsae*	XP_001678368	5.89E-21
contig03510	293	1310	predicted protein	*Nematostella vectensis*	XP_001623948	1.22E-05
contig01225	539	1147	elegans protein confirmed by transcript evidence	*Brugia malayi*	XP_001897850	1.21E-13
contig03771	499	1037	hypothetical protein F59A6.10	*Caenorhabditis elegans*	NP_001022216	7.57E-07
contig00100	1052	998	histone deacetylase 3	*Brugia malayi*	XP_001895111	9.76E-51
contig00258	465	922	piwi domain containing protein	*Brugia malayi*	XP_001901579	2.00E-08
contig03590	595	901	endothelin-converting enzyme 1	*Ancylostoma caninum*	AAG29105	3.61E-24
contig03749	518	817	potassium chloride cotransporter isoform a	*Caenorhabditis elegans*	ACN62949	3.26E-23
contig00493	557	788	short chain 3-hydroxyacyl-coa dehydrogenase	*Caenorhabditis briggsae*	XP_001667214	2.86E-08
contig01940	611	724	major sperm protein	*Ascaris suum*	AAP94885	3.84E-34
contig03515	579	722	elegans protein partially confirmed by transcript evidence	*Caenorhabditis elegans*	NP_001129807	3.84E-06
contig01218	472	707	elegans protein confirmed by transcript evidence	*Haemonchus contortus*	CAB40412	6.82E-09
contig00407	547	705	hyretin-related family domain family member (ttr-47)	*Caenorhabditis elegans*	NP_505304	1.27E-10
contig02020	903	699	briggsae cbr-frs-2 protein	*Caenorhabditis briggsae*	XP_001678661	4.74E-85
contig00742	552	687	cytoplasmic polyadenylation element binding protein 4	*Brugia malayi*	XP_001895028	2.32E-23
contig01905	548	674	hypothetical protein	*Brugia malayi*	XP_001897474	1.64E-13
contig03499	514	649	domain containing protein	*Caenorhabditis elegans*	AAB37835	2.58E-36
contig00146	211	635	aminopeptidase es-62 precursor	*Acanthocheilonema viteae*	AAC28365	1.01E-20
contig01681	546	629	transportin 3	*Brugia malayi*	XP_001900232	7.33E-22
contig01451	544	626	CG3579-PA	*Brugia malayi*	XP_001900825	1.85E-30
contig00092	529	622	26s proteasome non-atpase regulatory subunit 7	*Brugia malayi*	XP_001898659	8.79E-43
contig00753	511	603	briggsae cbr-tsn-1 protein	*Caenorhabditis briggsae*	XP_001678812	2.32E-10
contig03679	528	598	elegans protein partially confirmed by transcript evidence	*Brugia malayi*	XP_001902725	1.22E-52
contig00366	696	594	annexin family member (nex-2)	*Caenorhabditis briggsae*	XP_001664903	1.41E-15
contig03493	574	581	grim-19 protein	*Caenorhabditis elegans*	NP_492799	1.99E-39
contig00079	499	581	esophageal gland cell secretory protein 28	*Brugia malayi*	XP_001894809	2.38E-25
contig00458	563	580	protein disulfide isomerase family member (pdi-3)	*Caenorhabditis elegans*	NP_491995	1.28E-27
contig00213	490	563	pan domain containing protein	*Brugia malayi*	XP_001898702	1.06E-25
contig01164	544	559	elegans protein partially confirmed by transcript evidence	*Brugia malayi*	XP_001891860	2.02E-32

Table 5 – Transcripts expressed in the *D. siricidicola* entomophagous library that were not found in the mycetophagous library

Sequence name	# ESTs	Sequence desc.	length	Hit Identity	Hit ACC	E-Value	Database Hit
contig02051	341	reverse transcriptase	527	*Ascaris lumbricoides*	S60004	1.46E-21	nr
contig02334	161	unknown protein	483	*Globodera pallida*	GO251409	1.04E-91	NematodeEST
contig00821	102	solute carrier family 33 (acetyl- transporter) member 1	1116	*Caenorhabditis elegans*	NP_495969	1.52E-65	nr
contig01778	86	beta-lactamase family protein	655	*Brugia malayi*	XP_001892747	3.05E-30	nr, NematodeEST
contig00303	69	egf-like domain containing protein	1346	*Brugia malayi*	XP_001901879	2.64E-17	nr
contig02753	58	ke (drosophila actin-binding) homolog family member (ketn-1)	371	*Caenorhabditis elegans*	NP_503758	2.53E-19	nr, NematodeEST
contig02013	42	membrane associated guanylate ww and pdz domain containing 2	806	*Brugia malayi*	XP_001896933	2.37E-16	nr
contig01934	31	unknown gland cell protein	359	*Heterodera glycines*	CK349382	1.92E-18	NematodeEST
contig00384	24	transmembrane protein 68	384	*Caenorhabditis elegans*	NP_001023446	6.72E-49	nr, NematodeEST
contig02278	22	kinesin light chain 1 and 2	312	*Brugia malayi*	XP_001895440	1.27E-15	nr, NematodeEST
contig00096	17	mevalonate kinase	331	*Caenorhabditis briggsae*	XP_001666415	2.09E-42	nr, NematodeEST
contig03196	3	unknown protein	271	*Meloidogyne chitwoodi*	CB933565	2.11E-13	NematodeEST
contig03222	3	retrotransposon gag protein	278	*Caenorhabditis remanei*	DR779483	2.05E-11	NematodeEST
contig02466	2	unknown protein	447	*Caenorhabditis elegans*	BJ123622	7.56E-18	NematodeEST
contig02741	2	Glycogen phosphorylase	240	*Caenorhabditis elegans*	BJ135976	5.14E-22	NematodeEST
F0QM4P002GFZI9	1	hypothetical protein	455	*Caenorhabditis brenneri*	FF085908	1.85E-12	NematodeEST
F0QM4P002GJZ8B	1	Ribosomal Protein, Large subunit RPL-15	311	*Trichinella pseudospiralis*	FG580234	1.70E-28	NematodeEST
F0QM4P002GSMPX	1	unknown protein	441	*Ancylostoma caninum*	EX552706	2.75E-31	NematodeEST
F0QM4P002GZINR	1	Fructose-1 6-bisphosphatase	469	*Ascaris lumbricoides*	BU587064	3.44E-16	NematodeEST
F0QM4P002H2XWR	1	60s ribosomal protein l4	336	*Brugia malayi*	XP_001894323	8.14E-47	nr, NematodeEST
F0QM4P002HF7RW	1	Iron regulatory protein 1-like protein	314	*Strongyloides ratti*	BI073528	7.58E-13	NematodeEST
F0QM4P002HKSUD	1	Chaperonin	476	*Strongyloides stercoralis*	BE579280	1.51E-35	NematodeEST
F0QM4P002HSW3G	1	unknown protein	421	*Bursaphelenchus xylophilus*	CJ991992	5.11E-14	NematodeEST
F0QM4P002HYQD3	1	Cyclopropane-fatty-acyl-phospholipid synthase	465	*Litomosoides sigmodontis*	DN558176	1.98E-25	NematodeEST
F0QM4P002I0FMZ	1	tryptophanyl-trna synthetase	412	*Brugia malayi*	XP_001900656	5.51E-19	nr, NematodeEST
F0QM4P002I21MY	1	Tryptophanyl (W) tRNA Synthetase WRS-2 or CG7441-PA	505	*Ancylostoma caninum*	EX551131	3.44E-29	NematodeEST, nr
F0QM4P002I6Q44	1	Acetyl-coA carboxylase	401	*Ascaris suum*	BI781565	3.47E-14	NematodeEST
F0QM4P002IBRFL	1	Biotin synthase	437	*Heterodera glycines*	CB281758	4.24E-26	NematodeEST
F0QM4P002IENAY	1	unknown protein	412	*Ancylostoma caninum*	AW181737	7.35E-11	NematodeEST
F0QM4P002IHEMF	1	unknown protein	438	*Brugia malayi*	AA661216	2.92E-14	NematodeEST
F0QM4P002IN7WN	1	ras protein activator like 1 (gap1 like)	295	*Brugia malayi*	XP_001897948	1.21E-21	nr, NematodeEST
F0QM4P002IO3AS	1	Neurotransmitter-gated ion-channel subunit family member	389	*Heterorhabditis bacteriophora*	FG972612	2.25E-19	NematodeEST
F0QM4P002IOGJM	1	ubiquitin carboxyl-terminal hydrolase	504	*Trichuris vulpis*	CB188634	5.24E-11	NematodeEST
F0QM4P002IS5SK	1	Auxin response factor 8	173	*Meloidogyne hapla*	CN577507	4.09E-11	NematodeEST
F0QM4P002IYWO9	1	uncharacterized protein	378	*Osteragia osteragia*	BQ098901	1.24E-14	NematodeEST
F0QM4P002JN7FN	1	similar to CG10068 [*D. melanogaster]*	441	*Heterorhabditis bacteriophora*	FK808534	4.34E-15	NematodeEST
F0QM4P002JNSCP	1	unknown protein	467	*Globodera pallida*	GO251409	6.87E-69	NematodeEST
F0QM4P002JUV4U	1	unknown protein	275	*Heterorhabditis bacteriophora*	FF681843	7.93E-11	NematodeEST

Table 6 – Mycetophagous *D. siricidicola* transcripts similar to plant parasitic nematode secretory and putative parasitism genes

Sequence name	Sequence desc.	Organism	Hit ACC	E-Value
contig04657	14-3-3 product complete cds	*Meloidogyne incognita*	AF402309	9.37E-44
contig04658	14-3-3 product complete cds	*Meloidogyne incognita*	AF402309	3.14E-42
contig15225	14-3-3 product complete cds	*Meloidogyne incognita*	AF402309	9.71E-24
contig16203	14-3-3 product complete cds	*Meloidogyne incognita*	AF402309	6.18E-26
contig04402	annexin 4c10 complete cds	*Heterodera glycines*	AF469059	5.64E-25
F0QM4P001CYBTO	annexin 4c10 complete cds	*Heterodera glycines*	AF469059	4.20E-33
contig09864	calreticulin complete cds	*Meloidogyne incognita*	AF402771	1.38E-55
F0QM4P001DQ1YH	cellulase (eng4)	*Meloidogyne incognita*	AY422837	1.53E-26
contig08320	cellulase beta-1,4-endoglucanase (eng-1)	*Meloidogyne incognita*	AF100549	1.21E-48
contig12609	cellulase complete (eng4)	*Meloidogyne incognita*	AY422837	8.59E-40
contig07065	chitinase complete cds	*Heterodera glycines*	AF468679	1.18E-17
contig03346	esophageal gland cell secretory protein 2 complete cds	*Meloidogyne incognita*	AF531161	1.06E-21
contig00216	esophageal gland cell secretory protein 21 complete cds	*Meloidogyne incognita*	AY134440	1.81E-17
contig02136	esophageal gland cell secretory protein 21 complete cds	*Meloidogyne incognita*	AY134440	3.55E-23
contig03631	esophageal gland cell secretory protein 21 complete cds	*Meloidogyne incognita*	AY134440	3.88E-16
contig06494	esophageal gland cell secretory protein 21 complete cds	*Meloidogyne incognita*	AY134440	3.99E-36
contig07718	esophageal gland cell secretory protein 21 complete cds	*Meloidogyne incognita*	AY134440	3.06E-17
contig08885	esophageal gland cell secretory protein 21 complete cds	*Meloidogyne incognita*	AY134440	3.00E-40
contig09668	esophageal gland cell secretory protein 21 complete cds	*Meloidogyne incognita*	AY134440	9.95E-36
contig11633	esophageal gland cell secretory protein 21 complete cds	*Meloidogyne incognita*	AY134440	2.26E-20
F0QM4P001AHPAI	esophageal gland cell secretory protein 26 complete cds	*Meloidogyne incognita*	AY135362	2.18E-31
F0QM4P001ASHJ6	esophageal gland cell secretory protein 26 complete cds	*Meloidogyne incognita*	AY135362	9.17E-31
contig02278	esophageal gland cell secretory protein 28 complete cds	*Meloidogyne incognita*	AY135364	2.41E-43
contig12427	esophageal gland cell secretory protein 28 complete cds	*Meloidogyne incognita*	AY135364	4.44E-18
contig04403	hypothetical esophageal gland cell secretory protein 8 partial cds	*Heterodera glycines*	AF273735	8.60E-19
contig08923	hypothetical esophageal gland cell secretory protein 8 partial cds	*Heterodera glycines*	AF273735	1.55E-13
contig02949	mrna for annexin 2 (nex-2 gene)	*Globodera pallida*	AJ300178	2.47E-49
F0QM4P001B2YF1	mrna for annexin 2 (nex-2 gene)	*Globodera pallida*	AJ300178	2.64E-19
contig08263	mrna for annexin 2 (nex-2 gene)	*Globodera pallida*	AJ300178	3.77E-49
contig00952	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	6.72E-35
contig04816	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	1.44E-25
contig05355	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	3.62E-24
contig06212	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	8.35E-63
contig08795	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	7.25E-17
contig12494	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	2.18E-34
F0QM4P001DQUJF	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	3.54E-26
contig08793	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	2.05E-35
contig02007	mrna for peroxiredoxin (tpx gene)	*Globodera rostochiensis*	AJ243736	1.12E-107
contig06099	mrna for peroxiredoxin (tpx gene)	*Globodera rostochiensis*	AJ243736	3.37E-76
contig04432	mrna for sec-2 protein	*Globodera pallida*	Y09293	1.57E-11
contig05484	mrna for sec-2 protein	*Globodera pallida*	Y09293	1.16E-19
contig08210	mrna for sec-2 protein	*Globodera pallida*	Y09293	4.37E-38
contig02135	mrna for secreted glutathione peroxidase (gpx1 gene)	*Globodera rostochiensis*	AJ493677	3.23E-21
contig02204	mrna for secreted glutathione peroxidase (gpx1 gene)	*Globodera rostochiensis*	AJ493677	3.18E-21
contig06661	mrna for secreted glutathione peroxidase (gpx1 gene)	*Globodera rostochiensis*	AJ493677	1.67E-95
contig02414	myosin regulatory light chain complete cds	*Meloidogyne incognita*	AF402308	7.11E-35
contig01971	ubiquitin extension protein complete cds	*Heterodera glycines*	AF469060	1.71E-20
contig02442	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	1.09E-44
contig02995	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	1.04E-29
contig05320	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	1.04E-39
contig06253	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	8.68E-45
contig07784	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	9.44E-45
contig10828	ubiquitin extension protein complete cds	*Heterodera glycines*	AF469060	2.07E-20
contig12059	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	4.21E-42
contig16056	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	7.88E-44
contig15479	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	2.79E-40
F0QM4P001DDVNK	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	2.75E-26
contig09871	vap-1 complete cds	*Heterodera glycines*	AF374388	6.10E-16

Table 7 – Entomophagous *D. siricidicola* transcripts similar to plant parasitic nematode secretory and putative plant parasitism genes

Sequence name	Sequence Description	Organism	Hit ACC	E-Value
contig00284	14-3-3 product complete cds	*Meloidogyne incognita*	AF402309	3.33E-71
contig01512	14-3-3 product complete cds	*Meloidogyne incognita*	AF402309	6.86E-73
F0QM4P002ILBLY	14-3-3 product complete cds	*Meloidogyne incognita*	AF402309	5.87E-52
contig00600	calreticulin complete cds	*Meloidogyne incognita*	AF402771	1.94E-19
contig01783	cellulase complete cds	*Meloidogyne incognita*	AY422837	1.52E-19
contig01784	cellulase complete cds	*Meloidogyne incognita*	AY422837	8.68E-12
contig00507	esophageal gland cell secretory protein 21 complete cds	*Meloidogyne incognita*	AY134440	2.39E-36
contig00079	esophageal gland cell secretory protein 28 complete cds	*Meloidogyne incognita*	AY135364	2.72E-27
contig00366	mrna for annexin 2 (nex-2 gene)	*Globodera pallida*	AJ300178	2.59E-22
contig03585	mrna for annexin 2 (nex-2 gene)	*Globodera pallida*	AJ300178	2.40E-49
contig01431	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	3.45E-32
contig02917	mrna for cathepsin l protease (cpl-1 gene)	*Meloidogyne incognita*	AJ557572	7.64E-15
contig02116	mrna for peroxiredoxin (tpx gene)	*Globodera rostochiensis*	AJ243736	9.09E-91
F0QM4P002GPFZ9	mrna for peroxiredoxin (tpx gene)	*Globodera rostochiensis*	AJ243736	3.88E-23
contig00038	mrna for sec-2 protein	*Globodera pallida*	Y09293	7.36E-50
contig01032	mrna for secreted glutathione peroxidase (gpx1 gene)	*Globodera rostochiensis*	AJ493677	1.50E-68
F0QM4P002I5GGF	mrna for secreted glutathione peroxidase (gpx1 gene)	*Globodera rostochiensis*	AJ493677	7.22E-46
contig00479	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	2.04E-43
contig01080	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	1.23E-26
contig02227	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	7.69E-45
contig03171	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	4.58E-45
contig03521	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	1.10E-44
contig03799	ubiquitin extension protein complete cds	*Heterodera schachtii*	AY286305	2.00E-23
contig01190	ubiquitin extension protein complete cds	*Heterodera glycines*	AF469060	1.69E-20
contig02596	ubiquitin extension protein complete cds	*Heterodera glycines*	AF469060	1.59E-20

Figure 1

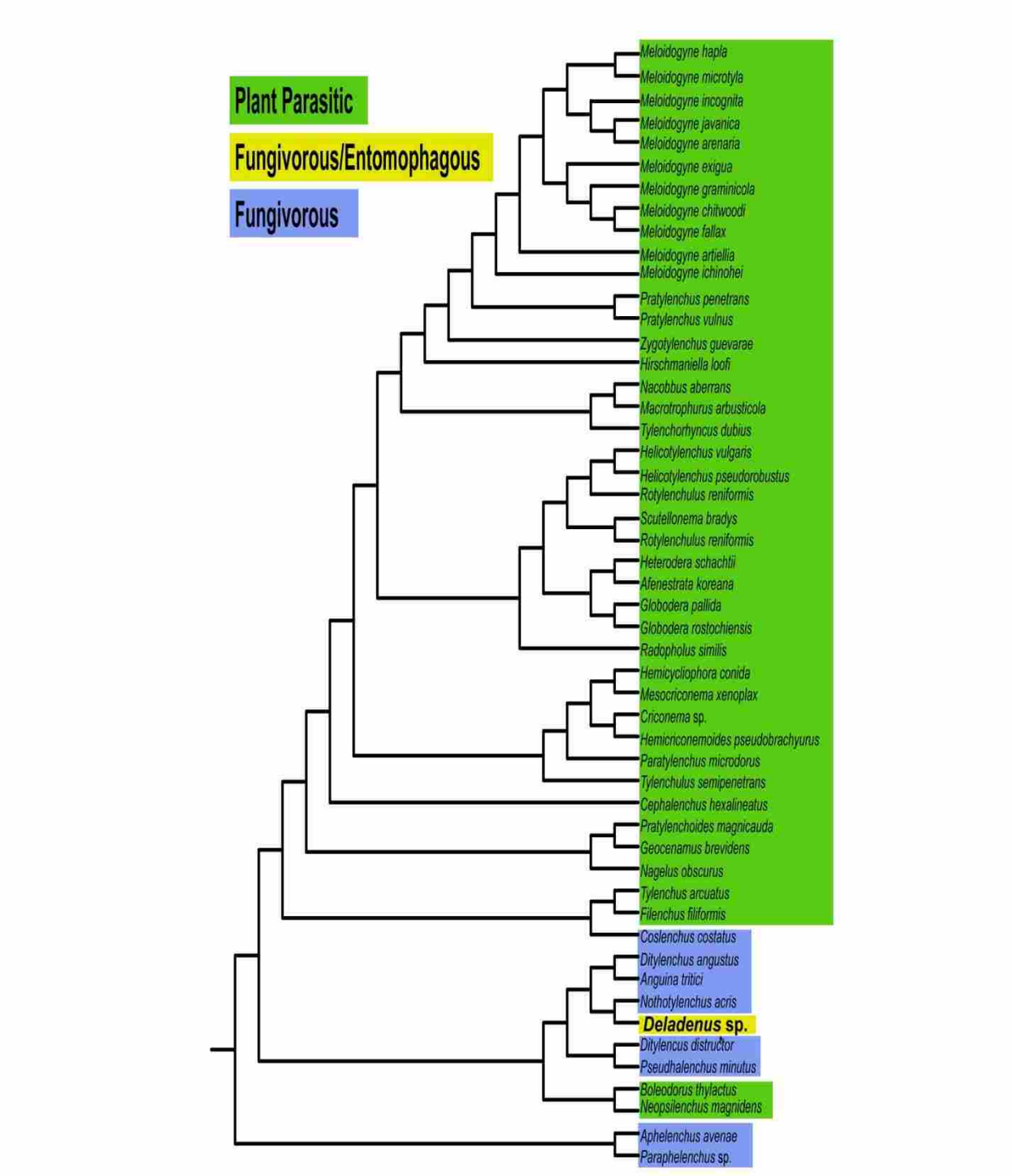

Plant Parasitic
Fungivorous/Entomophagous
Fungivorous
Meloidogyne hapla
Meloidogyne microtyla
Meloidogyne incognita
Meloidogyne javanica
Meloidogyne arenaria
Meloidogyne exigua
Meloidogyne graminicola
Meloidogyne chitwoodi
Meloidogyne fallax
Meloidogyne artiellia
Meloidogyne ichinohei
Pratylenchus penetrans
Pratylenchus vulnus
Zygotylenchus guevarae
Hirschmaniella loofi
Nacobbus aberrans
Macrotrophurus arbusticola
Tylenchorhyncus dubius
Helicotylenchus vulgaris
Helicotylenchus pseudorobustus
Rotylenchulus reniformis
Scutellonema bradys
Rotylenchulus reniformis
Heterodera schachtii
Afenestrata koreana
Globodera pallida
Globodera rostochiensis
Radopholus similis
Hemicycliophora conida
Mesocriconema xenoplax
Criconema sp.
Hemicriconemoides pseudobrachyurus
Paratylenchus microdorus
Tylenchulus semipenetrans
Cephalenchus hexalineatus
Pratylenchoides magnicauda
Geocenamus brevidens
Nagelus obscurus
Tylenchus arcuatus
Filenchus filiformis
Coslenchus costatus
Ditylenchus angustus
Anguina tritici
Nothotylenchus acris
Deladenus sp.
Ditylencus distructor
Pseudhalenchus minutus
Boleodorus thylactus
Neopsilenchus magnidens
Aphelenchus avenae
Paraphelenchus sp.

Figure 2

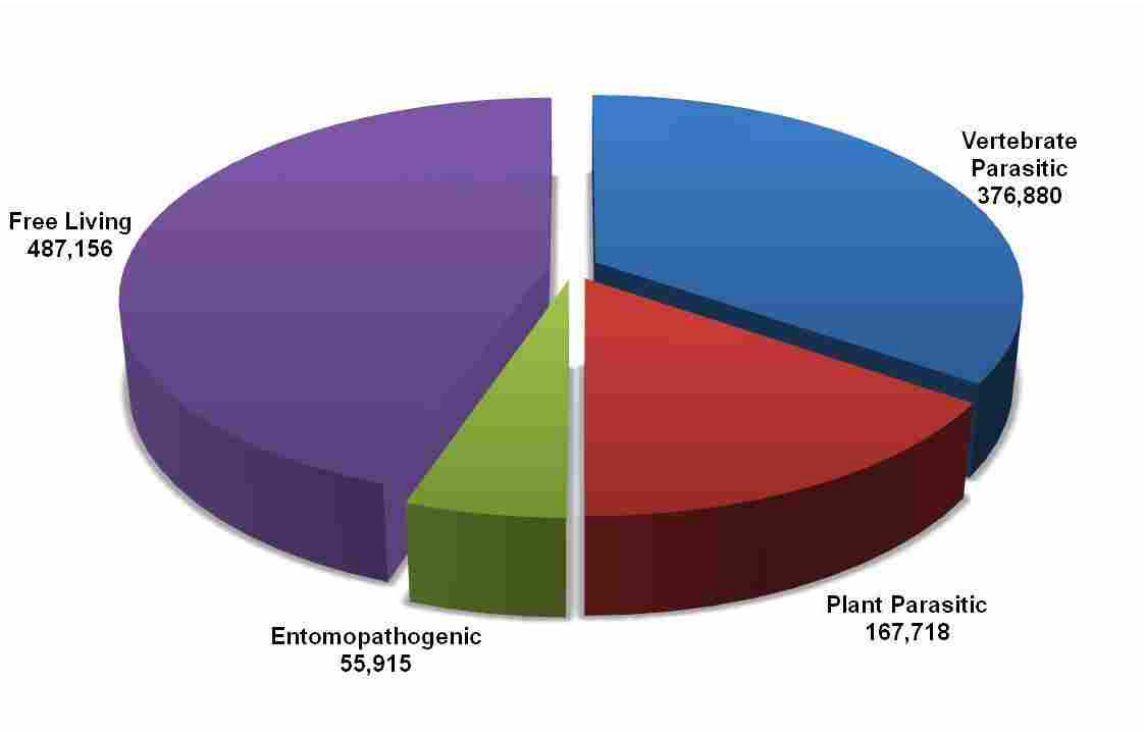

Vertebrate
Parasitic
376,880
Free Living
487,156
Plant Parasitic
167,718
Entomopathogenic
55,915

Figure 3

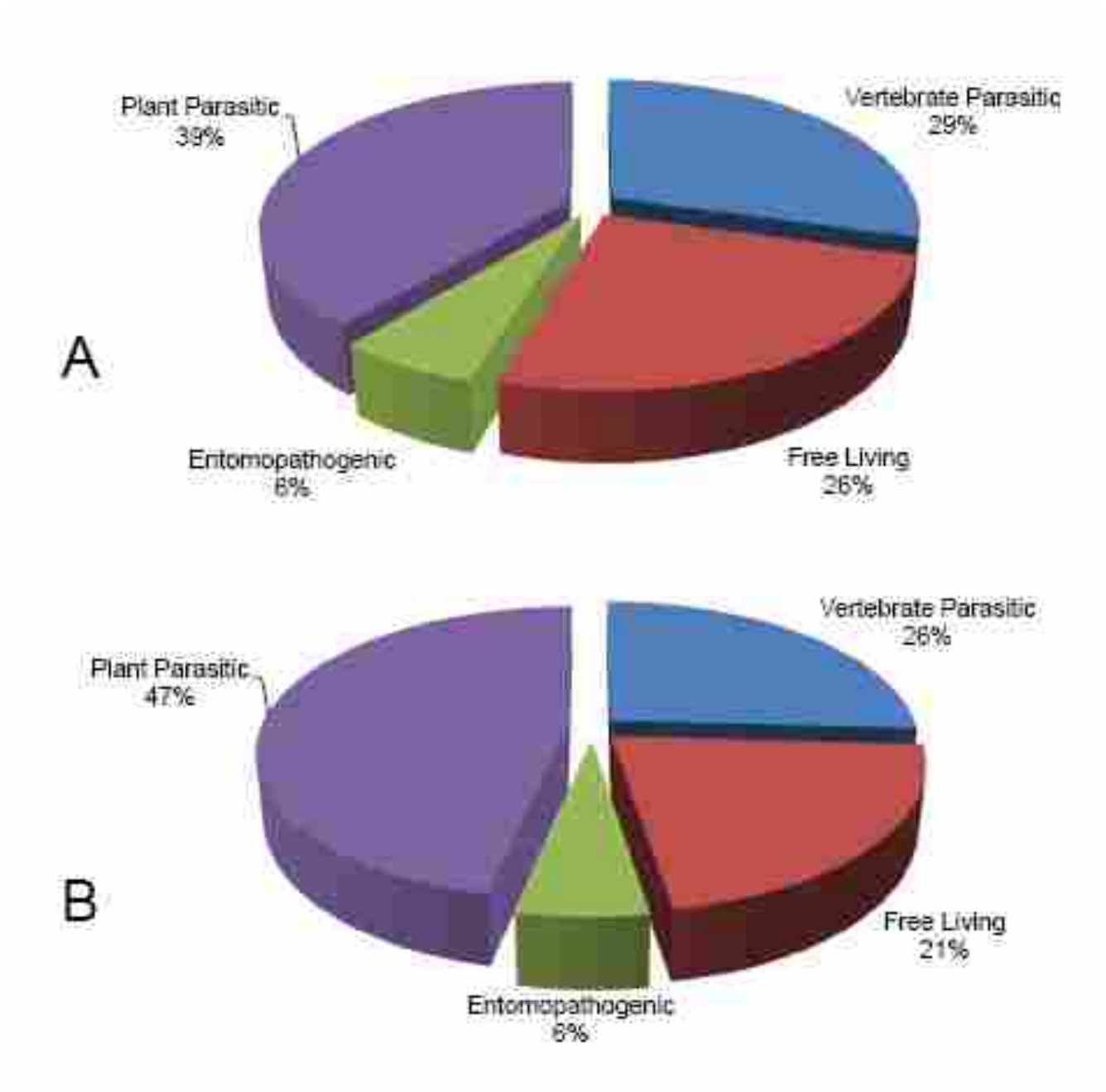

A
Plant Parasitic
39%
Vertebrate Parasitic
29%
Entomopathogenic
6%
Free Living
26%
B
Plant Parasitic
47%
Vertebrate Parasitic
26%
Free Living
21%
Entomopathogenic
6%

Figure 4

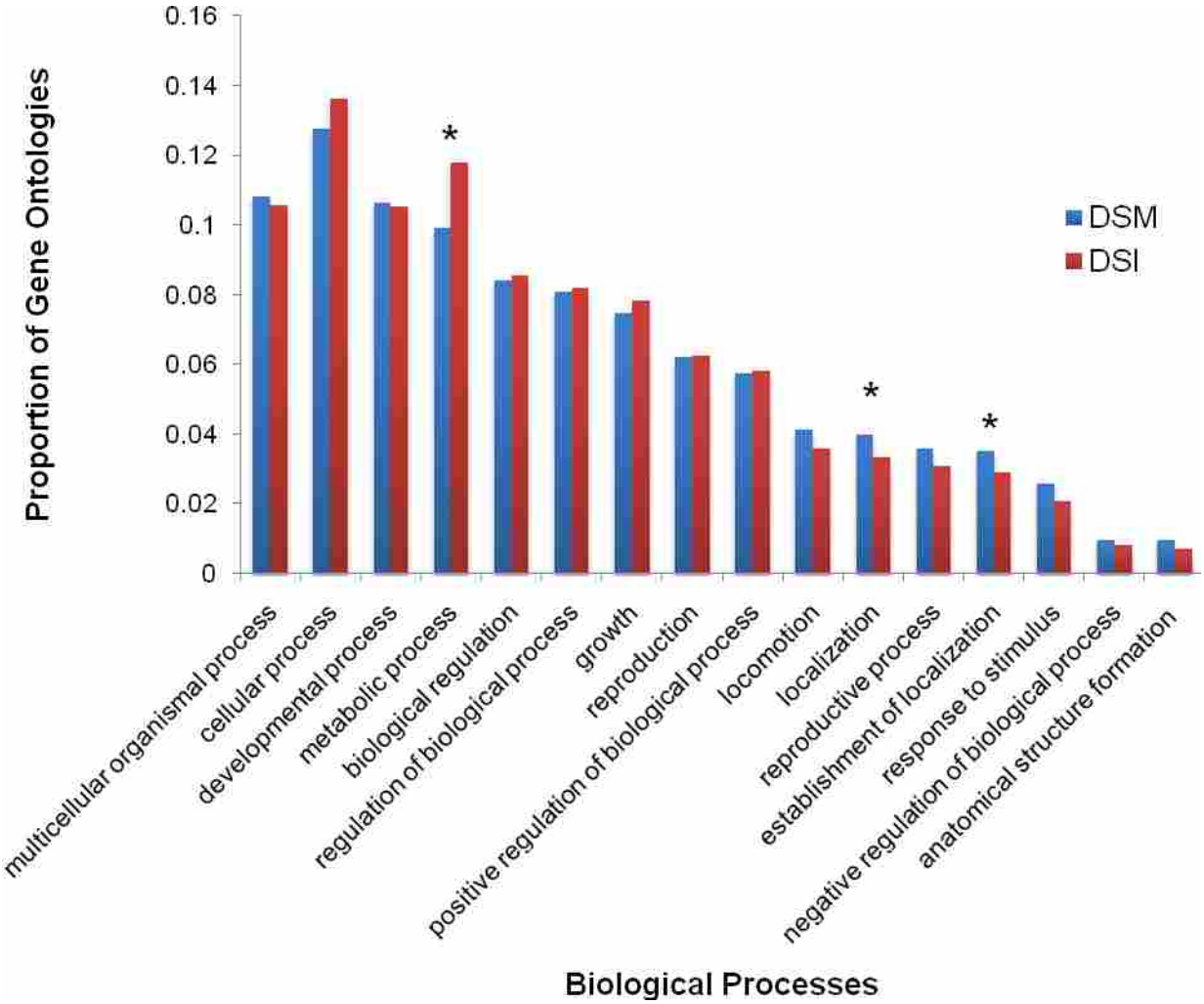

Proportion of Gene Ontologies
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
DSM
DSI
*
*
*
multicellular organismal process
cellular process
developmental process
metabolic process
biological regulation
regulation of biological process
growth
reproduction
positive regulation of biological process
locomotion
localization
reproductive process
establishment of localization
response to stimulus
negative regulation of biological process
anatomical structure formation
Biological Processes

Figure 5

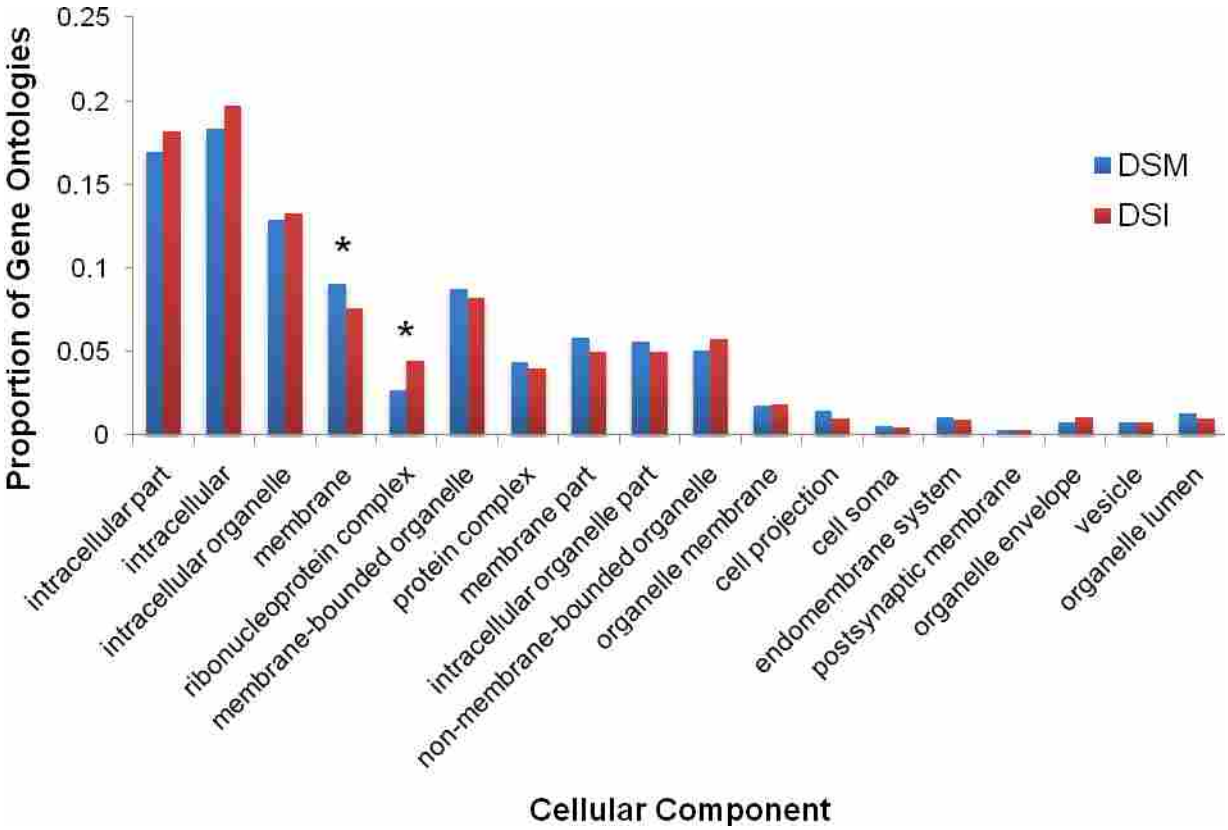

Proportion of Gene Ontologies
0.25
0.2
0.15
0.1
0.05
0
DSM
DSI
*
*
intracellular part
intracellular
intracellular organelle
membrane
ribonucleoprotein complex
membrane-bounded organelle
protein complex
membrane part
intracellular organelle part
non-membrane-bounded organelle
organelle membrane
cell projection
cell soma
endomembrane system
postsynaptic membrane
organelle envelope
vesicle
organelle lumen
Cellular Component

Figure 6

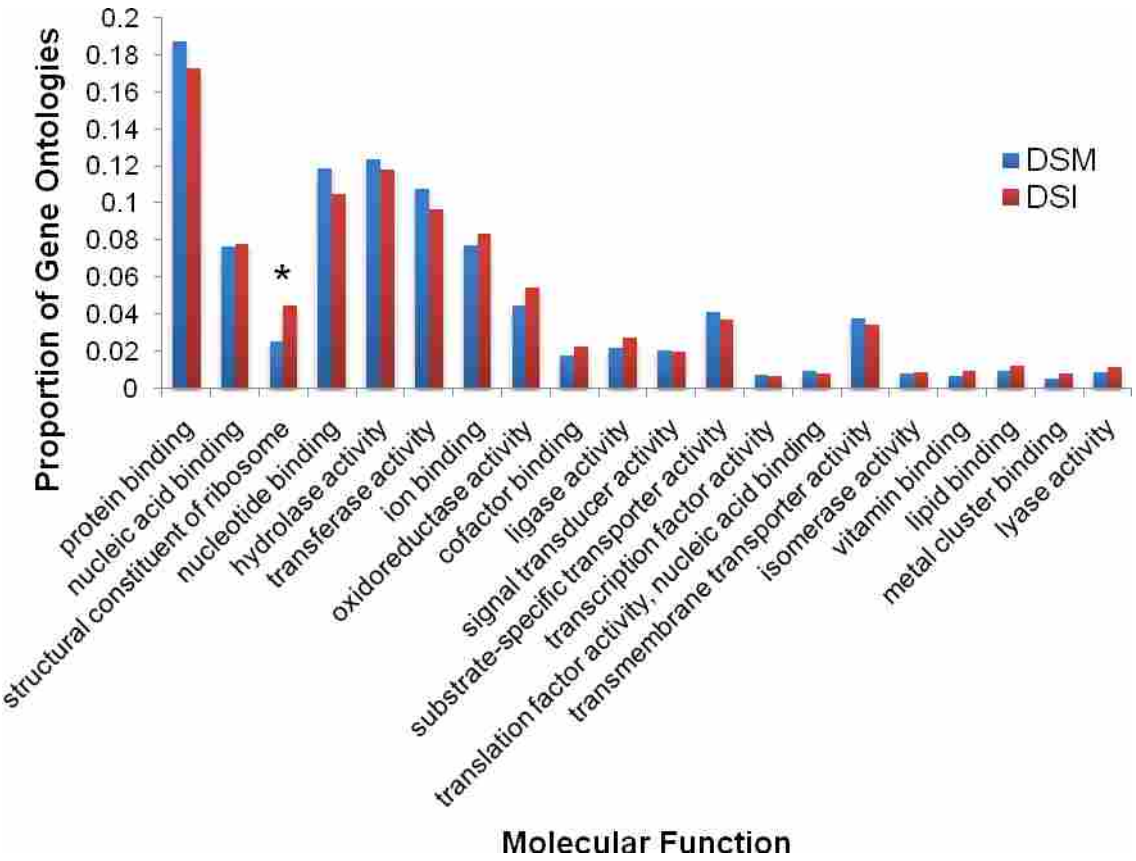
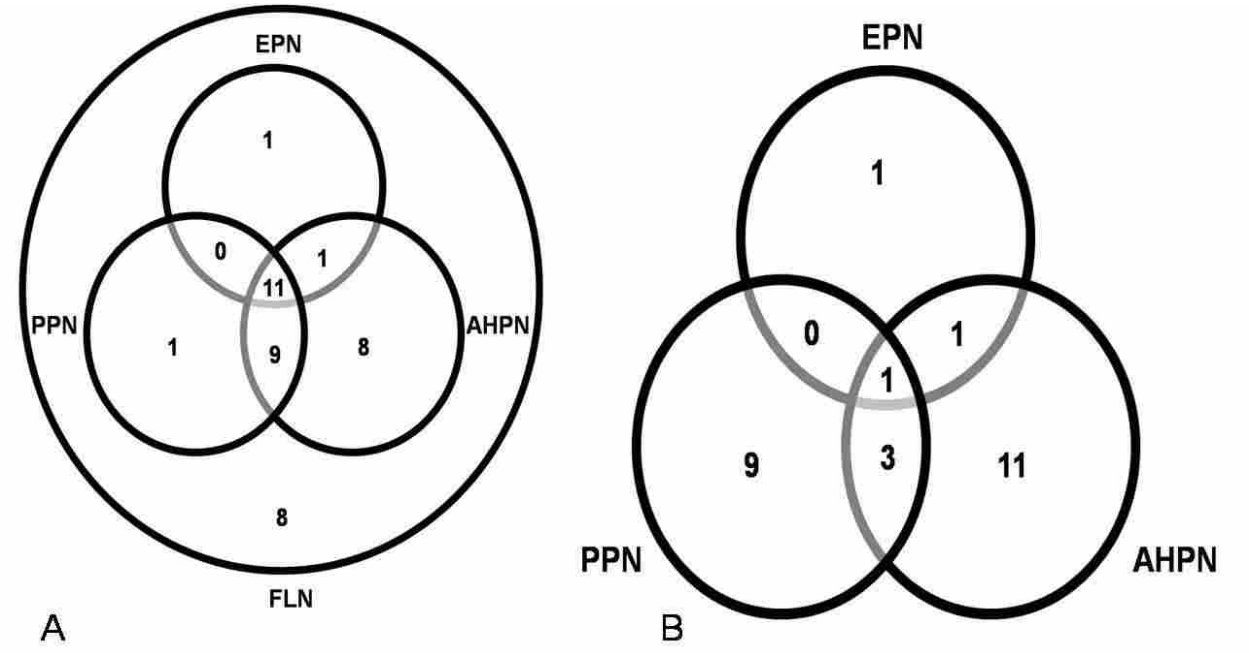

Proportion of Gene Ontologies
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
*
DSM
DSI
protein binding
nucleic acid binding
structural constituent of ribosome
nucleotide binding
hydrolase activity
transferase activity
ion binding
oxidoreductase activity
cofactor binding
ligase activity
signal transducer activity
substrate-specific transporter activity
transcription factor activity
translation factor activity, nucleic acid binding
transmembrane transporter activity
isomerase activity
vitamin binding
lipid binding
metal cluster binding
lyase activity
Molecular Function

Figure 7



Chapter 6

A Phylogenomic Analysis of the Nematode Infraorder Tylenchomorpha and a Framework for the Study of Parasitism Gene Evolution

Scott M. Peat, Tyler Collete, and Byron J. Adams

Department of Biology, Brigham Young University, Provo, Utah, 84602

Abstract

The nematode infraorder Tylenchomorpha contains fungal feeding nematodes, insect parasitic nematodes, and a large group of plant parasitic nematodes. Previous attempts to reconstruct the evolutionary history of this group of agriculturally important nematodes utilizing primarily ribosomal DNA data have produced phylogenies with poor resolution at the basal nodes of the tree. More rigorous analyses are needed to better understand the evolutionary relationships within this extremely important group of nematodes. As such, we conducted a phylogenomic analysis of the nematode infraorder Tylenchomorpha using expressed sequence tag data. Thirty orthologous datasets were selected using the program OrthoSelect, aligned, concatenated into a supermatrix, and Bayesian, likelihood, and parsimony phylogenetic analyses were conducted. Bayesian analysis provides strong support for Tylenchomorpha, Tylenchoidea, and Sphaerularioidea, though Aphelenchoidea appears paraphyletic. Parsimony analysis showed strong support for the placement of the outgroup *S. ratti* within the ingroup and as sister to *Meloidogyne*, though relative rates and long-branch extraction tests suggest that this may be an artifact of long-branch attraction.

Introduction

Nematodes are a relatively ancient group of organisms, with their origin believed to be sometime around the Cambrian or Precambrian period (Baldwin et al., 2004; Poinar et al., 2008). While most nematode diversity is represented by free-living nematodes in marine, soil, or freshwater environments, some of the more economically important nematodes are those that live a parasitic lifestyle. Nematodes parasitize a wide range of hosts from plants, to arthropods and vertebrates. The exact origin of the parasitic lifestyle in nematodes is unknown, though a fossil of *Cretacimermis libani* parasitizing adult midges in 135 million year old amber demonstrates that the animal parasitic lifestyle was around at least during the Cretaceous period (Poinar, 2003; Poinar et al., 1994). Like their animal counterparts, the origin of plant parasitic nematodes (PPNs) is unknown. A recent discovery of eggs, juveniles, and adults of an early Devonian nematode within the plant tissue of the early land plant *Aglaophyton major* suggests that nematodes had already formed associations with plants some 396 million years ago (Poinar et al., 2008)(Poinar et al., 2008). Paleontological evidence indicates that nematodes have co-inhabited earth with plants and other animals for well over 100 million years, and as such it is not surprising that the parasitic lifestyle in nematodes has arisen multiple independent times throughout the evolution of the phylum Nematoda (Blaxter et al., 2000; Blaxter et al., 1998; Dorris et al., 1999; Holterman et al., 2006).

Numerous genes found in plant parasitic nematodes are believed to play a role in plant parasitism (Davis et al., 2000; Huang et al., 2006; Huang et al., 2004; Huang et al., 2003). Currently, a popular area of research in nematode parasite management involves using RNAi to knockdown parasitism genes in plant parasitic nematodes by using genetically engineered plants. While these methods have shown promise (Huang et al., 2006), lack of knowledge of the evolutionary mode and tempo of parasitism genes in pest nematodes as well as the conservation of parasitism genes across taxa will greatly limit the success of RNAi in controlling nematodes. The reason for this is that a single mutation in the target gene of the nematode will virtually eliminate the effectiveness of the genetically engineered plant to inhibit parasitism by the nematode. As such, researchers need to be able to identify parasitism genes that evolve under extremely strong selection, in order to develop RNAi gene targets that will allow for long term host resistance and thus provide a more cost efficient and effective solution for the control of parasitic nematodes. Given the specificity of the RNAi pathways in nematodes and knowledge of selection on parasitism genes, we would be in a position to model/experiment predictions regarding the evolution of resistance to engineered RNAi-based resistance plants.

Tylenchomorpha is an infraorder of nematodes that contains free living fungal feeding nematodes, insect parasitic nematodes, and a large group of plant parasitic nematodes. Multiple attempts have been made to resolve relationships within the Tylenchomorpha using primarily ribosomal RNA (rRNA) data (Bert et al., 2008; Holterman et al., 2009; Subbotin et al., 2006). While these studies have shown strong support for terminal clades, many of the deeper nodes remain poorly supported and/or unresolved, and more genetic data from different loci are needed to fully resolve the relationships within Tylenchomorpha. Expressed sequence tag data provides a vast, largely untapped resource to conduct a large scale phylogenomic analysis of the

Tylenchomorpha, with 165,101 EST sequences available on GenBank for 20 Tylenchomorpha taxa (18 of these taxa have at least 1,900 ESTs available). Additionally, Peat and Adams (2010; in review) have generated the first EST dataset for an insect parasitic member of the Tylenchomorpha that contains over 25,000 ESTs. As such, a phylogenomic analysis of the Tylenchomorpha utilizing available EST data should prove useful in resolving clades that to date are poorly resolved/supported based on rRNA data alone.

When conducting phylogenomic and gene evolution studies, it is imperative that the genes being compared are orthologous (evolved from a common ancestor) and not paralogous (related due to duplication events) (Fitch, 1970, 2000; Li et al., 2003). Once duplicated, paralogous genes may be subjected to differing evolutionary constraints and perform their own biologically distinct functions (Dolinski and Botstein, 2007; Koonin, 2005). As such, when conducting phylogenetic analyses or molecular evolutionary analyses on a gene across multiple species, comparison of genes that have evolved from a common ancestor (i.e. orthologs) is necessary in order to accurately identify phylogenetic relationships, discover regions under selection, and infer rates of molecular evolution. Thus, initial steps need to be taken to ensure that gene orthologs are identified before further analyses of orthologs are conducted. Multiple methods exist for identifying orthologs including phylogeny based programs such as RIO (Zmasek and Eddy, 2002)and LOFT (van der Heijden et al., 2007), reciprocal best blast hit (RBH) methods such as those implemented in OrthoMCL (Li et al., 2003) and InParanoid (Remm et al., 2001), a profile hidden markov model based search method as implemented in HaMStR (Ebersberger et al., 2009), and OrthoSelect (Schreiber et al., 2009), a method that utilizes predefined orthologous groups to conduct preliminary ortholog assignment, followed by additional refinement steps to eliminate redundancies (i.e. duplicate sequences, paralogs, etc.).

Best blast hit methods are best applied toward whole genome data rather than EST data due problems with length of genes and incomplete representation of species' gene set in EST datasets (Ebersberger et al., 2009). Many tree based methods rely on knowledge of a species tree, require the selection of an appropriate outgroup species, and depend on pre-defined protein families (Kuzniar et al., 2008), all of which are not always available.

While the evolution of the parasitic lifestyle in nematodes has been addressed in numerous studies (Baldwin et al., 2004; Blaxter et al., 2000; Blaxter et al., 1998; Dorris et al., 1999; Holterman et al., 2006), little data exists addressing the origin of specific genes involved in parasitism and how these genes are maintained within individual clades containing parasitic nematodes. Within the Tylenchomorpha, Bert et al. (2008) suggests that fungal feeding is the most likely ancestral feeding state, with plant parasitism and insect parasitism evolving later. Many plant parasitism genes are believed to have been transferred via horizontal gene transfer from bacteria and fungi (Bird et al., 2003; Jones et al., 2005), though the exact evolutionary origin of these genes within the infraorder Tylenchomorpha is unknown. Were these genes acquired prior to the divergence of the insect and plant parasitic life histories? If the answer to this question is yes and these genes no longer serve a function in insect parasites, the selective constraints on these plant parasitism genes should be relaxed, and as such the rate of mutations in the insect parasite should be much higher. If the plant parasitism genes still serve a similar function in the insect parasite, evolutionary constraints should be similar in both the insect and the plant parasites, and as such these genes should have similar mutation rates. As such, Tylenchomorpha provides an excellent model system to explore the origin and maintenance of plant parasitism genes and to investigate what factors (ecological, biological, life history) are associated with the levels of selection that are exerted onto parasitism genes, though a well

resolved phylogenetic hypothesis is required before analyses of parasitism gene evolution can be undertaken.

To this end, we utilized a phylogenomic approach to construct a phylogenetic hypothesis for the Tylenchomorpha using 30 genetic datasets extracted from EST data from 19 Tylenchomorpha taxa.

Methods

Identification of Orthologous Genes from Tylenchomorpha Nematode EST Datasets

Identification of orthologous genes for the phylogenomic analysis of Tylenchomorpha was conducted for 19 Tylenchomorpha EST datasets (Table 1) using the program OrthoSelect (Schreiber et al., 2009). Orthology searches in OrthoSelect were conducted against the eukaryotic clusters of orthologous genes database (KOG) (Tatusov et al., 2003). Statistics on the resulting clusters were summarized and orthologous groups containing sequence data from at least 16 of the 19 Tylenchomorpha taxa were carried on further through the OrthoSelect pipeline to remove redundancies and paralogs. Following removal of redundant sequences, each candidate amino acid dataset proposed by OrthoSelect was aligned in Muscle to check for falsely identified orthologs and non-overlapping sequences. Poorly aligned and/or anomalous sequences were checked for gene identity and location within the gene using NCBI's BlastX. Problem sequences were either removed from the alignment or if many sequences appeared to be problematic, the whole dataset was removed from the group of candidate datasets.

Alignment and Phylogenomic Analysis

Orthologous nucleotide sequences (selected by orthoselect and confirmed via inspection of preliminary amino acid alignments) for each dataset were trimmed to the reading frame predicted by ESTScan (Iseli et al., 1999)(as part of the OrthoSelect pipeline). Nucleotide sequences were converted into amino acids using AlignmentHelper (http://www.bigelow.org/research/facilities/srs_laboratories/david_mcclellan_laboratory/alignmenthelper/), aligned using Muscle (Edgar, 2004), and then re-translated back into nucleic acids using AlignmentHelper. Following alignment, each dataset was concatenated into a super matrix in MacClade 4.05 (Maddison and Maddison, 2002).

Parsimony analyses were conducted in TNT (Goloboff et al., 2008) under the new technology search with drift, ratchet, and pruning and 1000 random addition sequences. A strict consensus tree was assembled for the multiple most parsimonious (MP) trees recovered from the heuristic searches. The program TreeRot v3 (Sorenson and Franzosa, 2007) was used to calculate partitioned Bremer support values, which were mapped onto the strict consensus tree to assess the contribution each dataset made to the overall topology. Bootstrap analyses were conducted in TNT using 1000 replications.

For model based phylogenetic analyses, best fit models of evolution were calculated for each of the 30 genes (table 2) using ModelTest 3.1 (Posada and Crandall, 1998) under the AIC selection criterion. Mixed model Bayesian analyses were conducted in MrBayes 3.08 (Ronquist and Huelsenbeck, 2003). Two runs were conducted for each dataset using 10,000,000 generations sampled every 1000 generations. Stationarity was estimated using Tracer v1.4 (Rambaut and Drummond, 2007), with the 30 gene supermatrix having a burn-in value of 40,000. Posterior probability values were generated using the "sumt" command in MrBayes.

Log likelihood values for each run were compared to ensure that each Bayesian run converged on similar log likelihood mean values for each of the two independent runs for each gene.

Maximum likelihood analyses were conducted in RAxML (Stamatakis, 2006) using the GTRGAMMAI model with partitioning by gene. Bootstrap values for the likelihood tree were calculated in RAxML using 1000 bootstrap replicates.

Identification of Orthologous plant parasitism genes

Parasitism gene candidates were selected based on genes that have been identified as potential parasitism genes in previous studies (Davis et al., 2000; Huang et al., 2006; Huang et al., 2004; Huang et al., 2003; Hussey et al., 2002; Jones et al., 2005; Ledger et al., 2006). Parasitism gene datasets were assembled from GenBank's EST and non-redundant databases using BLAST searches. These datasets were run through OrthoSelect (Schreiber et al., 2009) to identify orthologous sequences. As most plant parasitism genes are believed to have arisen via horizontal gene transfer from bacteria, the assignment of orthologous groups step in OrthoSelect was conducted with both the prokaryotic clusters of orthologous groups database (COG) and the KOG (Tatusov et al., 2003). One of the candidate parasitism gene datasets, a GHF5 cellulase gene (beta-1,4 endoglucanase 1 (eng1)) was utilized in molecular evolutionary based analyses.

Rates of Evolution in eng1

Orthologs for the putative plant parasitism gene beta 1,4-endoglucanase 1 (eng1) were used to construct a phylogeny using TNT for taxa with available gene sequence data . This data was then utilized to test for differential rates of diversification in the eng1 gene of Tylenchomorpha nematodes using the program SymmeTree (Chan and Moore, 2005).

Identification of Selection in Parasitism Genes

Analyses of selection on the eng1 gene of Tylenchomorpha nematodes was conducted in HYPHY (Pond et al., 2005), DataMonkey (Pond and Frost, 2005), and TreeSAAP (Woolley et al., 2003) using relationships inferred from the phylogenomic analyses (Bayesian and likelihood tree with *Strongyloides ratti* removed) in the present study. Analyses of individual codon sites under positive and negative selection were analyzed in DataMonkey using the SLAC, FEL, and REL methods. Additionally, analysis of overall selection within the dataset was conducted using the PARRIS method as implemented in DataMonkey.

TreeSAAP analyses were conducted using a sliding window of 15. Positive selection (categories 6, 7, and 8) was mapped onto branches of the phylogenetic hypothesis of Tylenchomorpha (from the phylogenomic analysis), to assess patterns of selection across the topology.

Results and Discussion

Orthology Assignment

From the OrthoSelect analysis of all 19 Tylenchomorpha EST datasets, 2 genes had representative sequence from all 19 taxa, 10 genes had representative sequence from 18 taxa, 25 genes had representatives from 17 taxa, and 43 genes had representatives from 16 taxa. From these 80 orthologous dataset candidates, 30 datasets (table 2) were selected for use in the phylogenomic analysis of Tylechomorpha following preliminary alignment of datasets. Upon preliminary alignment of each candidate dataset, multiple candidate datasets appeared to possess numerous paralogous sequences. We believe that in many of these cases, OrthoSelect may select a paralog for inclusion into a dataset when no ortholog is present. As such, close evaluation of

each dataset output from OrthoSelect is suggested to aid in utilizing datasets with limited to no paralogous data.

Phylogenetic Analyses

The nucleotide supermatrix used in the present analyses contained 19 ingroup taxa and two outgroup taxa, *Strongyloides ratti* and *Caenorhabditis briggsae*. Outgroups were selected based on availability of genomic/expressed sequence tag data as well as relationships recovered in a previous phylogenetic data based on 18S ribosomal DNA (Bert et al., 2008). Outgroup sequence data was obtained from GenBank and from the Sanger Trust Institute's *Strongyloides* (http://www.sanger.ac.uk/sequencing/Strongyloides/) and *Caenorhabditis* (http://www.sanger.ac.uk/Projects/C_briggsae/) sequencing project websites. Of the 19,116 characters and 401,436 total nucleotides included in the supermatrix, 50.4% of the data was coded as missing. A breakdown of the amount of data coded as missing for each taxon is shown in table 3.

Bayesian analysis (Figure 1) shows strong support (PP = 1.00) for a number of previously proposed Tylenchomorpha superfamilies (De Ley and Blaxter, 2002), including the monophyly of Tylenchoidea (*Meloidogyne, Pratylenchus, Globodera, Rotylenchus, Heterodera, and Radopholus*), the monophyly of Tylenchoidea with *Deladenus siricidicola + Ditylenchus africanus+ Aphelencus avenae*, the monophyly of Sphaerularioidea, and the monophyly of Tylenchomorpha. Bayesian analysis does not support the monophyly of Aphelenchoidea, as *Bursaphelenchus* spp. and *A. avenae* do not for a monophyletic group. *Bursaphelenchus* appears to be the most basal Tylenchomorpha taxon in the present analysis, while *Aphelenchus* forms a

poorly supported (PP = 0.69) monophyletic group with *Deladenus siricidicola* and *Ditylenchus africanus*.

Likelihood analysis (Figure 2) recovered similar relationships, though with *A. avenae* as sister to the monophyletic Tylenchoidea rather than to *D. siricidicola* + *D. africanus*. Likelihood bootstrap analyses were in disagreement with the best likelihood tree, with the placement of *S. ratti* within the ingroup being supported in the bootstrap analysis but not showing up in the best likelihood tree. As such, bootstrap support values are only shown for some of the terminal relationships and no support is given to the monophyletic grouping of the Tylenchomorpha in the best likelihood tree. This discrepancy between the best likelihood tree and the bootstrap analysis may be due to either missing data or it could be a combination of long branch attraction artifact and the way that bootstraps are calculated in RaxML.

Parsimony analysis of the Tylenchomorpha supermatrix resulted in two equally parsimonious trees with a score of 25,528. The strict consensus of the two most parsimonious trees (figure 3) shows *S. ratti* (one of the outgroups) forming a monophyletic group with the *Meloidogyne* clade. Similar to both the Bayesian and likelihood trees, *Bursaphelenchus* is basal to all other Tylenchomorpha taxa and Aphelenchoidea (*A. avenae*, *B. mucronatus*, and *B. xylophilus*) is not monophyletic. The placement of *S. ratti* within the ingroup, and more specifically as sister taxon to the *Meloidogyne* clade was a bit puzzling. Initially, it was believed that missing data may be playing a role in this grouping, and that the low missing data percentages for *S. ratti* and many of the *Meloidogyne* spp. may be contributing to the artificial grouping of the two, though we believe this is unlikely as both *R. similis* and *D. siricidicola* have lower missing data percentages than all of the *Meloidogyne* spp. used in this study (Table 3).

Additional Bayesian (figure 5) and likelihood (figure 6) analyses following removal of *S. ratti* show strong support for the monophyly of Tylenchoidea and Sphaerularioidea, as well as for the placement of *Aphelenchus* as sister to Tylenchoidea +Sphaerularioidea. Parsimony analysis following removal of *S. ratti* (figure 7) does not support the monophyly of Sphaerularioidea. Both our Bayesian and likelihood trees are in agreement with the 18S rRNA tree of Bert et al. (2008) and Holterman et al. (2009) on the monophyly of *Meloidogyne* and *Pratylenchus* as well as the monophyly of the Tylenchoidea. Anomolous placement of *S. ratti* within the ingroup in parsimony analyses seems to be alleviated by removing *S. ratti* from the analyses, though disagreement between the model and parsimony based tree reconstruction methods still exist regarding the placement of *A. avenae, D. siricidicola,* and *D. africanus*.

Partitioned Bremer Support

Partitioned bremer support analysis the 30 gene supermatrix (Table 4) was conducted, in an attempt to identify possible sources causing the grouping of *S. ratti* with *Meloidogyne* spp. Based on the partitioned bremer results, numerous genes were found that supported the grouping of *S. ratti* with *Meloidogyne* (node 6 of the parsimony tree (Figure 3)) as well as the node supporting *S. ratti* + Tylenchoidea (node 12). These genes include all of the ribosomal protein genes, troponin C, cyclophilin 3, alpha tubulin, translationally controlled tumor protein, Dynein light chain 1, and arginine kinase. To further explore the effect these genes were having on the phylogeny, genes supporting nodes 6 were removed one-by-one, starting with the highest positive bremer support value. Following the removal of seven genes (six ribosomal protein genes and 2-cystein peroxiredoxin), *S. ratti* became the sister taxon to *Meloidogyne* + *Pratylenchus penetrans*, though *P. vulnus* became the sister taxon to the *Globodera, Heterodera,*

Rotylenchus, and *Radopholus* clade. Following the removal of seven additional genes that provided the highest remaining support to node 12, two monophyletic groups formed. Clade one consisted of *Meloidogyne* spp. + *P. penetrans* + *S. ratti* while clade two was composed of all other taxa. From these dataset removal experiments, we were unable to break up the grouping of *Meloidogyne*, *P. penetrans*, and *S. ratti*. As such, it appears that the cause of the *S. ratti*/*Meloidogyne* problem is likely not a result of one or a few anomalous datasets.

Long Branch Attraction

An alternative explanation for the disagreement between tree reconstruction methodologies and the formation of a *Meloidogyne*/*S. ratti* clade is that of "long branch attraction (LBA)", the grouping of long branches in a phylogenetic tree based on methodological artifacts (Bergsten, 2005). An examination of branch lengths in the likelihood tree reveals the presence of long branches leading to *S. ratti*, *P. penetrans*, as well as leading to the *Meloidogyne* clade. As such, the artificial attraction of *S. ratti* to *Meloidogyne* spp. could very well be due to similarly high rates of evolution.

To test the hypothesis of similar rates of evolution, a relative rates test was conducted in the software program HyPhy (Pond et al., 2005). Results from the relative rates tests with all possible pairwise comparisons from the Tylenchomorpha dataset show a lack of significant difference (alpha = 0.001) between *M. hapla* and *S. ratti* (p = 0.0020), *M. incognita* and *S. ratti* (p = 0.0029), and *P. penetrans* and *S. ratti* (0.0011). Furthermore, no other taxa in the analyses showed significant similarity in relative rates of evolution with *S. ratti.* As such, this provides preliminary evidence that evolutionary rates of *S. ratti* and some *Meloidogyne* spp. may be playing a role in the artificial grouping of these taxa.

To further investigate the role of long-branch attraction in the grouping of *S. ratti* and *Meloidogyne* spp., analyses were conducted to investigate the effects of taxa removal on the final tree topology. Siddall and Whiting (1999)posited that if each of the two branches (believed to be attracted to their position in the tree via LBA) individually group in the same location within the phylogenetic hypothesis when the other branch has been removed from the analysis, than LBA is not the cause of the grouping. Thus, if *Meloidogyne* and *S. ratti* each forms a monophyletic group with *Pratylenchus* spp. when the other is removed from the analysis, then LBA is likely not the cause of the grouping. With this in mind, I employed the long-branch extraction (Siddall and Whiting, 1999) test to check for LBA in the Tylenchomorpha dataset. Upon removal of *S. ratti* from the dataset, the resulting tree topology (Figure 4a) looked similar to the Bayesian and likelihood trees, with Tylenchoidea forming a monophyletic group and *Meloidogyne* spp. and *Pratylenchus* spp. forming a monophyletic group. A lack of monophyly for the Tylenchoidea exists when *C. briggsae* is removed from the analysis and the tree is rooted with *S. ratti* (Figure 4b). When all *Meloidogyne* spp. are removed from the analysis and the tree is rooted with *C. briggsae* (figure 4c), Tylenchomorpha and Tylenchoidea are monophyletic, and *S. ratti* does not form a monophyletic group with *Pratylenchus* spp. but instead is basal to the Tylenchomorpha. Additionally, when a single *Meloidogyne* sp. is added to the analysis (tree not shown), *S. ratti* always forms a monophyletic group with that *Meloidogyne* sp., Tylenchomorpha is paraphyletic, and Tylenchoidea is paraphyletic. Finally, when both outgroups are removed and the tree is rooted with *Bursaphelenchus* (figure 4d), Tylenchoidea is monophyletic. As such, figures 4a and 4c illustrate that based on the long-branch extraction test, the grouping of *S. ratti* and *Meloidogyne* in the parsimony analysis are likely attributed to LBA, as *S. ratti* is only drawn into the ingroup when at least one of the *Meloidogyne* spp. are present.

Future analyses on this dataset will focus on alleviating this long branch attraction issue to enable a better understanding of the relationships within the Tylenchomorpha. The inclusion of additional taxa has been suggested as a method for avoiding LBA (Bergsten, 2005; Hendy and Penny, 1989; Hillis, 1996, 1998), though while addition of more ingroup taxa will likely not be feasible until new Tylenchomorpha datasets become available, addition of more outgroup taxa is a possibility and will be explored. Adding additional data to the analysis is another possible solution that has been suggested (Bergsten, 2005; Xiang et al., 2002), though the addition of linked genes is not recommended (Rokas et al., 2003). As such, future analyses will be conducted in OrthoSelect to obtain more orthologous datasets.

Missing data

While a large portion of the current dataset is composed of missing data, Weins and Moens (Wiens and Moen, 2008)have suggested that as long as the number of characters in a given dataset is large, incomplete taxa (taxa that are missing large amounts of data) should still be able to accurately be placed in the tree. This is well illustrated in the present analysis by the fact that *R. reniformis* is consistently placed in a clade with *Heterodera* spp. and *Globodera* spp., while having approximately 93% missing data. Additionally, it should be noted that all three methods of tree reconstruction used in the present study (Bayesian, likelihood, and parsimony) produced differing tree topologies. Much of this can be attributed to the presence of long branch attraction between *Meloidogyne* and *S. ratti* though the inconsistent placement of *A. avenae* within all three phylogenies may be attributed to the differential treatment of missing data by different tree reconstruction methods, though further analyses are needed to confirm this hypothesis.

Molecular evolution of eng1

Analysis of *eng1* diversification rates across Tylenchomorpha taxa using SymmeTree (Chan and Moore, 2005) showed no significant difference in diversification rates. Indicating similar rates of evolution for this gene across diverse Tylenchomorpha taxa. Selection based analyses conducted in HYPHY (Pond et al., 2005), and DataMonkey (Pond and Frost, 2005) indicate a lack of positive selection across the entire *eng1* dataset as well as at individual codon positions. Analysis of selection in TreeSAAP (Woolley et al., 2003) showed the presence of statistically significant positive destabilizing selection at numerous branches within the Tylechomorpha phylogeny (figure 8), with the greatest amounts of destabilizing selection occurring at the branches leading to *R. similis*, *M. incognita*, *D. siricidicola*, *Ditylenchus* spp., and *A. avenae*. These results show that selection on *eng1* in Tylenchomorpha nematodes is occurring at the interface between Tylenchomorpha species and their specific host. As such, it is likely the specific interaction between nematode and host that is driving the evolution of this parasitism gene rather than the more general trophic interactions (i.e. plant vs insect vs fungus).

Acknowledgements

This work was supported by a Brigham Young University Mentored Environment Grant, and by a National Research Initiative of the USDA Cooperative State Research, Education and Extension Service grant (number 2002-01974) to BJA. We would like to thank K. Crandall, M. Whiting, A. Harker, and G. Poinar for advice and guidance on this project.

References

Abubucker, S., Zarlenga, D.S., Martin, J., Yin, Y., Wang, Z.Y., McCarter, J.P., Gasbarree, L., Wilson, R.K., Mitreva, M., 2009. The transcriptomes of the cattle parasitic nematode Ostertagia ostartagi. Veterinary Parasitology 162, 89-99.

Baldwin, J.G., Nadler, S.A., Adams, B.J., 2004. Evolution of plant parasitism among nematodes. Annual Review of Phytopathology 42, 83-105.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163-193.

Bert, W., Leliaert, F., Vierstraete, A.R., Vanfleteren, J.R., Borgonie, G., 2008. Molecular phylogeny of the Tylenchina and evolution of the female gonoduct (Nematoda : Rhabditida). Molecular Phylogenetics and Evolution 48, 728-744.

Bird, D.M., Opperman, C.H., Davies, K.G., 2003. Interactions between bacteria and plant-parasitic nematodes: now and then. International Journal for Parasitology 33, 1269-1276.

Blaxter, M., Dorris, M., De Ley, P., 2000. Patterns and processes in the evolution of animal parasitic nematodes. Nematology 2, 43-55.

Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., Vida, J.T., Thomas, W.K., 1998. A molecular evolutionary framework for the phylum Nematoda. Nature 392, 71-75.

Chan, K.M.A., Moore, B.R., 2005. SYMMETREE: whole-tree analysis of differential diversification rates. Bioinformatics 21, 1709-1710.

Davis, E.L., Hussey, R.S., Baum, T.J., Bakker, J., Schots, A., 2000. Nematode parasitism genes. Annual Review of Phytopathology 38, 365-396.

De Ley, P., Blaxter, M., 2002. Systematic position and phylogeny. In: Lee, D.L. (Ed.), Biology of Nematodes. Taylor & Francis, Florence, KY, pp. 1 - 30.

Dolinski, K., Botstein, D., 2007. Orthology and functional conservation in eukaryotes. Annual Review of Genetics 41, 465-507.

Dorris, M., De Ley, P., Blaxter, M.L., 1999. Molecular analysis of nematode diversity and the evolution of parasitism. Parasitology Today 15, 188-193.

Ebersberger, I., Strauss, S., von Haeseler, A., 2009. HaMStR: Profile hidden markov model based search for orthologs in ESTs. BMC Evolutionary Biology 9, 157.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792-1797.

Fitch, W.M., 1970. Distinguishing homologous from analagous proteins Systematic Zoology 19, 99.

Fitch, W.M., 2000. Homology - a personal view on some of the problems. Trends Genet. 16, 227-231.

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774-786.

Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Systematic Zoology 38, 297-309.

Hillis, D.M., 1996. Inferring complex phylogenies. Nature 383, 130-131.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Systematic Biology 47, 3-8.

Holterman, M., Karssen, G., van den Elsen, S., van Megen, H., Bakker, J., Helder, J., 2009. Small Subunit rDNA-Based Phylogeny of the Tylenchida Sheds Light on Relationships Among Some High-Impact Plant-Parasitic Nematodes and the Evolution of Plant Feeding. Phytopathology 99, 227-235.

Holterman, M., van der Wurff, A., van den Elsen, S., van Megen, H., Bongers, T., Holovachov, O., Bakker, J., Helder, J., 2006. Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. Molecular Biology and Evolution 23, 1792-1800.

Huang, G.Z., Allen, R., Davis, E.L., Baum, T.J., Hussey, R.S., 2006. Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. Proceedings of the National Academy of Sciences of the United States of America 103, 14302-14306.

Huang, G.Z., Dong, R.H., Maier, T., Allen, R., Davis, E.L., Baum, T.J., Hussey, R.S., 2004. Use of solid-phase subtractive hybridization for the identification of parasitism gene candidates from the root-knot nematode *Meloidogyne incognita*. Molecular Plant Pathology 5, 217-222.

Huang, G.Z., Gao, B.L., Maier, T., Allen, R., Davis, E.L., Baum, T.J., Hussey, R.S., 2003. A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode *Meloidogyne incognita*. Molecular Plant-Microbe Interactions 16, 376-381.

Hussey, R.S., Davis, E.L., Baum, T.J., 2002. Secrets in secretions: Genes that control nematode parasitism of plants. Brazilian Journal of Plant Physiology 14, 183-194.

Iseli, C., Jongeneel, C.V., Bucher, P., 1999. ESTScan: A Program for Detecting, Evaluating, and Reconstructing Potential Coding Regions in EST Sequences. Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology. AAAI Press, pp. 138-158.

Jones, J.T., Furlanetto, C., Kikuchi, T., 2005. Horizontal gene transfer from bacteria and fungi as a driving force in the evolution of plant parasitism in nematodes. Nematology 7, 641-646.

Koonin, E.V., 2005. Orthologs, paralogs, and evolutionary genomics. Annual Review of Genetics 39, 309-338.

Kuzniar, A., van Ham, R.C.H.J., Pongor, S., Leunissen, J.A.M., 2008. The quest for orthologs: finding the corresponding gene across genomes. Trends Genet. 24, 539-551.

Ledger, T.N., Jaubert, S., Bosselut, N., Abad, P., Rosso, M.N., 2006. Characterization of a new beta-1,4-endoglucanase gene from the root-knot nematode *Meloidogyne incognita* and evolutionary scheme for phytonematode family 5 glycosyl hydrolases. Gene 382, 121-128.

Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Research 13, 2178-2189.

Maddison, W.P., Maddison, D.R., 2002. MacClade version 4.0.5. Sinauer, Sunderland, Massachusetts.

Poinar, G., 2003. Trends in the evolution of insect parasitism by nematodes as inferred from fossil evidence. Journal of Nematology 35, 129-132.

Poinar, G., Kerp, H., Hass, H., 2008. *Palaeonema phyticum* gen. n., sp n. (Nematoda : Palaeonematidae fam. n.), a Devonian nematode associated with early land plants. Nematology 10, 9-14.

Poinar, G.O., Acra, A., Acra, F., 1994. Earliest fossil nematode (Mermithidae) in Cretaceous lebanese amber. Fundamental and Applied Nematology 17, 475-477.

Pond, S.L.K., Frost, S.D.W., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531-2533.

Pond, S.L.K., Frost, S.D.W., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676-679.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer

Remm, M., Storm, C.E.V., Sonnhammer, E.L.L., 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. Journal of Molecular Biology 314, 1041-1052.

Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798-804.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Schreiber, F., Pick, K., Erpenbeck, D., Worheide, G., Morgenstern, B., 2009. OrthoSelect: a protocol for selecting orthologous groups in phylogenomics. Bmc Bioinformatics 10, 219.

Siddall, M.E., Whiting, M.F., 1999. Long-Branch Abstractions. Cladistics 15, 9-24.

Sorenson, M.D., Franzosa, E.A., 2007. TreeRot version 3. Boston University, Boston, MA.

Stamatakis, A., 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688-2690.

Subbotin, S.A., Sturhan, D., Chizhov, V.N., Vovlas, N., Baldwin, J.G., 2006. Phylogenetic analysis of Tylenchida Thorne, 1949 as inferred from D2 and D3 expansion fragments of the 28S rRNA gene sequences. Nematology 8, 455-474.

Tatusov, R., Fedorova, N., Jackson, J., Jacobs, A., Kiryutin, B., Koonin, E., Krylov, D., Mazumder, R., Mekhedov, S., Nikolskaya, A., Rao, B.S., Smirnov, S., Sverdlov, A., Vasudevan, S., Wolf, Y., Yin, J., Natale, D., 2003. The COG database: an updated version includes eukaryotes. Bmc Bioinformatics 4, 41.

van der Heijden, R., Snel, B., van Noort, V., Huynen, M.A., 2007. Orthology prediction at scalable resolution by phylogenetic tree analysis. Bmc Bioinformatics 8.

Wiens, J.J., Moen, D.S., 2008. Missing data and the accuracy of Bayesian phylogenetics. Journal of Systematics and Evolution 46, 307-314.

Woolley, S., Johnson, J., Smith, M.J., Crandall, K.A., McClellan, D.A., 2003. TreeSAAP: Selection on Amino Acid Properties using phylogenetic trees. Bioinformatics 19, 671-672.

Xiang, Q.Y., Moody, M.L., Soltis, D.E., Fan, C.Z., Soltis, P.S., 2002. Relationships within Cornales and circumscription of Cornaceae - matK and rbcL sequence data and effects of outgroups and long branches. Molecular Phylogenetics and Evolution 24, 35-57.

Zmasek, C., Eddy, S., 2002. RIO: Analyzing proteomes by automated phylogenomics using resampled inference of orthologs. Bmc Bioinformatics 3, 14.

Table Legend

Table 1

List of Tylenchomorpha taxa with expressed sequence tag (EST) data available on GenBank, and the number of ESTs that are available for each taxon. Highlighted taxa indicate taxa that were utilized in the present phylogenomic analysis of Tylenchomorpha.

Table 2

Tylenchomorpha supermatrix information detailing each of the 30 genetic datasets used in the present analysis, their sequence length, position within the supermatrix, and best fit model of evolution as selected using the AIC in Modeltest.

Table 3

Total amount of missing data included in the Tylenchomorpha supermatrix for each taxon and the percent of missing data for each taxon.

Table 4

Partitioned bremer support values for each gene in the supermatrix. Node numbers correspond to nodes in the parsimony strict consensus tree (figure 1)

Figure Legends

Figure 1

Mixed models Bayesian tree for the Tylenchomorpha supermatrix with posterior probability values indicated above branches. Analyses were run for 10,000,000 generations, sampling every 1000 generation, and partitioning by gene.

Figure 2

Maximum likelihood tree for the Tylenchomorpha supermatrix constructed in RAxML using the GTRGAMMA model of nucleotide evolution with partitioning by gene. Likelihood bootstrap (1000 replicates) values are indicated above branches where concordant.

Figure 3

Parsimonious strict consensus tree of two most parsimonious trees, constructed using the new technology search in TNT with ratcheting, fusing, and drifting and 1000 random addition sequences. Numbers at the nodes correspond to node numbers in the partitioned bremer support table (table 1). Bootstrap values are indicated on or under branches.

Figure 4

Investigation of long-branch attraction between *Strongyloides ratti* and *Meloidogyne* spp. using the long-branch extraction method. All trees were constructed in TNT using the new technology search with ratcheting, fusing, and drifting and 1000 random addition sequences. *Strongyloides ratti* was removed from the analysis that produced tree A, *Caenorhabditis briggsae* was removed from the analysis that produced tree B, all *Meloidogyne* spp. were removed from the analysis that

produced tree C, and both *S. ratti* and *C. briggsae* were removed from the analysis that produced tree D.

Figure 5

Mixed models Bayesian tree for the tylenchomorpha supermatrix excluding *Strogyloides ratti,* with posterior probability values indicated above branches. Analyses were run for 10,000,000 generations, sampling every 1000 generation, and partitioning by gene.

Figure 6

Maximum likelihood tree for the Tylenchomorpha supermatrix excluding *Strongyloides ratti,* constructed in RAxML using the GTRGAMMA model of nucleotide evolution with partitioning by gene. Likelihood bootstrap (1000 replicates) values are indicated above branches where concordant.

Figure 7

Parsimony strict consensus tree of two most parsimonious trees for the Tylenchomorpha supermatrix excluding *Strongyloides ratti*, constructed using the new technology search in TNT with ratcheting, fusing, and drifting and 1000 random addition sequences. Bootstrap values are indicated on or under branches.

Figure 8

Tylenchomorpha phylogenetic tree with total instances of positive destabilizing selection, as inferred from TreeSAAP, mapped onto each branch. The figures shows the greatest amounts of

destabilizing selection occurring at the branches leading to *R. similis*, *M. incognita*, *D. siricidicola*, *Ditylenchus* spp., and *A. avenae*.

Table 1

Plant Parasitic	*Bursaphelenchus mucronatus*	3,193
	Bursaphelenchus xylophilus	13,340
	Ditylenchus africanus	4,847
	Globodera pallida	9,020
	Globodera rostochiensis	11,851
	Heterodera glycines	24,444
	Heterodera schachtii	2,812
	Meloidogyne arenaria	5,042
	Meloidogyne chitwoodi	12,218
	Meloidogyne hapla	24,452
	Meloidogyne incognita	20,334
	Meloidogyne javanica	7,587
	Melooidogyne paranaensis	3,710
	Pratylenchus penetrans	1,916
	Pratylenchus vulnus	5,812
	Radopholus similis	7,380
	Rotylenchulus reniformis	2,004
	Globodera mexicana	17
	Heterodera avenae	1
Insect Parasitic	*Deladenus siricidicola*	24,745
Fungal Feeding	*Aphelenchus avenae*	2,586

Table 2

Gene	# of Taxa	Seq. Length	Position in Super Matrix	Model
Ubiquitin, UBQ family	18	666	1-666	GTR+I+G
Ubiquitin/Ribosomal L40 protein	15	390	667-1056	TrNef+I+G
Calmodulin	10	447	1057-1503	TIM+G
Troponin C, EF Hand Family	10	507	1504-2010	GTR+I+G
Myosin Light Chain 3, Myosin Light Chain Family	17	453	2011-2463	TrN+I+G
Myosin essential light chain, EF-Hand protein superfamily	16	819	2464-3282	GTR+I+G
40S ribosomal protein S13, Small subunit family member	17	465	3283-3747	TIMef+I+G
26S protease regulatory subunit, proteasome-like protein	16	807	3748-4554	GTR+I+G
Actin, Actin Family Member	17	858	4555-5412	SYM+I+G
60S ribosomal protein L3, Large subunit family member	17	1215	5413-6627	TIM+I+G
60S acidic ribosomal protein P0	13	699	6628-7326	TrN+I+G
2-cysteine peroxiredoxin, PeRoxireDoXin family member	15	681	7327-8007	GTR+I+G
Cyclophilin 3, CYclophyliN family member	17	525	8008-8532	TrNef+I+G
60S ribosomal protein L5, Large subunit family member	18	930	8533-9462	TIM+I+G
40S ribosomal protein S20, Small subunit family member	17	387	9463-9849	TrN+I+G
casein kinase 1, alpha 1, KINase family member	11	885	9850-10734	GTR+I+G
beta-tubulin	14	999	10735-11733	TrN+I+G
alpha tubulin, TuBulin, Alpha family member	15	1350	11734-13083	TIM+I+G
small subunit ribosomal protein 1	16	786	13084-13869	TIM+I+G
Translationally controlled tumor protein	16	561	13870-14430	GTR+I+G
40S ribosomal protein S11	15	510	14431-14940	TrN+G
histone H2B	16	432	14941-15372	K81uf+I+G
histone H3.3, HIStone family member	14	414	15373-15786	K80+I+G
Histone H2A, Histone family member	13	345	15787-16131	TVM+I+G
40S ribosomal protein S25	17	381	16132-16512	TrN+I+G
large subunit ribosomal protein 2	17	792	16513-17304	GTR+I+G
large subunit ribosomal protein 9	20	567	17305-17871	GTR+I+G
Dynein light chain 1, cytoplasmic	15	270	17872-18141	TrN+I+G
arginine kinase	20	627	18142-18768	SYM+I+G
Heat Shock Protein family member	12	348	18769-19116	TIM+I+G

Table 3

Species	Missing Data	Percent Missing
Aphelenchus avenae	11,970	62.62
Bursaphelenchus mucronatus	8,883	46.47
Bursaphelenchus xylophilus	9,630	50.38
Deladenus siricidicola	6,900	36.10
Ditylenchus africanus	12,469	65.23
Globodera pallida	8,763	45.84
Globodera rostochiensis	7,794	40.77
Heterodera glycines	8,028	42.00
Heterodera schachtii	12,809	67.01
Meloidogyne arenaria	8,835	46.22
Meloidogyne chitwoodi	8,706	45.54
Meloidogyne hapla	8,364	43.75
Meloidogyne incognita	7,171	37.51
Meloidogyne javanica	10,483	54.84
Melooidogyne paranaensis	8,896	46.54
Pratylenchus penetrans	12,384	64.78
Pratylenchus vulnus	11,706	61.24
Radopholus similis	6,732	35.22
Rotylenchulus reniformis	17,892	93.60
Strongyloides ratti	4,281	22.39
Caenorhabditis briggsae	9,855	51.55

Table 4

Node	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Gene																
Ubiquitin, UBQ family	-7	16	29	1	-4	-22	-17.86	0	-25	0	0	-25	-25	-25	-25	-2
Ubiquitin/Ribosomal L40 protein	-7	0	2	0	2.5	-7	-5.71	0	-8	0	0	-8	-8	-8	-8	15
Calmodulin	0	0	10	1	-6	-1	0	0	0	0	0	0	0	0	0	3
Troponin C, EF Hand Family	11	0	0	0	0	9	7.86	0	11	0	0	11	11	11	11	0
Myosin Light Chain 3, Myosin Light Chain Family	-13	0	1	0	18	-25	-10.71	0	-15	0	0	-15	-15	-15	-15	-3
Myosin essential light chain, EF-Hand protein superfamily	-27	0	0	0	-10	-36	-19.29	0	-27	0	0	-27	-27	-27	-27	8
40S ribosomal protein S13, Small subunit family member	17	0	9.5	4	10.5	18	12.14	0	17	0	0	17	17	17	17	28
26S protease regulatory subunit, proteasome-like protein	-3	-1	14	0	1	-6	-2.14	0	-3	0	0	-3	-3	-3	-3	2
Actin, Actin Family Member	-22	-10	7	-7	-3	-22	-15.71	0	-22	0	0	-22	-22	-22	-22	0
60S ribosomal protein L3, Large subunit family member	13	0	0	0	0	27	9.29	0	13	0	0	13	13	13	13	0
60S acidic ribosomal protein P0	-2	0	11	5	16	2	1.43	0	2	0	0	2	2	2	2	0
2-cysteine peroxiredoxin, PeRoxireDoXin family member	-1	1	7	2	2	15	-0.71	0	-1	0	0	-1	-1	-1	-1	0
Cyclophilin 3, CYclophyliN family member	8.5	2	8.25	1.5	6.5	10.5	6.21	0	8.5	0.25	0.25	8.5	8.5	8.5	8.5	-0.5
60S ribosomal protein L5, Large subunit family member	9	1	1	11	15.5	8	6.43	0	9	0	0	9	9	9	9	34
40S ribosomal protein S20, Small subunit family member	13	0	4	16	7	13	10	0	14	0	0	14	14	14	14	0
casein kinase 1, alpha 1, KINase family member	-13	0	0	0	0	-13	-9.29	0	-13	0	0	-13	-13	-13	-13	0
beta-tubulin	-9.8	0	2.5	1	-7	-18	-4.14	2	-10	0	0	-5.8	-5.8	-5.8	-5.8	3
alpha tubulin, TuBulin, Alpha family member	4.5	0	0.25	0.5	0	7.5	3.36	0	4.5	0.25	0.25	4.5	4.5	4.5	4.5	-0.5
small subunit ribosomal protein 1	4	-2	3	5	35	22	2.86	0	4	0	0	4	4	4	4	10
Translationally controlled tumor protein	12.5	0	0.25	1.5	8	1.5	9.07	0	12.5	0.25	0.25	12.5	12.5	12.5	12.5	16.5
40S ribosomal protein S11	15.5	0	0.25	3.5	5	20.5	11.21	0	15.5	0.25	0.25	15.5	15.5	15.5	15.5	-0.5
histone H2B	-6.4	0	-1	0	-1	-1	-6.86	-1	-13	-3	-3	-8.4	-8.4	-8.4	-8.4	2
histone H3.3, HIStone family member	-12	0	7.5	-1	13	-5	-8.57	0	-12	0	0	-12	-12	-12	-12	0
Histone H2A, Histone family member	-6	0	0	0	12	-8	-4.29	0	-6	0	0	-6	-6	-6	-6	31
40S ribosomal protein S25	1	0	5	5	8	1	2.14	0	3	0	0	3	3	3	3	0
large subunit ribosomal protein 2	18	0	0	0	-1	20	12.86	0	18	0	0	18	18	18	18	19
large subunit ribosomal protein 9	12	3	5.5	7	22.5	17	7.29	0	11	-1	-1	11	11	11	11	2
Dynein light chain 1, cytoplasmic	2.8	0	-1	6	-5	1	1.43	2	4	-2	-2	2.8	2.8	2.8	2.8	0
arginine kinase	11.4	-2	13	1	8.5	13	10.14	-2	19	7	7	11.4	11.4	11.4	11.4	8
Heat Shock Protein family member	-9	0	8	-2	15	-9	-6.43	0	-9	0	0	-9	-9	-9	-9	0

Figure 1

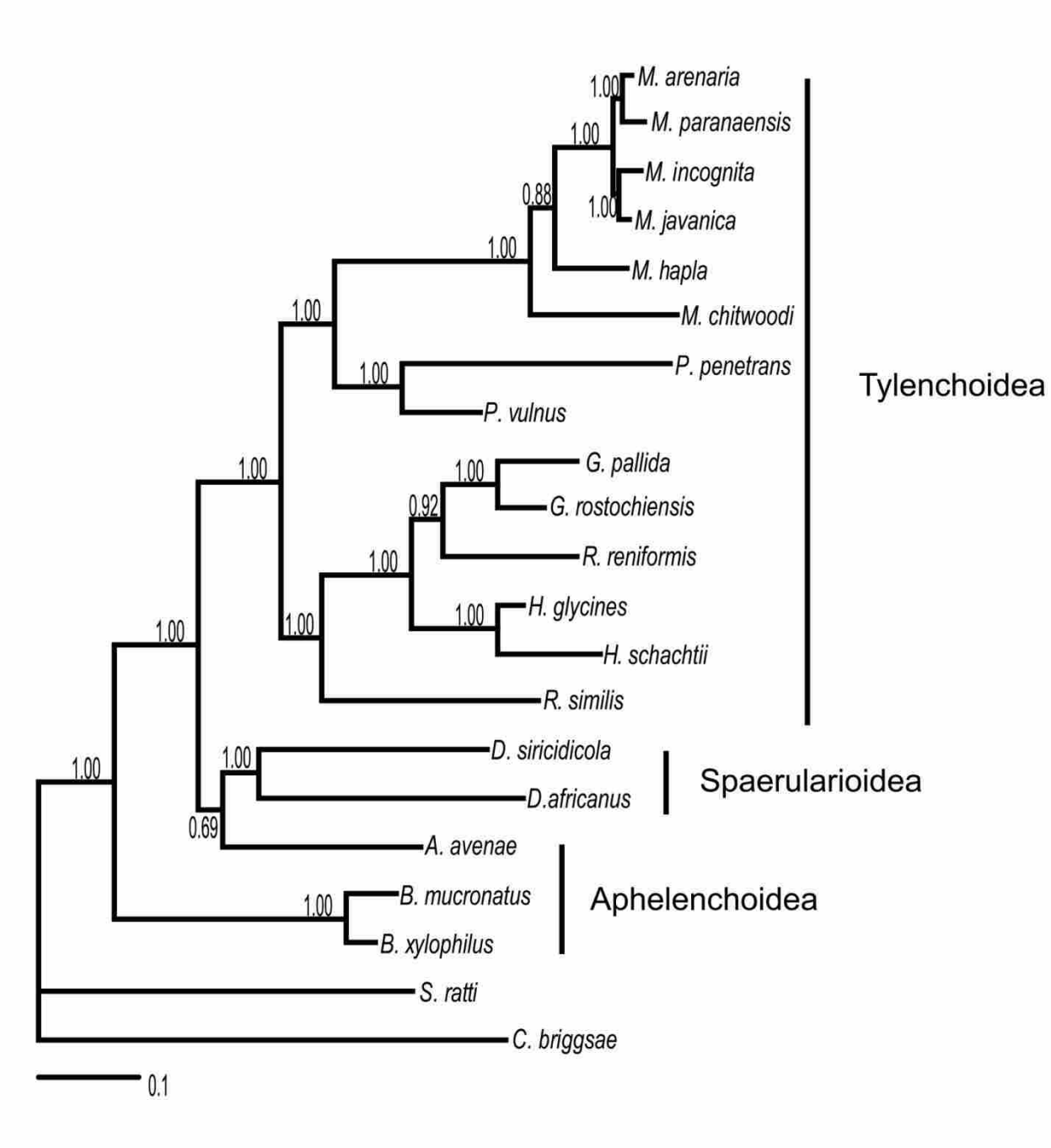

M. arenaria
M. paranaensis
M. incognita
M. javanica
M. hapla
M. chitwoodi
P. penetrans
P. vulnus
G. pallida
G. rostochiensis
R. reniformis
H. glycines
H. schachtii
R. similis
D. siricidicola
D.africanus
A. avenae
B. mucronatus
B. xylophilus
S. ratti
C. briggsae
Tylenchoidea
Spaerularioidea
Aphelenchoidea
1.00
0.88
0.92
0.69
0.1

Figure 2

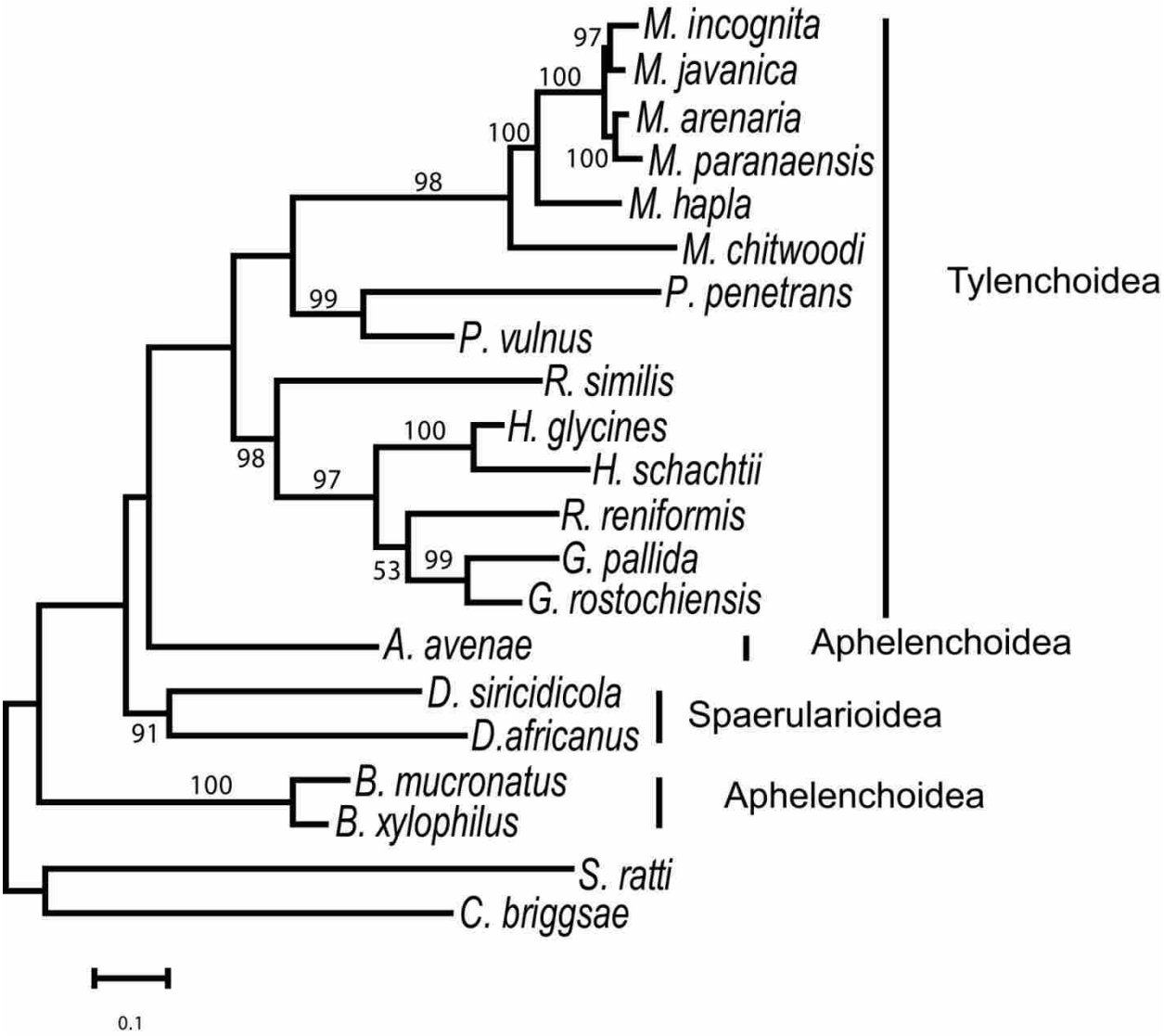

M. incognita
M. javanica
M. arenaria
M. paranaensis
M. hapla
M. chitwoodi
P. penetrans
P. vulnus
R. similis
H. glycines
H. schachtii
R. reniformis
G. pallida
G. rostochiensis
A. avenae
D. siricidicola
D.africanus
B. mucronatus
B. xylophilus
S. ratti
C. briggsae
97
100
100
100
98
99
98
100
97
53
99
91
100
0.1
Tylenchoidea
Aphelenchoidea
Spaerularioidea
Aphelenchoidea

Figure 3

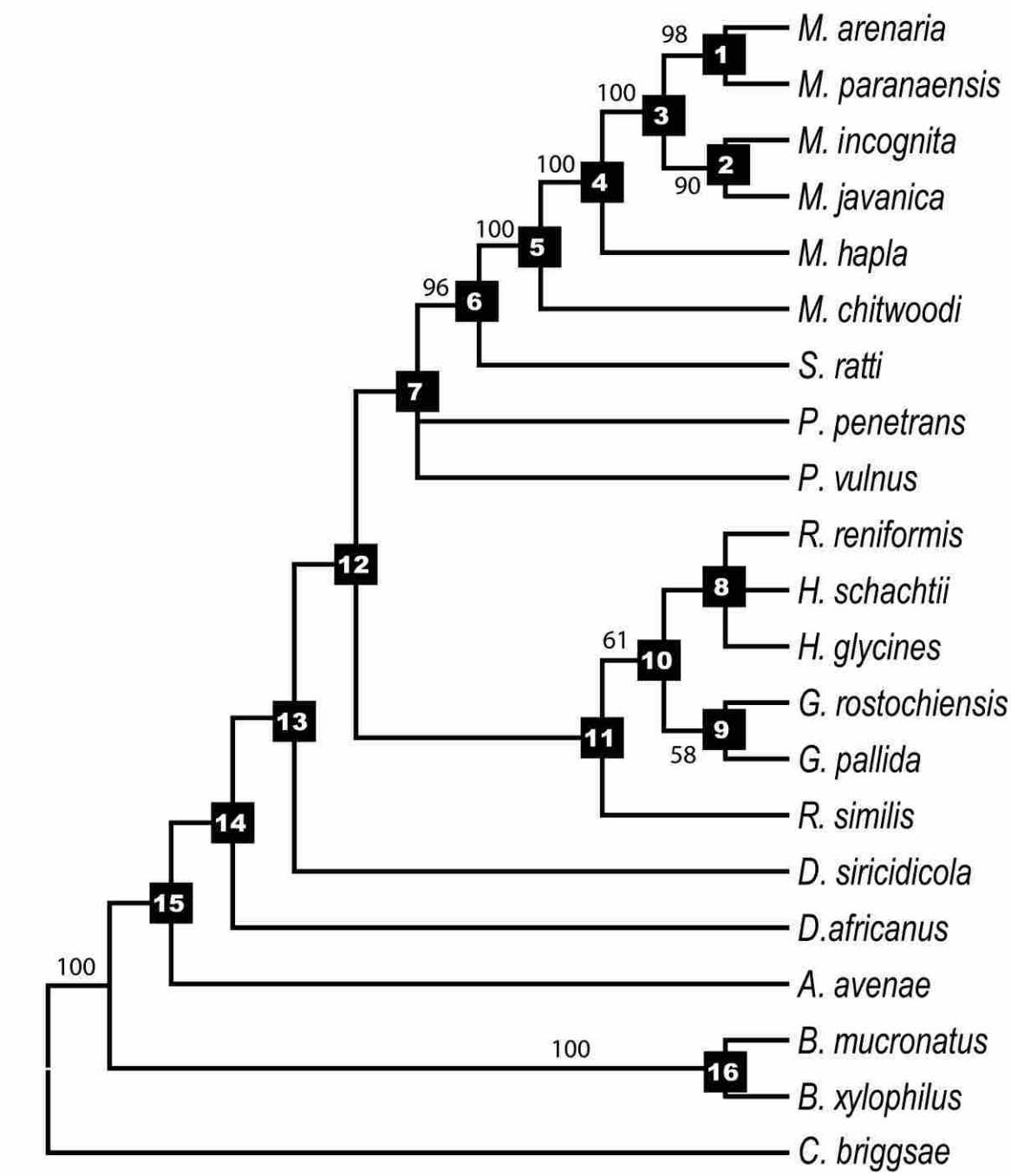

M. arenaria
M. paranaensis
M. incognita
M. javanica
M. hapla
M. chitwoodi
S. ratti
P. penetrans
P. vulnus
R. reniformis
H. schachtii
H. glycines
G. rostochiensis
G. pallida
R. similis
D. siricidicola
D.africanus
A. avenae
B. mucronatus
B. xylophilus
C. briggsae
98
100
100
90
100
96
61
58
100
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Figure 4

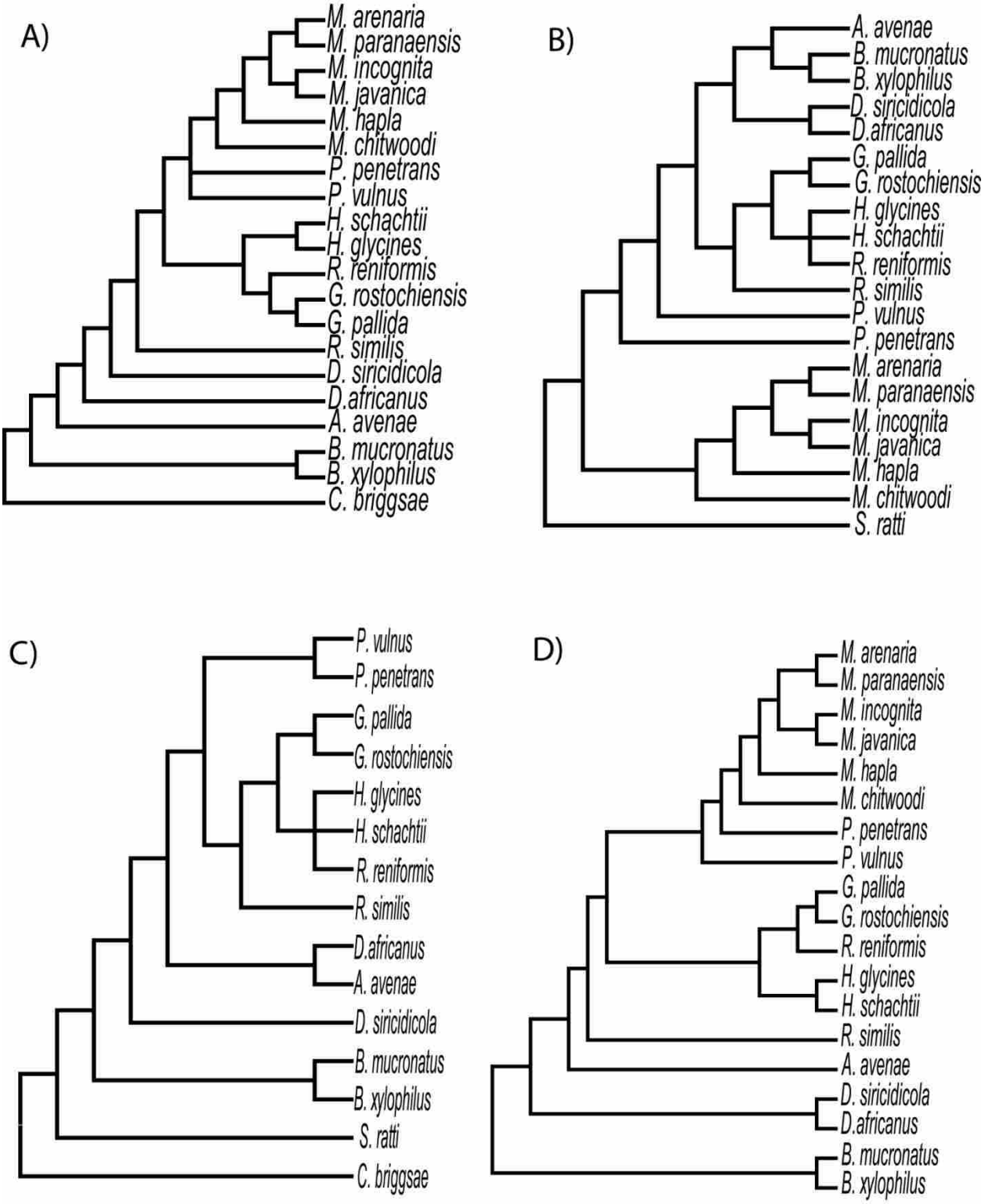

A)
M. arenaria
M. paranaensis
M. incognita
M. javanica
M. hapla
M. chitwoodi
P. penetrans
P. vulnus
H. schachtii
H. glycines
R. reniformis
G. rostochiensis
G. pallida
R. similis
D. siricidicola
D.africanus
A. avenae
B. mucronatus
B. xylophilus
C. briggsae
B)
A. avenae
B. mucronatus
B. xylophilus
D. siricidicola
D.africanus
G. pallida
G. rostochiensis
H. glycines
H. schachtii
R. reniformis
R. similis
P. vulnus
P. penetrans
M. arenaria
M. paranaensis
M. incognita
M. javanica
M. hapla
M. chitwoodi
S. ratti
C)
P. vulnus
P. penetrans
G. pallida
G. rostochiensis
H. glycines
H. schachtii
R. reniformis
R. similis
D.africanus
A. avenae
D. siricidicola
B. mucronatus
B. xylophilus
S. ratti
C. briggsae
D)
M. arenaria
M. paranaensis
M. incognita
M. javanica
M. hapla
M. chitwoodi
P. penetrans
P. vulnus
G. pallida
G. rostochiensis
R. reniformis
H. glycines
H. schachtii
R. similis
A. avenae
D. siricidicola
D.africanus
B. mucronatus
B. xylophilus

Figure 5

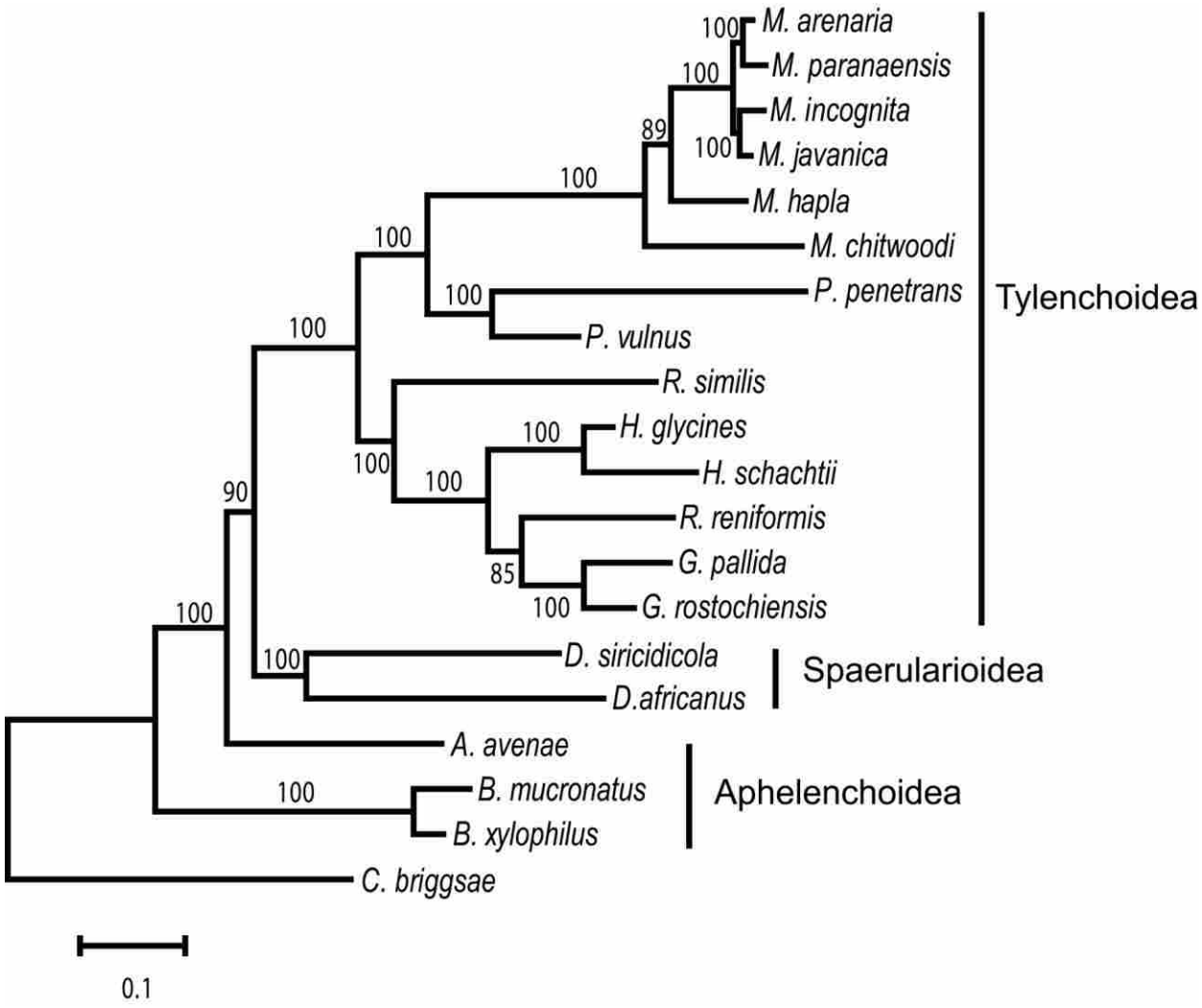

M. arenaria
M. paranaensis
M. incognita
M. javanica
M. hapla
M. chitwoodi
P. penetrans
P. vulnus
R. similis
H. glycines
H. schachtii
R. reniformis
G. pallida
G. rostochiensis
D. siricidicola
D.africanus
A. avenae
B. mucronatus
B. xylophilus
C. briggsae
Tylenchoidea
Spaerularioidea
Aphelenchoidea
100
100
89
100
100
100
100
100
100
100
90
100
85
100
100
100
0.1

Figure 6

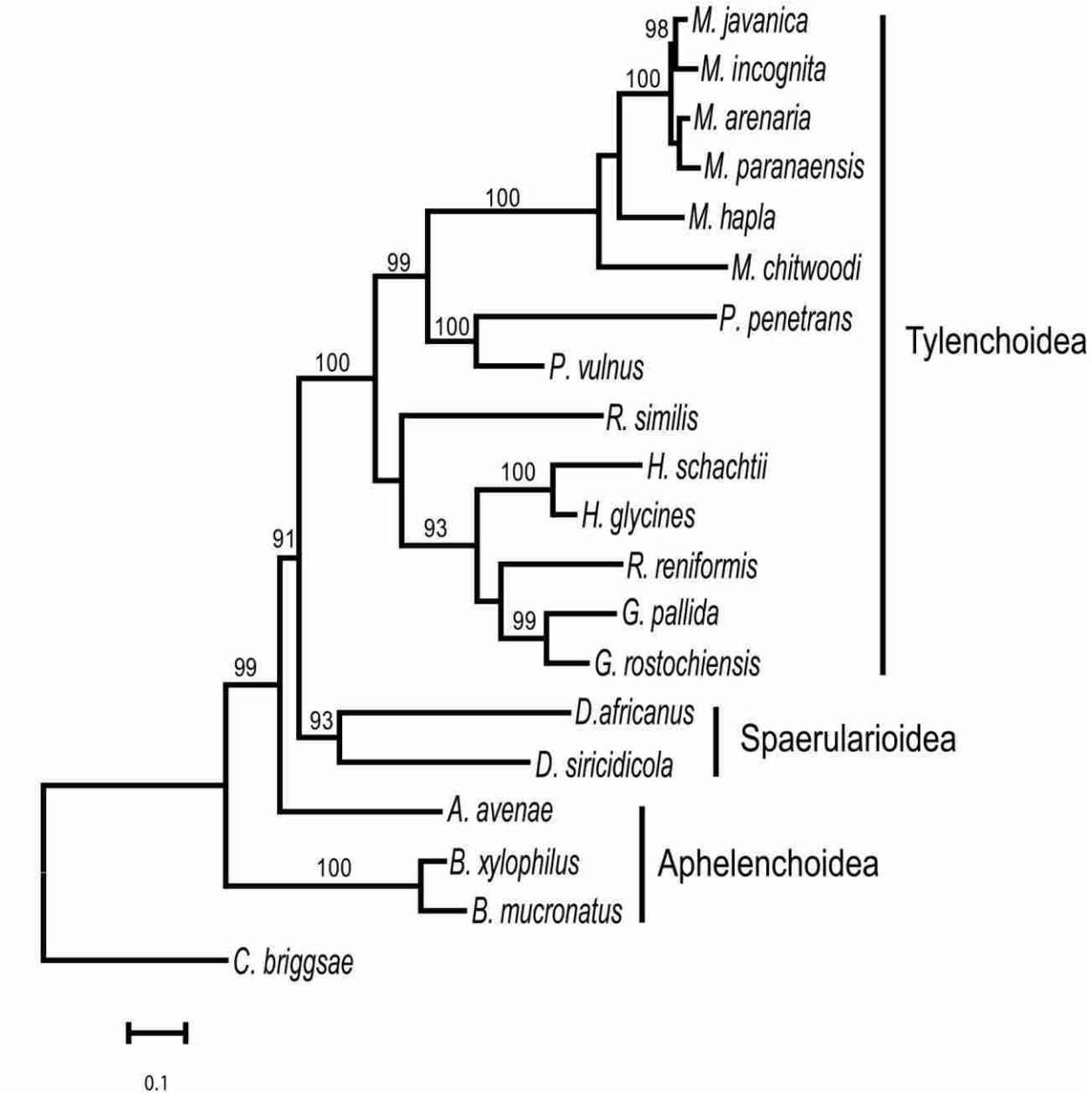

M. javanica
M. incognita
M. arenaria
M. paranaensis
M. hapla
M. chitwoodi
P. penetrans
P. vulnus
R. similis
H. schachtii
H. glycines
R. reniformis
G. pallida
G. rostochiensis
D.africanus
D. siricidicola
A. avenae
B. xylophilus
B. mucronatus
C. briggsae
98
100
100
99
100
100
91
100
93
99
99
93
100
Tylenchoidea
Spaerularioidea
Aphelenchoidea
0.1

Figure 7

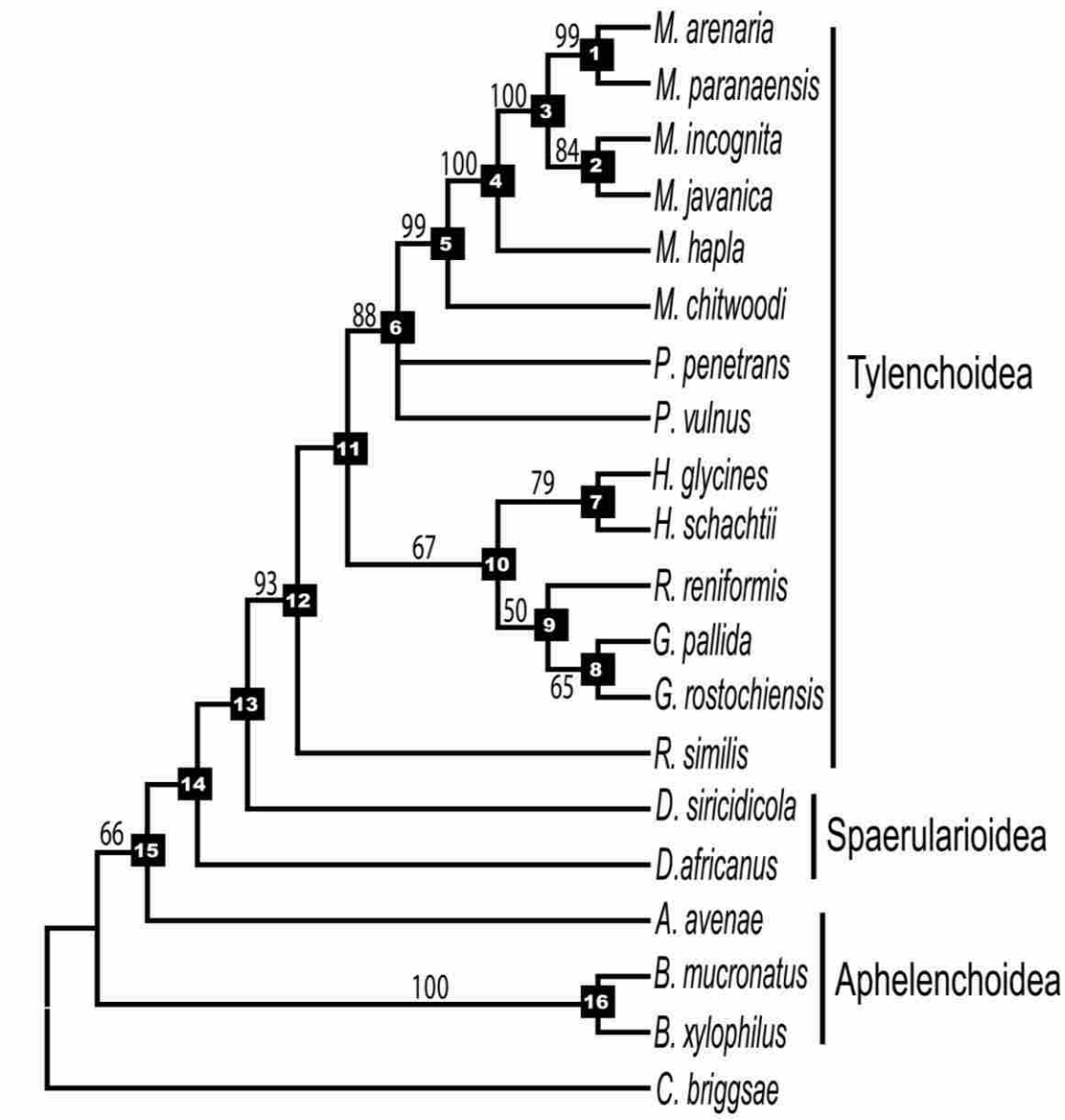
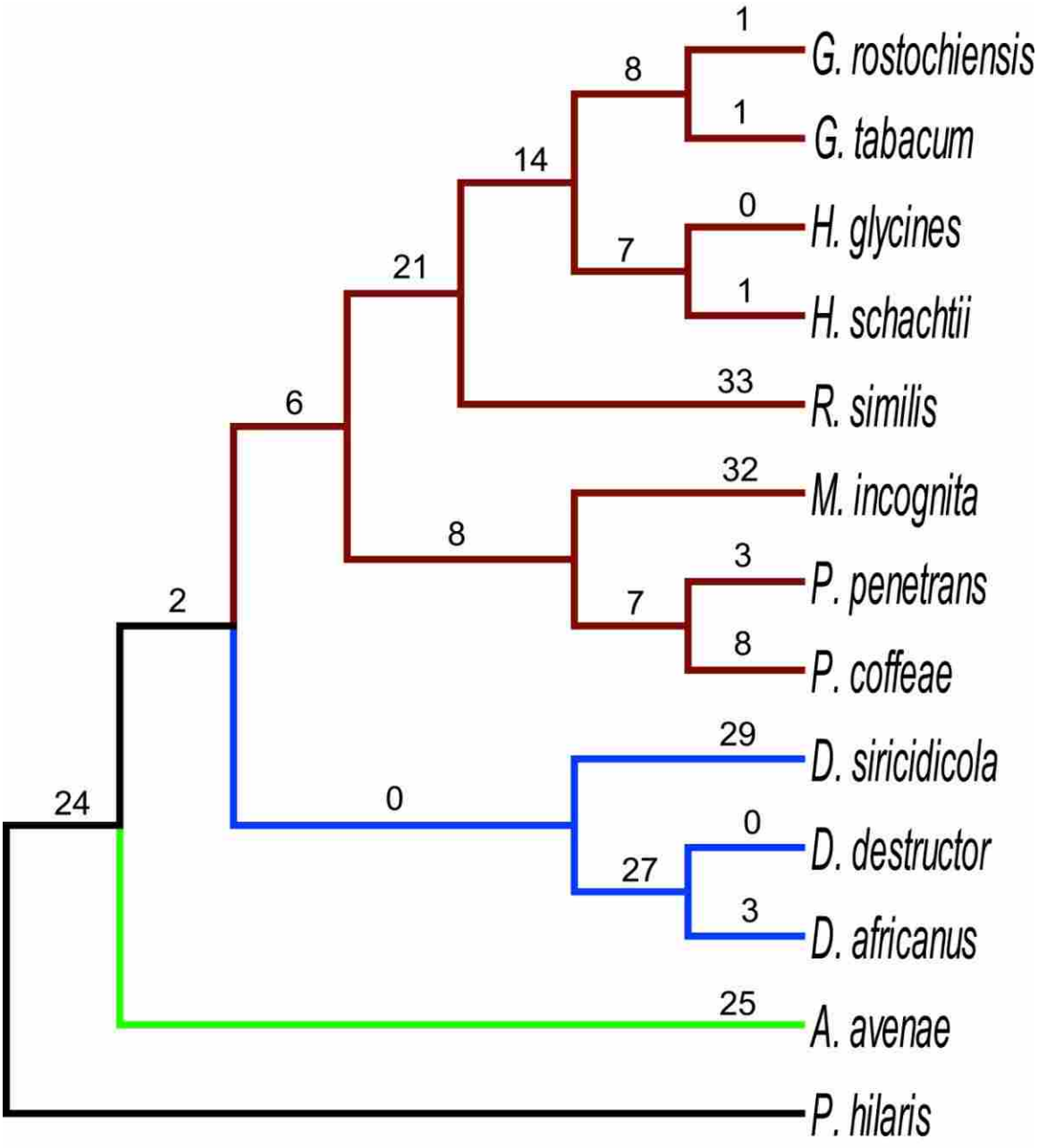

M. arenaria
M. paranaensis
M. incognita
M. javanica
M. hapla
M. chitwoodi
P. penetrans
P. vulnus
H. glycines
H. schachtii
R. reniformis
G. pallida
G. rostochiensis
R. similis
D. siricidicola
D.africanus
A. avenae
B. mucronatus
B. xylophilus
C. briggsae
Tylenchoidea
Spaerularioidea
Aphelenchoidea
99
100
84
100
99
88
79
67
93
50
65
66
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Figure 8


1
G. rostochiensis
8
1
G. tabacum
14
0
H. glycines
7
21
1
H. schachtii
33
6
R. similis
32
M. incognita
8
3
P. penetrans
7
2
8
P. coffeae
29
D. siricidicola
24
0
0
D. destructor
27
3
D. africanus
25
A. avenae
P. hilaris