AN ABSTRACT OF THE DISSERTATION OF

Michael J. Raboin for the degree of Doctor of Philosophy in Molecular and Cellular Biology presented on June 11, 2012.

Title: The Genus *Caenorhabditis*: A System for Testing Evolutionary Questions

Abstract approved:

_____

Dee R. Denver

Caenorhabditis elegans is arguably the best understood animal on the planet. Used for over 50 years to study development, we have a vast amount of knowledge of the inner workings of this worm. Our knowledge is incomplete, however, without placing this organism in its evolutionary and ecological context. In this body of work, I focused on examining the evolutionary forces shaping *Caenorhabditis* nematodes, with a particular emphasis on *C. briggsae*. In the first part, I examined the evolution of mitochondrial genomes throughout the genus. I tested for signatures of selection and examined the evolution of mitochondrial genome architecture. Through this, I have shown that the mitochondrial genomes of *Caenorhabditis* nematodes appear to be primarily influenced by purifying selection and that molecular evolutionary inference is greatly limited by mutational saturation. The evolutionary forces acting

on mitochondrial genomes have been examined before, however, this study, extensively examining this within a single genus, provides a much better characterization than any of the studies to date. In the second part, I characterized the evolutionary dynamics of mitochondrial pseudogenes in *C. briggsae* and its closest relatives. I showed that these elements, while they might not evolve under strictly neutral terms, are still quite useful in uncovering cryptic diversity and population structure. I also observed that they appear/disappear in a manner that appears inconsistent with one commonly held model for mitochondrial pseudogene evolution. In the final part, I examined the evolution of *C. briggsae* in response to a biotic environment. I showed that fitness in a parasite-containing environment incurs a trade-off with fitness in the absence of parasites. Together, the chapters of this dissertation demonstrate the strength of *Caenorhabditis*, and in particular *C. briggsae*, for examining evolutionary questions and advances this system as a tool for evolutionary biology research.

The Genus *Caenorhabditis*: A System for Testing Evolutionary Questions

by
Michael J. Raboin

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 11, 2012
Commencement June 2013

Doctor of Philosophy dissertation of Michael J. Raboin presented on June 11, 2012.

APPROVED:

_____

Major Professor, representing Molecular and Cellular Biology

_____

Director of the Molecular and Cellular Biology Program

_____

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

_____

Michael J. Raboin, Author

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my major professor Dee Denver for giving me the opportunity to study under his guidance. His intellectual input has been invaluable. Without his advice, encouragement, and support this dissertation would not have been possible. I would like to thank my committee for their dedication and interest in my research. I would like to thank the members of the Denver lab for their encouragement, support, and for the many useful discussions over the years. Finally, I would like to thank my parents, sisters, in-laws, and especially my wife for their unceasing encouragement.

CONTRIBUTION OF AUTHORS

Dee R. Denver contributed to the design and data interpretation in chapters 2, 3 and 4. He also helped in preparation of associated publications for chapters 3 and 4. Dana K. Howe conducted the DNA sequencing and phylogenetic analysis in chapter 2, and aided in data collection and analysis in chapter 3. Marie-Anne Félix collected many of the C. briggsae strains used in chapter 3 and provided us with them. Larry J. Wilhelm performed assembly of the mitochondrial sequence data for chapter 2. Ashley F. Timko designed DNA primers used for chapter 3. Katie A. Clark assisted with DNA preparation in chapter 2.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

The Genus *Caenorhabditis*: A System for Testing Evolutionary Questions

Chapter 1

Introduction and literature review

Evolutionary analyses

"Nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973). This oft quoted title by Theodosius Dobzhansky was the subject of at least two papers in which he described the necessity for evolutionary thought (Dobzhansky, 1964, 1973). In opposition to the reductionist views of molecular biology, which sought to break things down to their most basic components, he argued that only through evolution does the unity and diversity of life make sense. At every level of organization, nothing exists in a vacuum, but interacting with other things and frequently coming together in various patterns to make up the next level of organization, from molecules to ecosystems.

> *Nevertheless, at the lower levels of integration the type of question most frequently asked is "how things are," while at the higher levels an additional question insistently obtrudes on the mind of the investigator - "how things got to be that way."* (Dobzhansky, 1964)

This could then be said to be the basic goal of evolutionary analysis: to understand how things change over time, and to understand the forces and interactions shaping this change. In examining the natural history of a species, there are many aspects that we might desire to investigate to understand the question of what shaped a species into what we observe today. We might desire to know about migrations, population bottlenecks, and population and range expansions; interactions between populations or

with other organisms, including parasites; changes from a free-living lifestyle to a parasitic one, or vice versa; or changes in reproductive modes.

History of *Caenorhabditis* research

Over the past 50 years, the genus *Caenorhabditis* has developed into an invaluable tool to investigate biological phenomenon. In the early 1960's, Sydney Brenner began searching for an ideal system in which to study development and the nervous system. After initially working with *Caenorhabditis briggsae*, Brenner eventually chose *C. elegans* (Brenner, 1974, 1988). Initially, his proposal was to track the lineage of every cell and investigate the genetic control of development (Brenner, 1988). Since that initial proposal, a wealth of knowledge has been accumulated for *C. elegans*. The first *C. elegans* genetic map was published in 1974 (Brenner, 1974). The structure of the nervous system was determined (White et al., 1986), and the complete cell lineage of the worm was mapped (Sulston, 1988; Sulston et al., 1988). Finally, *C. elegans* had the distinction of being the first multicellular organism to have its genome completely sequenced (*C. elegans* Sequencing Consortium, 1998). Since then, there has been increasing work focused on evolutionary and ecological questions that have utilized not only *C. elegans*, but in some cases other members of the genus as well. This is perhaps best exemplified by the completion of the genome sequence of *C. briggsae* for comparison to *C. elegans* (Stein et al., 2003).

Extensive global sampling that has occurred in recent years has yielded not only numerous new species to analyze and compare, but has revealed new details of the evolutionary history of *Caenorhabditis* nematodes. In contrast to the initial description of *C. elegans* as a soil-nematode by Maupas in 1900 (as quoted in Félix and Braendle, 2010), *Caenorhabditis* are now thought to be fruit nematodes, growing on decaying plant matter (Félix and Braendle, 2010; Kiontke et al., 2011). Sampling has also produced the first naturally-occurring pathogens of *Caenorhabditis*, providing a system in which to explore evolution in response to parasitism (Troemel et al., 2008; Félix et al., 2011). Work on recently discovered viruses has demonstrated standing variations in susceptibility between populations of *C. elegans* (Félix et al., 2011). Global collections have also facilitated the examination of population genetic structure, and global evolutionary dynamics. While *C. elegans* seems to lack population structure (Denver et al., 2003; Cutter, 2006; Dolgin et al., 2008), in contrast to *C. briggsae* (Cutter et al., 2006b; Dolgin et al., 2008; Raboin et al., 2010), the populations none the less seem to be influenced by recurrent selective sweeps as large, megabase size haplotype blocks have swept through the global population (Andersen et al., 2012). This is all the more interesting given that *C. elegans* populations experience outbreeding depression (Dolgin et al., 2007), potentially limiting the amount of outcrossing that occurs. These features are in stark contrast with its relative, *C. remanei*, which rapidly breaks down linkage, possesses vastly higher

nucleotide diversity (Cutter et al., 2006a), and is affected by inbreeding depression in full-sibling crosses (Dolgin et al., 2007).

While Brenner originally utilized mutagenic approaches to examine the genes underlying behavioral traits (Brenner, 1974), more recently, interest has shifted to examining naturally-occurring variation and using it to uncover the basis of phenotypes. Recombinant inbred lines have been developed for both *C. elegans* and *C. briggsae* to facilitate associating variation in traits to their underlying genetic factors (Seidel et al., 2008; Ross et al., 2011); for *C. elegans*, higher resolution recombinant inbred advanced intercross lines have also been developed (Rockman and Kruglyak, 2009). In *C. elegans*, these approaches have proven to be quite fruitful. One of the first successes of this approach resulted in identifying the polymorphism responsible for variation in copulatory plug deposition seen in *C. elegans* (Palopoli et al., 2008). This behavior, which involves the deposition of a gelatinous plug on a hermaphrodite by a male during mating, varies between *C. elegans* isolates, with some exhibiting this behavior, and others not (Hodgkin and Doniach, 1997). Following this, studies were able to link variation in other behaviors to underlying polymorphisms. Variation in parasite resistance was associated with a gene involved in aerotaxis, demonstrating that resistance in this case is conferred through behavioral changes (Reddy et al., 2009). Furthermore, polymorphism in a G-protein-coupled receptor was linked to variation in foraging behavior (Bendesky et al., 2011). Beyond examining the levels and distributions of polymorphisms, or connecting them to phenotypic

changes, *Caenorhabditis* species have been used to examine the patterns of one of the most basic of evolutionary processes: mutation.

C. elegans and *C. briggsae* have found a place in investigating mutational processes. *C. elegans* has been used to measure direct estimates of nuclear and mitochondrial mutation rates (Denver et al., 2000, 2004, 2009), with similar estimates following in *C. briggsae* (Howe et al., 2010; Denver et al., 2012). These studies reported higher mutation rates in both the nuclear and mitochondrial genomes than had previously been reported, and described the types of mutations that were occurring. While the mitochondrial and nuclear genome mutation rates between these species are similar (Howe et al., 2010; Denver et al., 2012), the mitochondrial genomes of these two species experience different, predominant mutational processes (Howe et al., 2010). In a further examination of mutational processes, *C. briggsae* has also been utilized to examine how organisms might recover mutationally from losses of fitness; providing evidence for compensatory mutations and epistatic interactions in the recovery process (Denver et al., 2010).

Caenorhabditis briggsae

The genus *Caenorhabditis* is a diverse, and globally distributed group (Dolgin et al., 2008; Raboin et al., 2010; Kiontke et al., 2011) that is particularly well-suited to investigating a variety of evolutionary phenomenon (eg. Graustein et al., 2002; Kiontke et al., 2004; Wang et al., 2004; Hill et al., 2006). Nematodes of this genus

vary in their ranges and habitats with some being very restricted, while others are cosmopolitan, differ in reproductive modes, and cover a wide range of genetic divergence (Kiontke and Sudhaus, 2006; Cutter, 2008; Kiontke et al., 2011). There are also several closely-related species pairs available for comparison (Raboin et al., 2010; Woodruff et al., 2010; Kiontke et al., 2011). While *C. elegans* is the historical model organism in the group, we contend that *C. briggsae* provides a better anchoring point for evolutionary studies. For evolutionary questions, *C. briggsae* possesses many advantages over *C. elegans*. Certainly not the least of these is the presence of closely-related sister taxa including *C.* sp. 9, *C.* sp. 5, and *C. remanei* (Raboin et al., 2010; Kiontke et al., 2011). Despite the discovery of numerous new species recently, no close relatives of *C. elegans* have been found (Kiontke et al., 2011). *C.* sp. 9 is the most closely related species to *C. briggsae* (Raboin et al., 2010; Kiontke et al., 2011), with crosses between the two species producing semi-fertile offspring (Woodruff et al., 2010; Kiontke et al., 2011). Crosses with *C.* sp. 5 also produces viable, though sterile, progeny (Kiontke et al., 2011). In contrast to it's close relatives, *C. briggsae* is an androdioecious hermaphrodite; most likely a derived character evolved from the dioecy seen in much of the rest of *Caenorhabditis*, including *C.* sp. 9, *C.* sp. 5 and *C. remanei* (Kiontke et al., 2004, 2011; Denver et al., 2011). In contrast to *C. elegans*, *C. briggsae* is roughly latitudinally structured (Denver et al., 2003; Cutter, 2006; Cutter et al., 2006b; Dolgin et al., 2008). Clade I strains of *C. briggsae* occur between the tropic of Cancer and the tropic of Capricorn, Clade II is restricted to the northern

hemisphere, and Clade III has been located in both equatorial Kenya and India (Dolgin et al., 2008; Raboin et al., 2010). While there is some overlap between Clades I and III, Clade II is found exclusively in the temperate regions of the northern hemisphere and does not overlap with the other two clades (Dolgin et al., 2008; Raboin et al., 2010). Clade II is proposed to have expanded into the northern hemisphere relatively recently (Cutter et al., 2006b). Additionally, some ecological differences have been proposed between the clades, particularly between the Clade I (tropical) and Clade II (temperate) strains (Joyner-Matos et al., 2009; Prasad et al., 2011).

Mitochondrial DNA as a tool for evolutionary analyses

Mitochondrial DNA has long been the marker of choice for addressing many questions about the evolutionary history of populations and species. In animals, the mitochondrial genome is inherited uniparentally, usually without recombination, and possesses a higher mutation rate than the nuclear genome, which allows distinguishing between closely related groups (reviewed in: Ballard and Whitlock, 2004; Galtier et al., 2009). As such, it has been used to examine phylogenetic relationships between populations, and to infer the histories of organisms (Avise, 2000; Zink and Barrowclough, 2008). It has found use in describing population structure, and some have even examined the mitochondrial genome for signs of adaptation to different environments (Ruiz-Pesini et al., 2004; da Fonseca et al., 2008; Hassanin et al., 2009; Foote et al., 2011; Tomasco and Lessa, 2011). Overlapping with this, there has been

research into how the mitochondrial genome itself evolves and what selective pressures are affecting it specifically (eg. Denver et al., 2000; Bazin et al., 2006; Galtier et al., 2006; Lynch et al., 2006). These are questions of great interest given the vital role of the mitochondrion in the cell. Due to the well sampled phylogeny (Kiontke et al., 2011), especially at the *C. briggsae* end of the tree (Raboin et al., 2010), this system is particularly well suited to examine the evolutionary forces acting on mitochondrial genomes throughout the genus. The closely related species provide a unique opportunity for comparative genomics in *Caenorhabditis*.

Dissertation outline

The goal of this body of work is to evaluate the evolutionary forces acting on *Caenorhabditis* nematodes, with particular focus on *Caenorhabdtis briggsae*. This work will be presented in five chapters. After the introductory chapter will be three chapters presenting data from several evolutionary analyses performed leveraging different aspects of the *Caenorhabditis* system. In the first data chapter, I will examine evolution on a broad scale, and will seek to elucidate what selective forces are acting on mitochondrial genomes throughout the genus *Caenorhabditis* (Chapter 2). I will proceed under the hypothesis that the mitochondrial genomes of nematodes throughout the genus are evolving according to the neutral theory. I expect to find purifying selection as the primary force acting on the genome. Additionally, I expect that any signatures of positive selection seen will be most likely the result of

mutational saturation, which has been observed in nematode species before (Blouin et al., 1998). In the second data chapter, I will move to a finer scale and examine the evolution of pseudogene elements in the mitochondrial genomes of *C. briggsae* and its close relatives, in addition to using these elements to reconstruct fine-scale population history (Chapter 3). I expect that these noncoding elements will evolve more rapidly than the rest of the genome, and will therefore be able to provide fine-scale resolution of population structure, but what can they tell us about the origin, proliferation, and demise of noncoding sequence in the mitochondrial genome? Due to the position of one specific noncoding element in *C. briggsae* and *C.* sp. 5 (Howe and Denver, 2008), I will examine the hypothesis that the noncoding elements behave in a manner more consistent with illegitimate recombination than the tandem-duplication random loss model. In the third data chapter, I will examine evolution in even more narrow terms: investigating the clades of *C. briggsae* for variation in fitness in a parasite-containing environment (Chapter 4). Certain phenotypic differences in *C. briggsae* appear to follow clade delineations (Joyner-Matos et al., 2009; Prasad et al., 2011). Additionally, the phylogenies indicate that there may have been a single colonization event of the temperate, northern latitudes (Cutter et al., 2006b, 2010; Raboin et al., 2010). Given that there is evidence that these nematodes can migrate over large distances (Denver et al., 2003; Raboin et al., 2010), why haven't more independent colonizations occurred? I hypothesize that the clades of *C. briggsae* differ in their biotic, ecological tolerances. More generally, I will test the expectation that trade-offs

will be observed between growth in a benign, laboratory environment, and a parasite-containing one. Finally, I will follow the data chapters with a discussion of the material with a focus on the conclusions derived from the overall body of work, and a look at the future prospects of research in the genus *Caenorhabditis* (Chapter 5).

Chapter 2

Characterizing signatures of selection in *Caenorhabditis* mitochondrial genomes

Michael J. Raboin, Dana K. Howe, Larry J. Wilhelm, Katie A. Clark, Dee R. Denver

Abstract

The mitochondrial genome is a widely used molecular marker in evolutionary biology. Studies seeking to examine the evolutionary forces acting on mitochondrial DNA (mtDNA) have produced evidence for the action of both positive and negative selection. A point of concern, however, has been the comparison of distantly-related taxa. Taking advantage of a large dataset consisting of 84 nearly-complete mitochondrial genome sequences from 23 species of *Caenorhabditis* nematodes, we have performed evolutionary and population genomic analyses to examine mtDNA evolution within this single nematode genus. We examined multiple levels of sequence including nearly-complete mitochondrial genomes, concatenated sequences of the 12 protein-coding genes, individual protein-coding sequences, and noncoding elements. At each of these levels we have calculated multiple population genetic statistics, such as synonymous (d_S) and nonsynonymous (d_N) mutations, transition/transversion ratios, nucleotide diversity, and the neutrality index. We tested for evidence of selection through plotting d_N vs. d_S, and performing the McDonald-Kreitman test. We tested for signatures of saturation by comparing d_S, neutrality indexes, and transition/transversion ratios. Our results provide evidence that purifying selection and silent-site mutation saturation are major features of these genomes. Saturation of silent-site mutations was observed between all but the most recently diverged species, leading to false signals of positive selection. Our study provides important insights into within- and between-species evolution of nematode

mitochondrial genomes, and indicates that caution must be used when performing mtDNA comparisons between species to avoid the confounding effects of saturation.

Introduction

The mitochondrial genome differs from nuclear DNA in a number of ways that have resulted in it being an attractive tool for use in ecological and evolutionary studies. The mitochondrial genome itself is small, with a conserved coding content across animals generally consisting of 13 protein-coding genes, 22-24 tRNA genes, and 2 ribosomal RNA genes (Gissi et al., 2008). It is most frequently inherited in a uniparental manner, with little recombination (Xu, 2005; Galtier et al., 2009), and possesses a mutation rate ~10 fold greater than the nuclear genome, though potentially much higher (Lynch et al., 2006). These features have led to it being utilized in elucidating phylogeny (Arnason et al., 2002; Raboin et al., 2010), calculating population size in conservation work (Roman and Palumbi, 2003), and identifing species (Hebert et al., 2003, 2004). With the widespread use of mitochondrial DNA, however, the question has arisen as to how it has evolved.

For many years the accepted view was that the mtDNA acted in a neutral manner (Ballard and Kreitman, 1995). Accordingly, studies have found evidence that the major force acting on the protein-coding genes of the mitochondrial genome is negative selection; consistent with the neutral theory of molecular evolution (Kimura, 1983). Studies in the nematode *C. briggsae* (Howe and Denver, 2008), humans (Sun

et al., 2007), and canines (Björnerfeldt et al., 2006; Rutledge et al., 2010) have reported nonsynonymous/synonymous mutation ratios lower than 1 and studies of mtDNA in cod (Marshall et al., 2009) have reported neutrality indexes greater than 1, both indicative of purifying selection. Despite this, other studies have reported evidence of positive selection acting on the genome (e.g. Oliveira et al., 2008; Hassanin et al., 2009; Foote et al., 2011; Tomasco and Lessa, 2011). These studies have not been without controversy, however. The most comprehensive of of these studies, by Bazin et al. (2006), was disputed primarily on the basis of comparing distant taxa (Wares et al., 2006). It should be noted that some have gone so far as to question the validity of the statistical methods themselves (Hughes, 2007; Nei et al., 2010).

In recent years, the genus *Caenorhabditis* has developed into a model system for evolutionary studies (Carvalho et al., 2006; Haag et al., 2007; Kammenga et al., 2008; Kiontke et al., 2011). Global sampling efforts have resulted in discovering numerous new species, and filling in the phylogeny of *Caenorhabditis* (Kiontke et al., 2011). As a result, 26 related species, isolated globally, are available which allows us to make comparisons at varying evolutionary distances. In several cases, we have species pairs diverging by less than 0.08 substitutions/site in RNA polymerase II (Kiontke et al., 2011). Importantly, we are beginning to accumulate genetic data about the members of the genus.

The availability of numerous, related species covering varying evolutionary distances makes the *Caenorhabditis* dataset ideal for facilitating comparisons both within and between species, and examining the evolutionary forces acting on the mitochondrial genome of *Caenorhabditis* nematodes. In the current study, our objective is to examine the mitochondrial genome for signs of neutral evolution, or selective pressure. Due to the unusual presence of mitochondrial DNA pseudogenes in the genus *Caenorhabditis* (Howe and Denver, 2008; Raboin et al., 2010), the evolution of these elements will also be examined and compared to that of the protein-coding genes. Additionally, the genus *Caenorhabditis* covers a wide span of evolutionary divergence. Estimated divergence time between the two most thoroughly studied members of the group, *C. elegans* and *C. briggsae* may be up to 30 million years (Cutter, 2008), equivalent to the divergence time between rats and mice (Kim et al., 2011). As others have warned of the danger of utilizing distant outgroups for comparison (Wares et al., 2006), we will examine the dataset for evidence of saturation, and determine which comparisons might prove the most informative.

Methods

mtDNA sequences

Nearly complete mitochondrial genomes were obtained for eighty-four natural isolate-derived strains of *Caenorhabditis* nematodes. (Sequences do not include the AT region.) Forty-one sequences were obtained from GenBank

(http://www.ncbi.nlm.nih.gov/genbank/). The remaining 43 were produced through Illumina sequencing. Briefly, mitochondrial genomes were amplified using the Roche Long Range dNTP Pack and an overnight reaction. The primers C_lsuA_F1 (5'-AATGGCAGTCTTAGCGTGA-3'), C_lsuB_R1 (5'-AATTTCCGAAGACTTATCTT-3'), C_ssuA_F1 (5'-GTTCCAGAATAATCGGCTA-3'), and C_ssuB_R1 (5'-GCAATTGATGGATGATTTG-3') were used to amplify two overlapping regions of the genome from the large subunit rRNA to the small subunit rRNA. PCR and sequencing were performed as done previously (Denver et al., 2000, 2003), and Illumina reads were assembled using the PERL script SCRAPE.pl (http://denverlab.cgrb.oregonstate.edu/node/14). SCRAPE.pl assembles Illumina reads by feeding multiple random samples into the assembly program Velvet (Zerbino and Birney, 2008). The contigs resulting from this first-pass assembly are then assembled using CAP3 (Huang and Madan, 1999), resulting in 1-5 contigs that were then manually assembled if needed. Altogether, this resulted in sequences for 23 species of *Caenorhabditis* nematodes including: 26 strains of *C. briggsae*, 20 strains of *C. elegans*, 2-6 strains for 10 species, and single representatives for an additional 12 species.

Sequence alignment and identification

Multiple alignments and sequence manipulation were conducted using MEGA5 (Tamura et al., 2011). Sequences were first aligned with ClustalW (Thompson et al.,

1994) using default settings. The multiple alignments were then examined for quality by visual inspection, and adjusted manually if necessary. Individual gene sequences for *C. elegans* were retrieved from GenBank and manually aligned to the multiple alignment using *C. elegans* strain N2 as a guide, and alignments were trimmed to contain only the sequence between the genes for *trnP* and *trnA*. Questionable sequences at the 5' and 3' ends were removed. Protein-coding gene regions in the multiple alignment identified in this manner were confirmed through translating the sequence to amino acids to ensure proper translation, and thus, the determination of the appropriate sequence and reading frame. This method was used to identify sequence-order rearrangements, such as the insertion of pseudogene elements. The 3' end of the alignment, encompassing from the stop codon of *nad5* onward, was further aligned using MUSCLE (Edgar, 2004) under default settings.

Based on the identified gene regions, and the translated amino acid sequences, appropriate sequences were extracted from the multiple alignment for analysis. Sequences for each of the 12 protein-coding genes were extracted from the alignment and concatenated. Sequences were adjusted to ensure that overlapping regions of adjacent protein-coding genes were not accounted for twice, and questionable sequence was removed where necessary. For further investigation, noncoding sequences identified through gene alignment were extracted on a species-by-species basis for analysis. The nonconserved, noncoding sequences identified in *C.* sp. 7 and *C.* sp. 14 were also submitted to NCBI BLASTN and BLASTX

(http://blast.ncbi.nlm.nih.gov/Blast.cgi) comparing against the nr database in order to identify the sequence origin.

Sequence evolution analysis

A variety of measures were calculated for different sets of sequences. Sequence analysis was conducted within DnaSP v5.1 (Librado and Rozas, 2009). Nucleotide polymorphism (π) was calculated for all species for which we have sequence for two or more strains available. Nucleotide polymorphism was calculated across all sequence from *trnP* to *trnA* using a scanning-window of 100 nucleotides and a step size of 25. Nucleotide polymorphism was also calculated for all noncoding elements. Similarly, within-species polymorphism at synonymous (π_S) and nonsynonymous (π_N) sites was calculated across the set of concatenated protein-coding genes for all species for which we have more than a single mtDNA sequence available. All π values were subjected to the Jukes-Cantor correction (Jukes and Cantor, 1969).

A number of measures were also calculated to provide comparison between species. Nucleotide divergence was calculated for synonymous (d_S) and nonsynonymous (d_N) sites at multiple levels of comparison, and corrected using the Jukes-Cantor model. These measures were calculated across the concatenated protein-coding genes for all pairwise comparisons of clades within *C. briggsae*, and species within the *Elegans* group, *Japonica* group, and *Drosophilae* supergroup (Fig. 2.1). Synonymous and nonsynonymous divergence was also calculated for each protein-

coding gene for all comparisons between *C. briggsae* clades, comparisons of species to the nearest outgroup within the *Elegans* group, and between *C. angaria* and *C.* sp. 12. The neutrality index (Rand and Kann, 1996) was calculated via the McDonald-Kreitman test (McDonald and Kreitman, 1991) for certain comparisons. The McDonald-Kreitman test was performed across the concatenated protein-coding genes between *C. briggsae* clades, for all pairwise comparisons between species of the *Elegans* group, and between *C. angaria* and *C.* sp. 12. It was also performed on a gene-by-gene basis for all pairwise comparisons between *C. briggsae*, *C.* sp. 9, *C.* sp. 5, and *C. remanei*.

Lastly, transition/transversion ratios were examined. Nucleotide substitution models were tested using MEGA5. Concatenated protein-coding sequences were tested in pairwise comparisons of *C. briggsae* with *C.* sp. 9, *C.* sp. 5, *C. remanei*, and *C. elegans*. The results were compared to determine the best overall model of nucleotide substitution. In accordance with these results, transition/transversion ratios were calculated using the Hasegawa, Kishino and Yano model (Hasegawa et al., 1985) with gamma distributed variation in rate among sites (HKY + Γ). Transition/transversion ratios were calculated for the same comparisons and datasets as the McDonald-Kreitman test.

Results

Study overview

Recent work has called the neutrality of the mitochondrial genome into question (see reviews in Dowling et al., 2008; Galtier et al., 2009), though evidence for positive selection, even in the nuclear genome, is not without controversy (Wares et al., 2006; Hughes, 2007; Nei et al., 2010). We, therefore, set out to examine the evolution of the mitochondrial genome within a single genus. The amount of divergence in the genus *Caenorhabditis* should be noted, as previous work has placed the divergence of *C. elegans* and *C. briggsae* at approximately 30 million years (Cutter, 2008).

The work described here represents one major facet of a larger study on mitochondrial genome evolution in *Caenorhabditis* that was a collaborative effort with Dana K. Howe and Larry J. Wilhelm. To facilitate this study, the mitochondrial genome sequences of 41 strains of *Caenorhabditis* that have been previously sequenced were retrieved from GenBank, while the remaining 43 sequences for this study were newly sequenced using the Illumina platform. The resulting dataset contained nearly-complete mitochondrial genomes (excluding the control region) for 84 natural isolate-derived strains of *Caenorhabditis* nematodes covering 23 different species. Altogether, this set consisted of 26 strains of *C. briggsae*, 20 strains of *C. elegans*, 2-6 strains for 10 species, and a single representative strain for the remaining 12 species. DNA preparation and sequencing was performed by DKH with the aid of the Oregon State University Center for Genome Research and Biocomputing,

assembly of Illumina reads was performed by LJW, phylogenetic analysis was conducted by DKH, identification of new noncoding elements was conducted by MJR and DKH, and characterization of noncoding elements and population genetic analyses were conducted by MJR. The work presented here will focus on my major contributions to this larger study: the evolutionary characterization of noncoding elements, and the population genetic analyses.

Statistical tests of evolution were conducted at several levels. The MEGA5 (Tamura et al., 2011) software package was used to align genome sequences. Single gene sequences for all mitochondrially encoded genes in *C. elegans* strain N2 were retrieved from GenBank, and aligned to the mitochondrial genome sequences to identify gene boundaries and structural rearrangements. For analysis, noncoding elements for each species were extracted. Additionally, protein-coding genes were extracted and examined individually, and as a concatenated sequence containing all 12 protein-coding genes. Transition/transversion ratios for the concatenated protein-coding sequences were calculated using MEGA5, while all other analyses were conducted in DnaSP v5.1 (Librado and Rozas, 2009). To examine within- and between-species diversity, π_N, π_S, d_N, and d_S values were calculated for both concatenated protein-coding sequences, and for individual gene sequences. Ratios of nonsynonymous to synonymous mutations are expected to be equal if a sequence is evolving neutrally, while synonymous mutations are expected to be more abundant if the sequence is subject to purifying selection ($d_N/d_S < 1$), and nonsynonymous

mutations expected to predominate under positive selection ($d_N/d_S > 1$) (Yang and Bielawski, 2000). Following this, we employed neutrality indexes (Rand and Kann, 1996) derived from the McDonald-Kreitman test (McDonald and Kreitman, 1991) to test for selection between lineages. The McDonald-Kreitman test compares within- versus between-species diversity. If the lineages are evolving neutrally, these two ratios are expected to be equal and will produce a neutrality index of 1, whereas the within-species ratio will be greater under purifying selection between the lineages (N.I. > 1), and the between-species ratio will be greater if positive selection is acting between the lineages (N.I. < 1) (Rand and Kann, 1996).

Structural variation

Animal mitochondrial genomes typically contain little in the way of noncoding sequence (Lynch et al., 2006; Gissi et al., 2008). The *Caenorhabditis* genus, however, is known to contain comparably large noncoding elements (Howe and Denver, 2008; Raboin et al., 2010). In examining the multiple alignments of the mitochondrial genomes, two new noncoding elements were discovered: one in *C.* sp. 7 spanning 195 basepairs, and one in *C.* sp. 14 spanning 117 basepairs, bringing the total number in the genus to eight (Fig. 2.2). Both fall between *trnQ* and *trnF*, which is the same location as Ψnad5-1 found in *C. briggsae* and *C.* sp. 5 (Howe and Denver, 2008; Raboin et al., 2010). Testing for homology against the nonredundant nucleotide database using both BLASTN and BLASTX yielded no matches. For other

Caenorhabditis species for which we have multiple mtDNA sequences, scanning-window analyses of intraspecies nucleotide polymorphism across the entire mitochondrial genome was used to place the variation present within these noncoding sequences in the context of the entire mitochondrial landscape.

Despite their noncoding nature, these elements do not generally appear to evolve in a manner distinct from the rest of the mitochondrial genome in those *Caenorhabditis* species that we were able to test. *C. briggsae* Ψnad5-1 represents one end of the spectrum, as it was readily observable in the scanning-window analysis producing a peak far higher than the rest of the genome (Fig. 2.3). In contrast to this, the noncoding element discovered in *C.* sp. 11 possesses no variation among the available isolates. The valley produced at this point is indistinguishable from other regions of the genome possessing no variability among isolates. The other noncoding sequences examined in *C. briggsae*, *C.* sp. 5, and *C.* brenneri produced peaks, but were not readily distinguishable from other peaks of nucleotide variation seen throughout the genomes. It is apparent in these cases that other regions of the mitochondrial genome possess as much, if not more, variation than the noncoding element. Of the four peaks of nucleotide polymorphism observed to exceed 0.09 among the six species examined, two correspond to noncoding elements in *C. briggsae* and *C.* sp. 5, while the remaining two in *C. remanei* and *C.* sp. 5 do not (Fig. 2.3). Interestingly, *C. remanei* and *C.* sp. 5 seem to possess a higher amount of nucleotide polymorphism across their genomes than other species, with numerous

regions of their genomes possessing $\pi > 0.06$. This is consistent with higher levels of nucleotide polymorphism seen in the nuclear genomes of these species (Cutter et al., 2006a; Wang et al., 2010).

Evolution of protein-coding sequences

The protein-coding sequences of the mitochondrial genomes were examined for signatures of natural selection, and also for indications of saturation. Plots of d_N vs. d_S for both the concatenated protein-coding genes, and on a gene-by-gene basis indicate strong purifying selection acting across the protein-coding portion of the mtDNA. Data points fall well below the line indicative of neutral evolution ($d_N/d_S = 1$; Fig. 2.4A & 2.4B). Interestingly, the d_N and d_S values in our dataset do not appear to have a linear relationship, but rather, d_N appears to grow more rapidly as d_S increases. Logically, this is the type of result we might expect when examining saturated data: at higher levels of d_S, d_S becomes saturated, and thus, underestimates the true divergence. This causes d_N to appear to increase more rapidly relative to d_S. Based on these results, mtDNA in *Caenorhabditis* nematodes appears to be dominated primarily by purifying selection, and is susceptible to rapid mutational saturation.

In examining the effects of mutational saturation, there are several groupings within *Caenorabditis* that could prove quite informative, due to the presence of closely-related sister species. Investigation of the species most closely related to *C. briggsae* appears to provide evidence for both positive and negative selection acting

on the mitochondrial genome. Of particular interest is the *C. briggsae*-*C.* sp. 9 comparison. These two species are the most closely related pair in this part of the phylogeny (Raboin et al., 2010; Kiontke et al., 2011), and accordingly, they have the lowest d_S in the group (0.42). The neutrality index is significantly greater than 1 (p-value < 0.001; Fisher's exact test) and the transition/transversion ratio is relatively high (Fig. 2.5). Transition/transversion ratios can also provide an indication of saturation as they tend to decrease as you compare more distantly related groups and experience saturation (Thomas and Wilson, 1991; Blouin et al., 1998). Three comparisons within this group (*C.* sp. 9-*C.* sp. 5; *C.* sp. 5-*C. remanei*; *C.* sp. 9-*C. remanei*) provide highly significant indications of positive selection through low neutrality indexes. In none of the other comparisons of this group, however, does d_S fall below 0.6, or the transition/transversion ratio exceed 2.76.

Further examining the relationship between the neutrality index and d_S indicates that the trend seen for d_S seems to extend to the rest of the genus, as d_S appears to greatly affect the neutrality index (Fig. 2.6). In comparisons that are known to be between closely-related groups (*C. briggsae* clades; *C. briggsae*-*C.* sp. 9; *C. angaria*-*C.* sp. 12; see Kiontke et al., 2011), the d_S values all fall lower than 0.5, and the neutrality indexes indicate purifying selection. All the more distant comparisons within the *Elegans* group cluster between 0.6 and 0.9. Evidence for positive selection is only seen in these more distant comparisons. Additionally, d_S values did not necessarily increase with increasing phylogenetic distance. Numerous cases were

observed of d_S values between more distant comparisons falling below values for more closely related species.

Evolution of noncoding DNA

Caenorhabditis mtDNA appears to be unusual, as compared to other animal groups, in the number of noncoding that have arisen throughout it's history (Fig. 2.2). These elements provide a unique view on the evolution of the mitochondrial genome as they are expected to evolve neutrally. Comparing estimates of nucleotide polymorphism at these elements, they possess more polymorphism, and do appear to be evolving closer to neutrality than the protein-coding sequence (Fig. 2.7). All species of *Caenorhabditis* surveyed thus far contain a conserved intergenic region that falls between *nad4* and *cox1*. These elements possess only slightly more polymorphism than the protein-coding sequence. The other noncoding elements tend to possess substantially more, though despite this, they still generally fall below where we would expect under complete neutrality. The intergenic region of *C.* sp. 11 falls on the line indicative of where we would expect under neutrality, and the noncoding elements of *C. briggsae* and *C. brenneri* both fall relatively close. The noncoding elements of *C.* sp. 5, however, appear to be far more constrained, based on the levels of polymorphism observed at synonymous sites in this species ($\pi_S = 0.176$). These results might indicate that these noncoding elements are under some kind of selective constraint.

Discussion

Our understanding of the natural history of *Caenorhabditis* nematodes is improving as numerous studies of evolution and ecology are being conducted on members of the group. Recent reports have explored facets such as natural habitats (Kiontke et al., 2011), parasitic interactions (Schulenburg and Müller, 2004; Troemel et al., 2008; Abebe et al., 2010; Schulte et al., 2010), local adaptations (Cutter et al., 2010; Prasad et al., 2011), and the effects of mating system on diversity (Dolgin et al., 2007). Understanding if there is adaptive or neutral evolution occurring in the mitochondrial genomes of this genus could shed light on the ecological differences between species. Mitochondrial genomes have previously been suggested to play a role in adaptation to thermal regimes (Mishmar et al., 2003; Balloux et al., 2009). Additionally, there have been, thus far, three transitions to hermaphroditism observed throughout the genus (Denver et al., 2011). Is there evidence in *Caenorhabditis* that the mitochondrial genome is involved in adaptation? Or perhaps, is the additional reduction in population size associated with this reproductive mode causing the mitochondrial genomes in these species to evolve more neutrally?

Despite other reports of the mitochondrial genome evolving under positive selection (eg. Bazin et al., 2006; Oliveira et al., 2008; Tomasco and Lessa, 2011), we found no convincing evidence of this. The neutrality indexes for all of the closest comparisons were indicative of purifying selection (Fig. 2.6), though the comparison of *C. angaria*-*C.* sp. 12 was not significant. While neutrality indexes indicative of

positive selection, some of them highly significant (p-value < 0.001), were observed, these cases were only observed in more distant comparisons where we believe saturation to be an issue. Under the effect of saturation, we might expect to see many of our d_S values cluster together at the upper range of d_S, as they become more affected by stochasticity than actual evolutionary distance. An important question we might ask of the data is at what point saturation starts to become an issue. The clustering of all but the closest of comparisons between d_S values of 0.6-0.9, seems to indicate that we cannot be confident in results obtained beyond a d_S value of 0.6. The relatively low level of d_S is most likely the effect of base composition and transition biases in the nematode mitochondrial genome. In an unbiased system, every site can exist in four possible polymorphic states. Compositional bias, however, can affect the levels of sequence divergence (discussed in Collins et al., 1994).

Saturation was a major point of concern in the study by Bazin et al. (2006; see discussion in Wares et al., 2006). Our dataset is particularly well suited to examine the effects of saturation. Though the most widely studied species of the genus, *C. elegans* and *C. briggsae*, are ~30 million years diverged (Cutter, 2008), the genus contains a series of species of varying levels of divergence that can be examined (Cutter, 2008; Kiontke et al., 2011). This includes at least three pairs of much more closely related species: *C. briggsae*-C. sp. 9, *C. angaria*-C. sp. 12, and *C. drosophilae*-*C.* sp. 2. These three species pairs are only diverged by 0.079, 0.068, and 0.057

substitutions/site in RNA polymerase II, respectively. While these represent the closest species pairs, it should be noted that other pairs of *Caenorhabditis* nematodes are at least close enough to produce young larvae, or sterile adult progeny (Kiontke et al., 2011). Additionally, ongoing discovery of new nematodes in the genus will hopefully provide more opportunities in the future for comparisons without saturation.

Though we find no evidence for positive selection in *Caenorhabditis* mitochondrial genomes through the use of the neutrality index, this does not discount the possibility of positive selection acting on the mitchondrial genome. Tomasco and Lessa (2011) reported evidence of positive selection acting against a backdrop of purifying selection. Interestingly, the best evidence for positive selection acting in *Caenorhabditis* mitochondrial genomes might come from the noncoding elements. Our results, particularly between *C. briggsae* and *C.* sp. 9, have indicated strong purifying selection acting (Fig. 2.5). The apparent strong purifying selection present in the mitochondrial genome makes the presence of noncoding elements all the more puzzling. Noncoding mtDNA elements are thought to be disadvantageous, and thus should be eliminated from the genome (Lynch, 2006; Lynch et al., 2006). The pseudogene Ψnad5-2 in *C. briggsae* has in fact been associated with mtDNA chromosomal deletions and decreased fitness (Howe and Denver, 2008). Yet noncoding elements are present both in hermaphroditic and outcrossing species (Howe and Denver, 2008). Some, such as the conserved intergenic region present in all species and Ψnad5-1 present in *C. briggsae* and *C.* sp. 5, are even maintained over

evolutionary time (Fig. 2.2). With the selection pressure we observed, these should be targeted for elimination.

While some of these elements appear to be evolving more neutrally than the protein-coding genes, others appear to be under purifying selection (Fig. 2.7), indicating that they may have a functional role. This seems most plausible for the conserved intergenic region, as it is present in all *Caenorhabditis* species and has a lower π estimate than most other, younger noncoding elements. In any case, that we have a signature of purifying selection on the mitochondrial genomes of *Caenorhabditis* and the presence of these noncoding elements suggests that other forces might be counteracting selection. Aside from selection based on a functional role there is the possibility that the mtDNA is acting in a selfish manner (Hastings, 1992; Galtier et al., 2009). This would not be unheard of for mtDNA (MacAlpine et al., 2001). In *C. briggsae*, we in fact have evidence for selfish behavior at Ψnad5-2 (Clark et al., in press). In mutation accumulation experiments, they observed a tendency for deletion levels to stay the same or increase, never decrease. This would seem to indicate that the levels seen in the populations are maintained by a balance of purifying selection and selfish drive.

The mitochondria plays a crucial role, generating up to 95% of the energy in a cell (stated in da Fonseca et al., 2008). Because of its important nature, disfunction in the mitochondria could be quite detrimental. Indeed, it has been implicated as having a role in many diseases (Wallace, 2008). Purifying selection acts to maintain the

function of the mitochondrial genome, and our analysis indicates that this is the primary selective pressure acting in these genomes. While our analysis cannot discount that positive selection has occurred within *Caenorhabditis*, we believe it cannot be disentangled from the effects of mutational saturation.

Figure 2.1. Phylogeny of the Caenorhabditis genus.
Phylogenetic analysis of available *Caenorhabditis* species based on slowly-evolving nucleotide sites. Sites considered for this analysis included both first and second codon positions of protein coding genes, and the large and small ribosomal RNA subunits. Bootstrap support values for the maximum likelihood (1000 replicates) are indicated to the left of each node, and were calculated using the TN93 model (Tamura and Nei, 1993) with the gamma distribution and invariant sites.

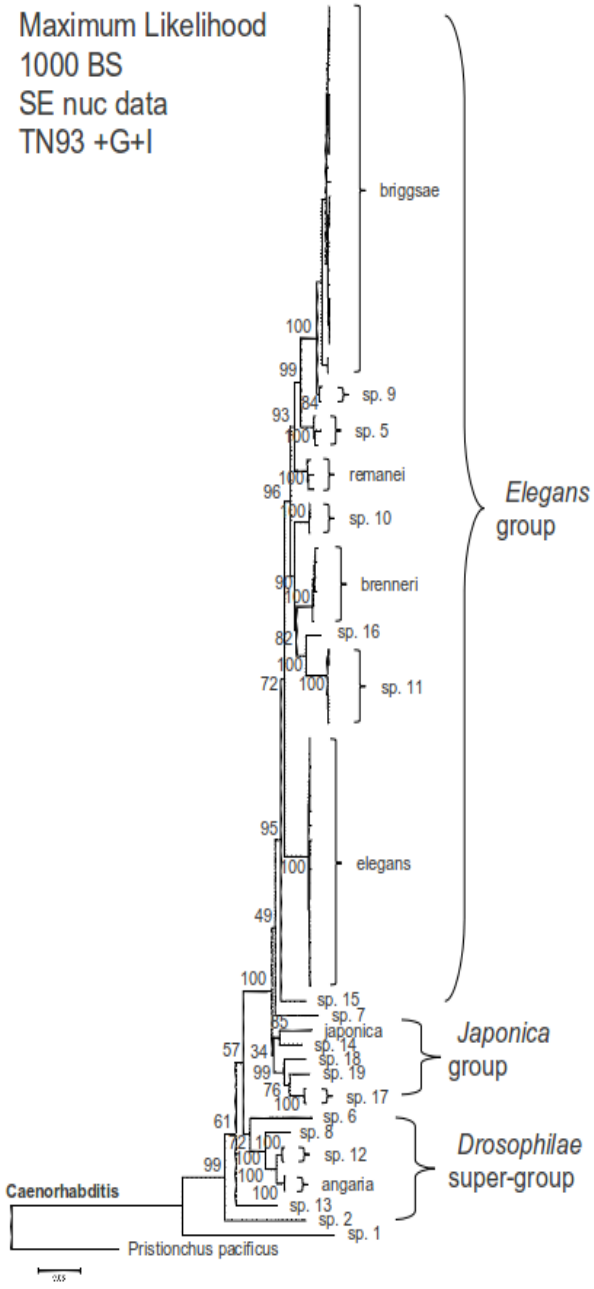


Figure 2.1. Phylogeny of the Caenorhabditis genus.

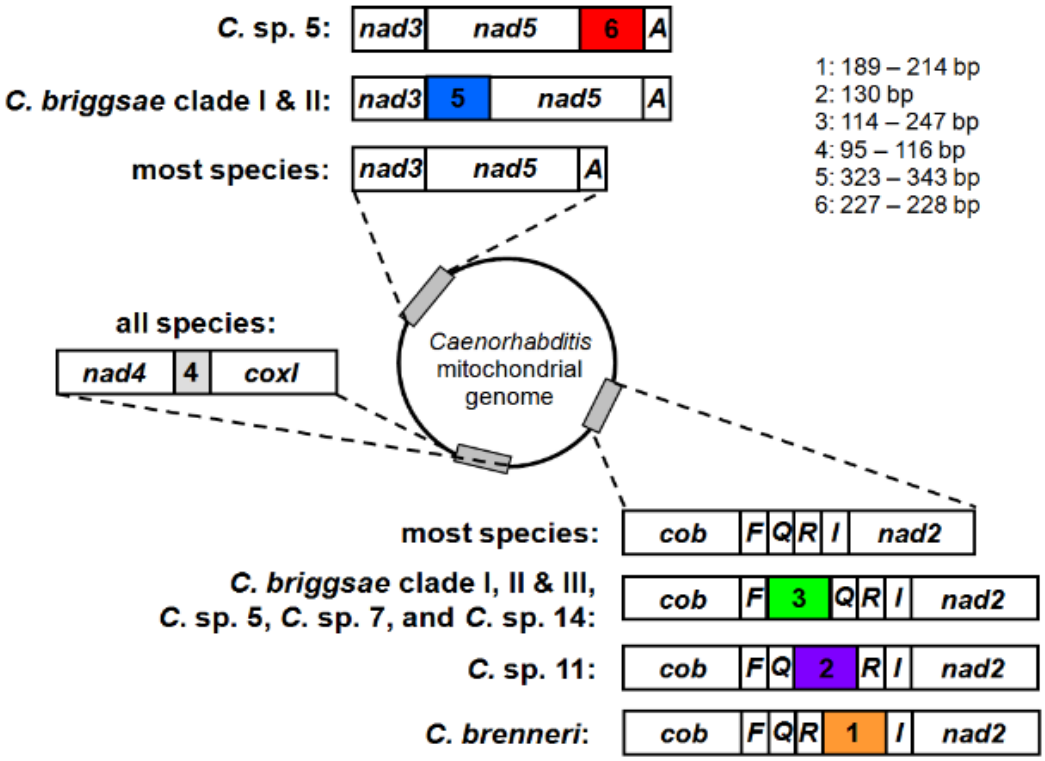


Figure 2.2. Noncoding elements in Caenorhabditis mtDNA.
Schematic shows the relative positions for the various noncoding elements found throughout the genus *Caenorhabditis* in the mitochondrial genome. Elements are numbered based on their positions relative to the control region.

Figure 2.3. Genome-wide polymorphism in Caenorhabditis species.
Scanning-window analysis of nucleotide polymorphism across the entire mitochondrial genome for species for which we have three or more representatives. Approximate locations of putative pseudogene elements shown as colored bars.

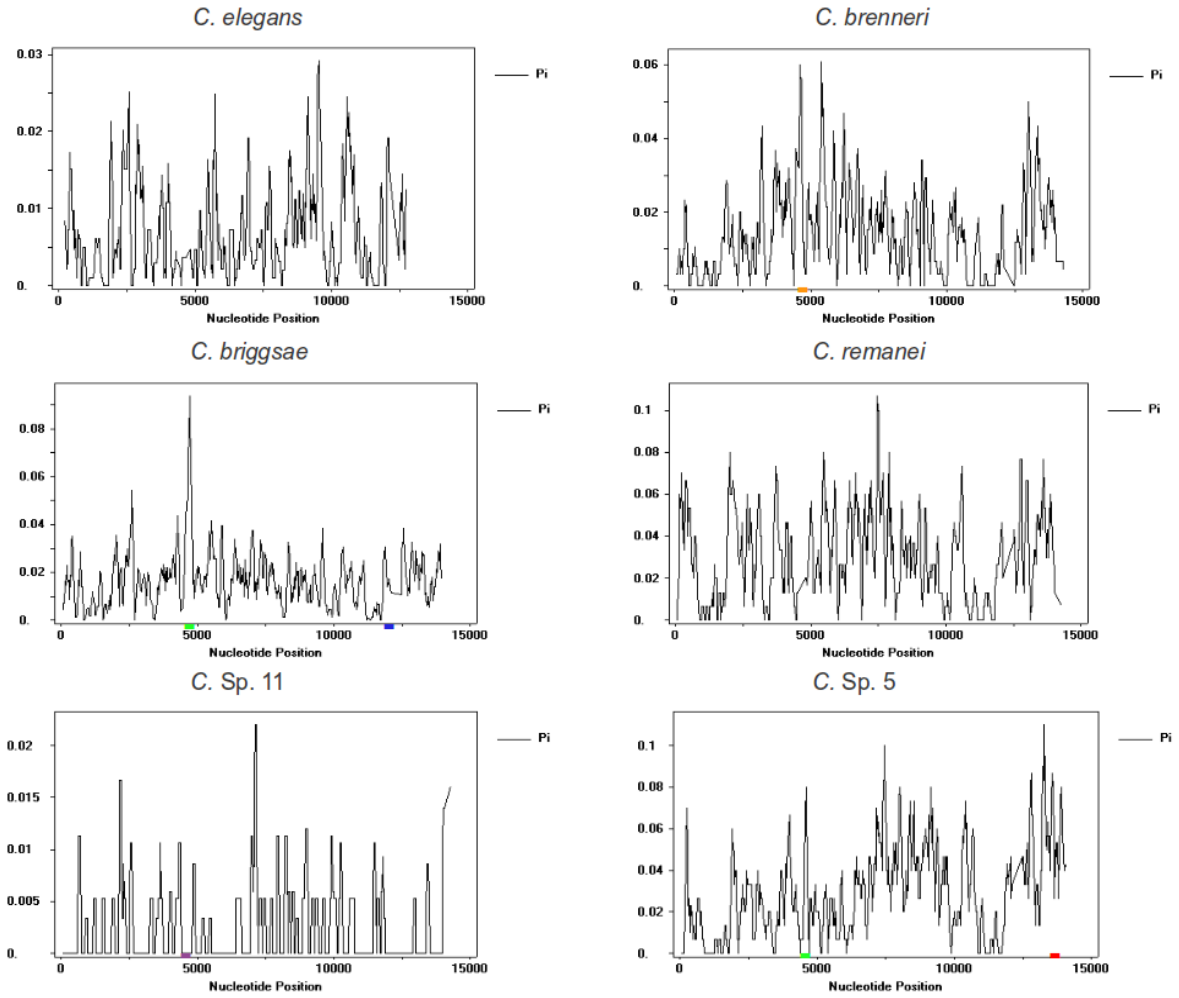


Figure 2.3. Genome-wide polymorphism in Caenorhabditis species.

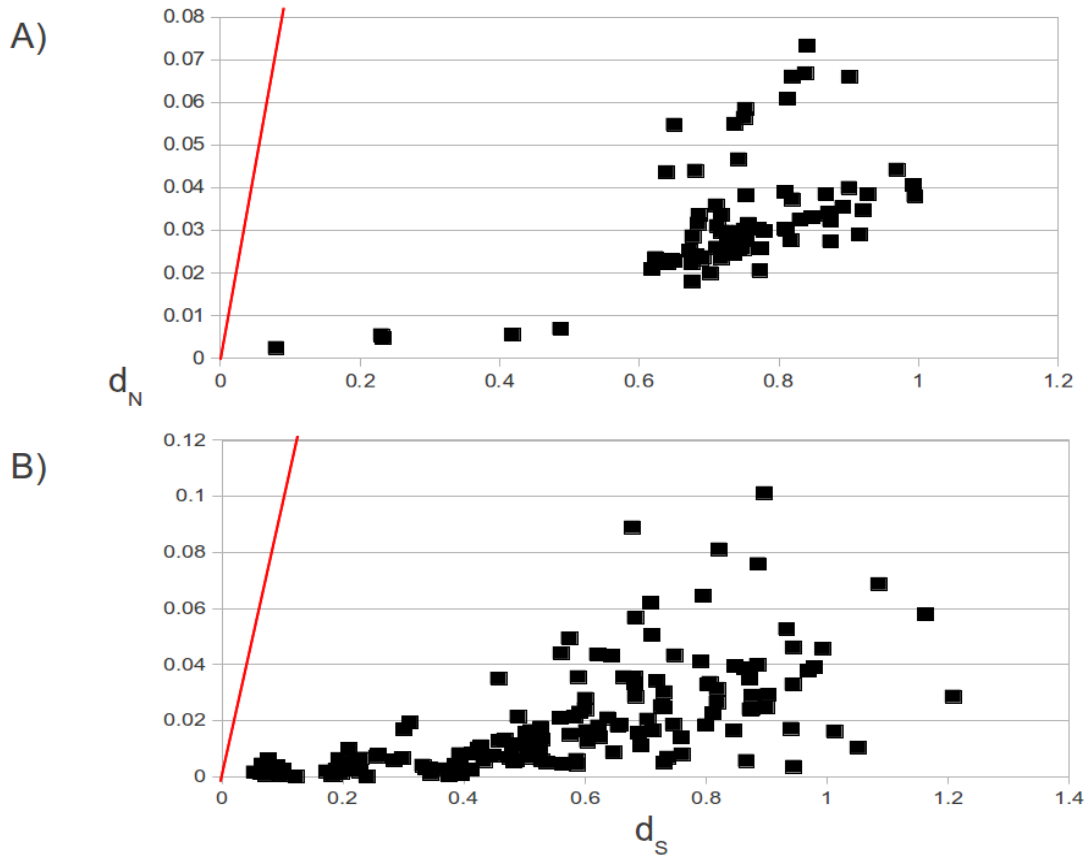


Figure 2.4. Synonymous and nonsynonymous polymorphism.
Scatter plot shows synonymous and nonsynonymous divergence values. (A) d_N and d_S values were calculated across the concatenated sequence of all 12 protein-coding genes. This was performed for all pairwise comparisons within: *C. briggsae* (between clades), the *Elegans* group, the *Japonica* group, and the *Drosophilae* supergroup. (B) d_N and d_S values were calculated gene by gene . Species in this set of comparisons included from *C. briggsae* out to *C.* sp. 15, with each species being compared to the nearest outgroup on the phylogenetic tree. Also included were the comparisons of *C. angaria* to *C.* sp. 12 and all pairwise comparisons between *C. briggsae* clades. In both (A) and (B) the red line indicates a d_N/d_S ratio of 1.

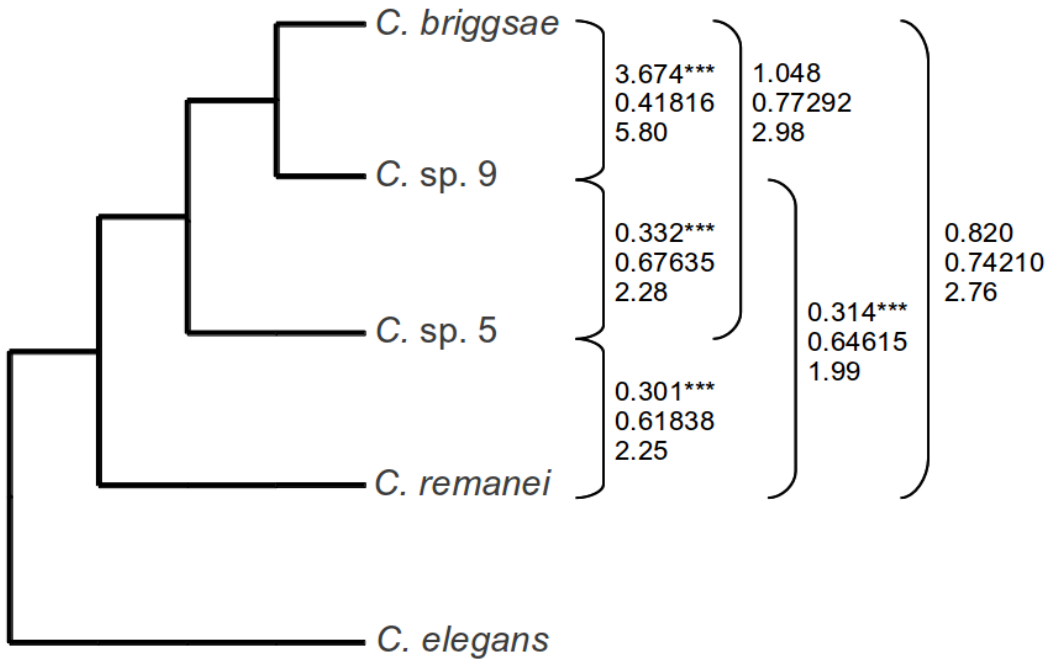


Figure 2.5. Population genetic statistics of close C. briggsae relatives.
A simplified phylogeny of the close relatives of *C. briggsae* with *C. elegans* as an outgroup. The numbers indicate the neutrality index, d_S (Jukes-Cantor corrected), and the transition/transversion ratio (ordered from top to bottom), for each pairwise comparison indicated by brackets. Four comparisons yielded a significant neutrality index (*** is indicative of $P < 0.001$).

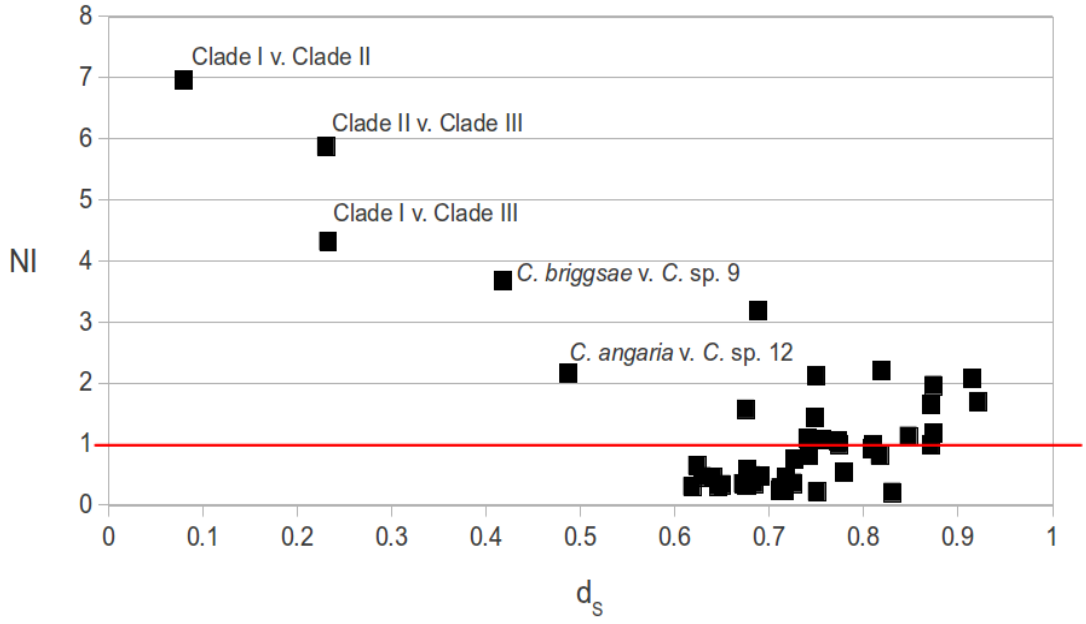


Figure 2.6. The neutrality index as a function of synonymous divergence.
The neutrality index plotted against d_S for the concatenated sequence of all 12 protein-coding genes. Statistics were calculated for all possible pairwise comparisons between *C. briggsae* clades, within the *Elegans* group, and between *C. angaria* and *C.* sp. 12. The red line indicates a neutrality index of 1. NI values above 1 indicate purifying selection, while values below 1 indicate positive selection.

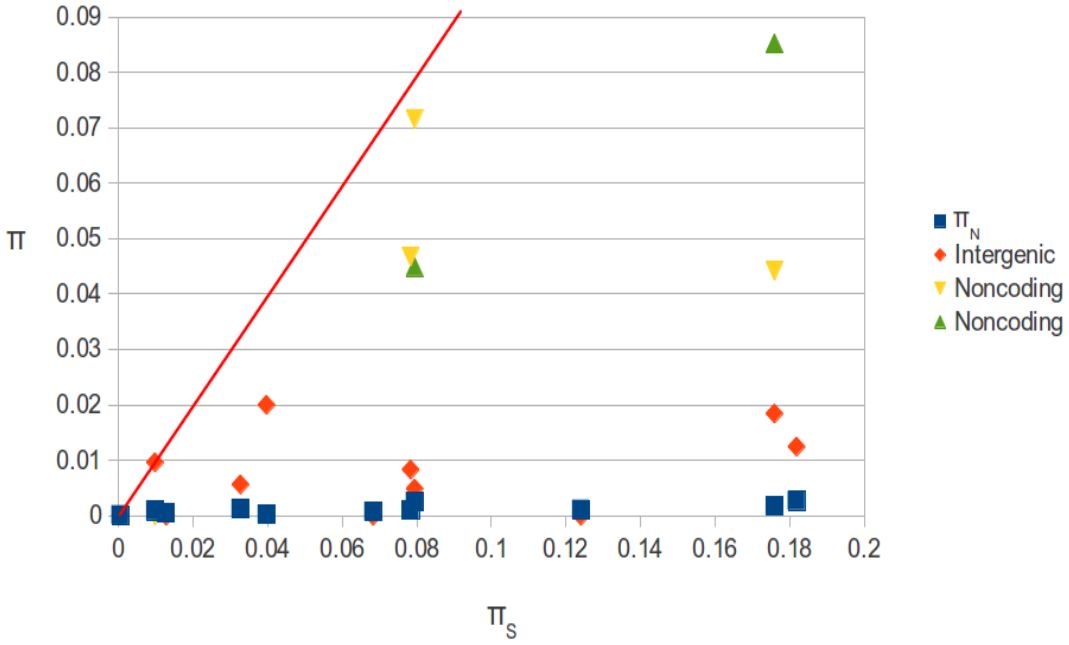


Figure 2.7. Polymorphism at noncoding elements as compared to silent sites. Within species comparison of various polymorphism measures against synonymous polymorphism. π_S serves as an approximation of neutral evolution. Comparisons made for every species for which two or more representatives are available. Polymorphism measures included: π_N for the concatenated sequence of all 12 protein-coding genes, π for the conserved intergenic element present in all *Caenorhabditis* species, and π for putative pseudogene elements. *C. briggsae* and *C.* sp. 5 each contain two putative pseudogene elements. The second element is in these species is represented by a green triangle. The red line indicates a π_N/π_S ratio of 1, indicative of neutral evolution.

Chapter 3

Evolution of *Caenorhabditis* mitochondrial genome pseudogenes and *C. briggsae* natural isolates

Michael J. Raboin, Ashley F. Timko, Dana K. Howe, Marie-Anne Félix and Dee R. Denver

Published

Molecular Biology and Evolution

May 2010, Vol. 27, No. 5, pp. 1087-96

Oxford University Press
2001 Evans Road
Cary, NC 27513

Abstract

Although most metazoan mitochondrial genomes are highly streamlined and encode little noncoding DNA outside of the "AT" region, the accumulation of mitochondrial pseudogenes and other types of noncoding DNA has been observed in a growing number of animal groups. The nematode species *Caenorhabditis briggsae* harbors two mitochondrial DNA (mtDNA) pseudogenes, named Ψnad5-1 and Ψnad5-2, presumably derived from the *nad5* protein-coding gene. Here we provide an in-depth analysis of mtDNA pseudogene evolution in *C. briggsae* natural isolates and related *Caenorhabditis* species. Mapping the observed presence and absence of the pseudogenes onto phylogenies suggests that Ψnad5-1 originated in the ancestor to *C. briggsae* and its recently discovered outcrossing relative species *Caenorhabditis* sp. 5 and *Caenorhabditis* sp. 9. However, Ψnad5-1 was not detected in *Caenorhabditis* sp. 9 natural isolates, suggesting a lineage-specific loss of this pseudogene in this species. Our results corroborated the previous finding that Ψnad5-2 originated within *C. briggsae*. The observed pattern of mitochondrial pseudogene gain and loss in *Caenorhabditis* was inconsistent with predictions of the tandem duplication-random loss model of mitochondrial genome evolution and suggests that intralineage recombination-like mechanisms might play a major role in *Caenorhabditis* mtDNA evolution. Natural variation was analyzed at the pseudogenes and flanking mtDNA sequences in 141 geographically diverse *C. briggsae* natural isolates. Although phylogenetic analysis placed the majority of isolates into the three previously

established major intraspecific clades of *C. briggsae*, two new and unexpected haplotypes fell outside of these conventional groupings. Ψnad5-2 copy number variation was observed among *C. briggsae* isolates collected from the same geographic site. Patterns of nucleotide diversity were analyzed in Ψnad5-1 and Ψnad5-2, and confidence intervals were found to overlap values from synonymous sites in protein-coding genes, consistent with neutral expectations. Our findings provide new insights into the mode and tempo of mitochondrial genome and pseudogene evolution both within and between *Caenorhabditis* nematode species.

Introduction

The mitochondrial genomes of animals are generally minimized in size, usually encoding the same set of 13 protein-coding and 24 structural RNA genes, with little duplicated or noncoding DNA outside of the control (a.k.a. "AT") region (Lynch et al., 2006; Gissi et al., 2008). Accumulation of noncoding DNA is thought to have a negative impact, increasing the susceptibility of these genomes to certain mutation types such as repeat-associated deletions (Townsend and Rand, 2004; Howe and Denver, 2008). The high mutation rates of animal mitochondrial DNA (mtDNA) are thought to provide an evolutionary environment prohibitive to the accumulation of noncoding DNA in animal mitochondrial genomes (Lynch et al., 2006). Some incidences of mtDNA noncoding elements preserved across long evolutionary timescales have been attributed to selection for their continued maintenance for

hypothesized functional reasons (Kumazawa et al., 1996; Bakke et al., 1999) or secondary structure features that prevent their loss (McKnight and Shaffer, 1997). Despite their overall rarity, pseudogenes have been described in a number of animal mitochondrial genomes including certain species of amphibians (Mueller and Boore, 2005; San Mauro et al., 2006; Kurabayashi et al., 2008), reptiles (Macey et al., 2004; Fujita et al., 2007), fish (Mabuchi et al., 2004) and nematodes (Tang and Hyman, 2007; Howe and Denver, 2008). Although in most of these cases the pseudogenes are short sequences derived from transfer RNAs (tRNAs), nematode species such as *Caenorhabditis briggsae* have been found to contain relatively large (>100 bp) pseudogenes derived from protein-coding genes (Howe and Denver, 2008). Most previous studies involving mtDNA pseudogenes have been centered on understanding their association with mitochondrial genome rearrangement events. Much remains to be understood about the evolution of animal mtDNA pseudogenes themselves, particularly within species.

C. briggsae is a hermaphroditic species in the *Elegans* group of *Caenorhabditis* nematodes (Kiontke and Fitch, 2005) and has been developed as a comparative development, genomics and evolution model for *C. elegans* (Stein et al., 2003; Cutter et al., 2006b; Zhao et al., 2008). *Caenorhabditis briggsae* natural isolates have a worldwide distribution characterized by distinct latitudinal phylogeographic population structure with isolates falling into three major clades (Cutter et al., 2006b; Dolgin et al., 2008). Although the deeper relationships of major

C. briggsae intraspecific clades are well-understood, knowledge on more local patterns of genetic diversity and phylogeography in this species remains enigmatic. *Caenorhabditis* sp. 5, a gonochoristic species found only in southeast Asia thus far, is one of many *Caenorhabditis* nematode species discovered over the last few years as a consequence of a renewed push to identify more species in the genus and a closely related sister species to the famous model *C. elegans*. *Caenorhabditis* sp. 9 is another recently discovered gonochoristic species that is a close relative to *C. briggsae*.

Large noncoding elements have been described in the mitochondrial genomes of *C. briggsae* and *Caenorhabditis* sp. 5 (Howe and Denver, 2008), providing an excellent model system to examine animal mtDNA pseudogene evolution. The two pseudogenes present in the mitochondrial genome of *C. briggsae* are both derived from the *nad5* protein-coding gene (88-89% identity between each pseudogene and *nad5*) and were named Ψnad5-1 and Ψnad5-2 (Howe and Denver 2008). Note that we are here switching to the more commonly used, standardized mtDNA gene naming scheme to facilitate easier comparisons to other mtDNA studies; Ψnad5-1 and Ψnad5-2 are the same elements that Howe and Denver (2008) referred to as ΨND5-1 and ΨND5-2, respectively. These mtDNA pseudogene elements are homologous to a central region of the *nad5* gene. Ψnad5-1 (214-223 bp) is present in all three intraspecific clades of *C. briggsae* as well as *Caenorhabditis* sp. 5, while Ψnad5-2 (325-344 bp) is only present in two *C. briggsae* intraspecific clades. Ψnad5-1 is located between *trnQ* and *trnF*, whereas Ψnad5-2 is located directly upstream of *nad5*,

on the opposite side of the mitochondrial genome (fig. 1). Direct repeats in Ψnad5-2 and *nad5* are associated with large, function-disrupting, and deleterious *nad5* gene deletions in the mtDNA of *C. briggsae* natural isolates (Howe and Denver, 2008). The *nad5* deletions are heteroplasmic, and isolate-specific levels vary from zero to ~50% of the total mtDNA pool within a nematode. Howe and Denver (2008) also described the presence of putative compensatory alleles in the Ψnad5-2 direct repeat of some isolates that renders the Ψnad5-2 repeat an imperfect match to the downstream repeat in *nad5* – isolates with these alleles showed significantly lower *nad5* deletion levels than those without the compensatory mutations. Although the putative compensatory alleles were initially identified exclusively in one *C. briggsae* intraspecific subclade (the "temperate" clade), the Howe and Denver (2008) study considered only 24 *C. briggsae* natural isolates. There are currently >100 geographically diverse *C. briggsae* natural isolates available for study – the patterns of mtDNA pseudogene conservation and distribution patterns of the putative Ψnad5-2 compensatory alleles in the larger set of *C. briggsae* isolates, and *Caenorhabditis* sp. 9, remains unknown. Furthermore, the patterns of intraspecific molecular genetic diversity of mtDNA pseudogenes have not been well studied in *C. briggsae* or any other animal species.

In the present study, we analyze the evolution of mtDNA pseudogenes and flanking sequence in 141 global natural isolates of *C. briggsae*, as well as related *Caenorhabditis* species. We report on the gain, conservation, and loss of pseudogenes in *C. briggsae* and its close relatives *Caenorhabditis* sp. 5 and *Caenorhabditis* sp. 9.

The pseudogene and flanking coding sequences are also used to reconstruct phylogenetic relationships among the nematodes analyzed. mtDNA pseudogene nucleotide diversity (π) is analyzed among different *C. briggsae* clades and compared to diversity patterns in mtDNA protein-coding sequences. Our analysis reveals mtDNA pseudogene gain and loss events in *Caenorhabditis*, high levels of molecular genetic diversity in *C. briggsae* mtDNA pseudogenes, new insights into global and local patterns of *C. briggsae* genetic diversity, and divergent *C. briggsae* mtDNA haplotypes that suggest substantial genetic diversity remains to be discovered in this species.

Materials and Methods

Nematode isolate culturing and species diagnosis

Natural geographic isolates of 141 *C. briggsae*, 5 *Caenorhabditis* sp. 5, and 2 *Caenorhabditis* sp. 9 were collected and examined (Table S3.1). Isolates were obtained from our own collection efforts and as gifts from helpful colleagues (see Acknowledgements). All isolates were expanded in culture, prepared for cryogenic storage, and collected for DNA extraction using standard techniques (Wood, 1988; Denver et al., 2003). Species diagnoses in *Caenorhabditis* are generally performed by testcrossing unknown hermaphrodites or females with males of a known species (Barrière and Félix, 2006); although certain crosses of *Caenorhabditis* sp. 9 and *C. briggsae* isolates yield viable progeny (Félix M.-A., unpublished data), they are

considered different species due to strong hybrid sterility and their different reproductive modes.

Polymerase chain reaction and DNA sequencing

Polymerase chain reaction (PCR), product purification, and direct DNA sequencing were performed as previously described (Denver et al., 2003; Howe and Denver, 2008). Table S3.2 provides all PCR and sequencing primers used for this study. For all isolates analyzed here, two mtDNA regions were amplified: the first was a ~810 bp amplicon containing *trnQ*, Ψnad5-1, *trnF* and the 5' end of *cob;* the second was a ~1,150 bp amplicon containing the 3' end of *nad3,* Ψnad5-2, and the 5' end of *nad5*. A single nuclear locus, primarily composed of the third intron of the *Cbr-polh-1* gene (DNA polymerase eta homolog) and ~850-1,000 bp in length (depending on isolate), was also amplified for comparative purposes with the mtDNA data. For *Caenorhabditis* sp. 5 and *Caenorhabditis* sp. 9, one additional region was analyzed for phylogenetic analysis: an ~400 bp segment composed of a small subunit ribosomal RNA gene segment, *trnS(ucn)*, *trnN*, and *trnY*. Sequence data were submitted to GenBank under accession numbers (GU451901-GU452323).

Phylogenetic and nucleotide diversity analyses

Multiple alignment and phylogenetic analyses were performed using the MEGA4 software package (Tamura et al., 2007). DNA sequence multiple alignments were

performed using the ClustalW function of MEGA4; the IUB DNA weight matrix was used and the gap opening and extension penalties were set to 15 and 6.66, respectively (default settings). Reliabilities of resultant multiple alignments were evaluated by visual inspection after running ClustalW. For a first phylogenetic analysis involving multiple nematode species, amplified regions from *Caenorhabditis* sp. 5 and *Caenorhabditis* sp. 9 were aligned with the homologous sequences from 1 *C. remanei* (strain EM464; complete mtDNA sequence provided as a gift from W. K. Thomas) and 24 *C. briggsae* sequenced mitochondrial genomes (Howe and Denver, 2008); 1,995 bp of aligned DNA sequence was used. For the first phylogenetic analysis, concatenated gene sequences were used and pseudogene sequences were excluded. The mtDNA sequence from *C. remanei* strain EM464 was used as the outgroup based on results from a previous phylogenetic analysis of *Caenorhabditis* using nuclear genes (Kiontke et al., 2007). For a second phylogenetic analysis involving only *C. briggsae* isolates, alignments were performed for the two amplified mtDNA regions that included pseudogene sequences – the two regions were concatenated for subsequent analysis (1,694 bp in the alignment). Redundant haplotypes were removed prior to phylogenetic analysis. In MEGA4, neighbor-joining (NJ) and maximum parsimony methods were employed for phylogenetic analyses. For NJ analyses, the maximum composite likelihood model of sequence evolution was implemented. For maximum parsimony analyses, the close-neighbor interchange method was used for tree searching. All substitution types were included in all phylogenetic analyses. The

reliabilities of tree topologies resulting from phylogenetic analyses were evaluated using bootstrap testing (1,000 replicates performed for each analysis).

π values were calculated for mitochondrial protein-coding sequences and both pseudogenes using DnaSP v4.0 (Rozas et al., 2003). Protein-coding sequence flanking Ψnad5-1 and Ψnad5-2 (portions of *cob*, *nad3*, and *nad5*) was concatenated in-frame – this data set was used to calculate nonsynonymous- and synonymous-site nucleotide diversity (π_N and π_S, respectively); 327 codons were analyzed. Populations were defined according to clades observed from the phylogenetic trees. Approximate 95% confidence intervals (CIs) for π estimates were determined in DnaSP 4.0 using coalescent simulations, following the methods reported in a previous analysis of mtDNA and nucleotide diversity in *Caenorhabditis* nematode species (Graustein et al., 2002). The simulations computed π value estimates and CIs given input theta estimates (values empirically calculated from corresponding data sets in DnaSP) and the assumption of no recombination. Ten thousand replicates were performed for each simulation.

Results and Discussion

Gain, conservation, and loss of Caenorhabditis mtDNA pseudogenes

To examine the evolutionary origins and fates of pseudogene sequences in the mitochondrial genomes of *Caenorhabditis* nematodes we analyzed two mtDNA regions: one containing Ψnad5-1 (in *C. briggsae* and *Caenorhabditis* sp. 5) and a

second containing Ψnad5-2 (in *C. briggsae*) – see figure 1. mtDNA analysis of crosses between divergent natural strains of *C. elegans* showed that mtDNA is inherited exclusively through the hermaphrodite oocyte lineage in this species (Morris K., Thomas W.K., personal communication). Thus, we applied a phylogenetic approach to investigate the evolution of *Caenorhabditis* mtDNA pseudogenes under the assumptions that mtDNA was inherited through the hermaphrodite lineages and that there is no heterologous mtDNA recombination. To construct a reference phylogeny for mapping pseudogene gain and loss events in *Caenorhabditis* (Fig. 3.2), 1,995 bp of aligned mtDNA protein-coding, tRNA, and ribosomal RNA gene sequence was analyzed from 24 isolates of *C. briggsae*, 5 isolates of *Caenorhabditis* sp. 5, and 2 isolates of *Caenorhabditis* sp. 9. Ψnad5-1 and Ψnad5-2 sequences were excluded from these analyses. We also included homologous sequence from *C. remanei* strain EM464 mtDNA as an outgroup sequence based on a previous phylogenetic analysis of *Caenorhabditis* nuclear DNA (Kiontke et al., 2007). Our phylogeny was consistent with the previous nuclear analysis (Kiontke et al., 2007) and *Caenorhabditis* sp. 9 was placed as a closely related sister taxon to *C. briggsae* (Fig. 3.2). Natural populations of *C. briggsae* have been subdivided into three major clades in previous analyses of both nuclear DNA and mtDNA (Dolgin et al., 2008; Howe and Denver, 2008). Here, we propose and apply a new, simplified, and generic clade naming scheme for *C. briggsae*: Clade I (previously "tropical" clade, contains AF16 reference nuclear genome strain), Clade II (previously "temperate" clade) and Clade III (previously

"equatorial" and "Kenya" clade). We propose this new scheme because the present study reveals a greater degree of latitudinal overlap between *C. briggsae* clades than was observed in previous smaller scale studies (discussed further below). *Caenorhabditis* sp. 5 is a recently-discovered, unnamed, outcrossing species found only in southeast Asia thus far (see Table S3.1) that is more closely related to *C. briggsae* than *C. remanei* and *C. elegans* (Kiontke et al., 2007; Cutter, 2008). *Caenorhabditis* sp. 9 is an even more recently discovered, unnamed, outcrossing species that is capable of crossing with *C. briggsae* to make fertile offspring. Two isolates of this species are currently available: one from India and the other from The Democratic Republic of The Congo (Table S3.1).

We found that Ψnad5-1 is present in all surveyed *C. briggsae* and *Caenorhabditis* sp. 5 isolates but was not found in either *Caenorhabditis* sp. 9 isolate. This pseudogene is also not present in *C. remanei* or *C. elegans* mitochondrial genomes (Denver et al., 2003; Howe and Denver, 2008). By contrast, Ψnad5-2 is present only in Clade I and II of *C. briggsae* and absent in all other surveyed *Caenorhabditis* species, suggesting a single origin event within *C. briggsae*. Given our results, we infer that an absence of Ψnad5-1 and Ψnad5-2 was the ancestral mitochondrial genome architecture state in *Caenorhabditis*. We mapped the Ψnad5-1 presence/absence data onto our phylogeny (Fig. 3.2), which revealed two possible paths of Ψnad5-1 evolution. First, it is possible that there were two independent origins of Ψnad5-1 on the phylogeny: one gain in *C. briggsae* and a separate gain in

Caenorhabditis sp. 5. A second possibility is that there was a single origin of Ψnad5-1 in the ancestor to (*C. briggsae* + *Caenorhabditis* sp. 9 + *Caenorhabditis* sp. 5) and a loss in *Caenorhabditis* sp. 9. We observed that Ψnad5-1 occupies identical junctions between $tRNA^{Q}$ and $tRNA^{F}$ in both *C. briggsae* and *Caenorhabditis* sp. 5. Although independent duplications of the same sequence are not unheard of in the mitochondrial genome (Fujita et al., 2007), it is unlikely that such events would share the same junctions. We thus conclude that the most likely explanation for the observed pattern of Ψnad5-1 presence/absence is a single gain in the ancestor to (*C. briggsae* + *Caenorhabditis* sp. 9 + *Caenorhabditis* sp. 5), with a subsequent loss in the *Caenorhabditis* sp. 9 branch. *Caenorhabditis* sp. 9 was found to possess 6 bp of unassignable intergenic sequence between $tRNA^{Q}$ and $tRNA^{F}$, where Ψnad5-1 resides in *C. briggsae* and *Caenorhabditis* sp. 5. Although we cannot determine its origin with certainty, it is possible that this small intergenic spacer is all that remains of Ψnad5-1 in *Caenorhabditis* sp. 9.

The probable loss of Ψnad5-1 from *Caenorhabditis* sp. 9, but not other lineages, raises questions regarding the evolutionary forces acting on these pseudogenes. Differences in effective population size may provide an apt explanation because this factor determines the relative strengths of selection versus drift (Charlesworth, 2009). However, there are currently too few isolates of *Caenorhabditis* sp. 5 and *Caenorhabditis* sp. 9 to estimate effective population size for these species. Differences in mating system (*C. briggsae* is hermaphroditic whereas

the two unnamed species are gonochoristic) might also be hypothesized to have an effect, though Ψnad5-1 is present in both *C. briggsae* and *Caenorhabditis* sp. 5. Lower mtDNA mutation rates in species with mtDNA pseudogenes relative to those without the elements provide another possible contributing factor. Although the mtDNA mutation process in *C. elegans* has been well-characterized (Denver et al., 2000), little is known about mtDNA mutation in other nematode species. The evolutionary reasons why Ψnad5-1 was eliminated entirely from one species and preserved in two others, one hermaphroditic and one gonochoristic, remain unknown.

The tandem duplication-random loss (TDRL) model of mitochondrial genome evolution (Boore and Brown, 1998) posits that tandem duplication of mtDNA stretches followed by random loss (via pseudogenization then degeneration) of redundant genes is responsible for between-lineage differences in mtDNA gene order. The TDRL model has gained wide empirical support, primarily in deuterostomes, with the observation of requisite "transition state" duplicated and pseudogenized sequences in a number of test cases (Pääbo et al., 1991; Arndt and Smith, 1998; Yu et al., 2008). For the *Caenorhabditis* mtDNA pseudogenes (homolgous to a central region of *nad5*), however, no instances of longer "precursor" pseudogene regions were detected for either Ψnad5-1 or Ψnad5-2 in any of the *Caenorhabditis* species and strains analyzed here. Ψnad5-1 occurs in a tRNA cluster on the opposite side of the mitochondrial genome relative to its presumed coding gene ancestor (*nad5*); a very large mtDNA duplication event involving 19 genes would be required to explain the presence of

Ψnad5-1 under the TDRL model. Furthermore, mtDNA gene order is identical in ancestral species lacking Ψnad5-1 (e.g. *C. remanei*, *C. elegans*) and those harboring the pseudogene (*C. briggsae*, *Caenorhabditis* sp. 5). Thus, the large duplication event required to invoke TDRL would also have to be followed by a duplicate gene pseudogenization/loss process that resulted in a mtDNA gene order identical to the ancestral order – highly unlikely considering the involvement of 19 duplicated genes. Although Ψnad5-2 is directly upstream of *nad5* which on the surface might seem to make the TDRL model more applicable to this element as compared with Ψnad5-1, Ψnad5-2 originated within *C. briggsae* (Fig. 3.2) and no natural isolates of this species show any evidence of a longer, more intact pseudogene copy predicted by the TDRL model. More specifically, if Ψnad5-2 originated as a consequence of a tandem duplication, sequences homologous to the 5' region of *nad5* would be a required evolutionary intermediate that we would expect to observe in some *C. briggsae* strain.

We hypothesize that the sudden and punctuated appearance and disappearance of these *Caenorhabditis* mtDNA pseudogene elements are driven by intra- and/or intergenomic mtDNA recombination-like processes that occur within hermaphrodite oocyte lineages – molecular exchange between nonallelic direct repeat sequence tracts from different molecules in the population of mitochondrial genomes within an organelle might lead to the gain or loss of sequences in a given mtDNA molecule. mtDNA recombination of this type has been documented in other nematodes species (Lunt and Hyman, 1997) but not in *Caenorhabditis*. Further studies targeting

predicted recombination intermediates (e.g., mtDNA subgenomic circles) will be required to evaluate the potential role of recombination-like mechanisms in shaping mtDNA evolution in *Caenorhabditis* and other animal taxa.

Natural mtDNA variation in C. briggsae

In addition to analyzing pseudogene gain and loss events in the genus *Caenorhabditis*, we used the pseudogene and flanking sequences to investigate within-species mtDNA variation in 141 geographically diverse *C. briggsae* isolates (Table S3.1). Phylogenetic relationships of *C. briggsae* hermaphrodite lineages were analyzed using a combined sequence set of 1,694 bp containing Ψnad5-1; Ψnad5-2; and sequences from *nad3*, *nad5*, *trnQ*, *trnF* and *cob*. These regions were amplified and sequenced for 117 isolates and combined with homologous regions from published mitochondrial genome sequences for 24 additional *C. briggsae* isolates (Howe and Denver, 2008). Among the 141 isolate-specific sequences analyzed, 44 unique haplotypes were observed and used in subsequent phylogenetic analyses. The three major intraspecific clades of *C. briggsae* identified in previous studies (Dolgin et al., 2008; Howe and Denver, 2008) were again identified here with strong bootstrap support (Fig. 3.3). Furthermore, Ψnad5-1 was observed in all *C. briggsae* natural isolates, whereas Ψnad5-2 was present only in isolates falling into the (I + II) superclade, also consistent with our previous analysis of 24 isolates (Howe and Denver, 2008). The worldwide geographic distribution patterns of the growing collection of *C. briggsae*

natural isolates are available on an online GoogleMap resource we have made with help from our colleague A. D. Cutter (Supplementary note 3.1).

Our analysis of mtDNA pseudogene regions in 141 *C. briggsae* natural isolates enabled the study of intraspecific patterns of genetic diversity in this species at both global and more local scales and revealed new and unexpected *C. briggsae* haplotypes and isolate groupings. One unexpected finding was that a new set of isolates from southern India were placed in Clade III, whereas in previous studies only equatorial isolates from Kenya composed this group (Dolgin et al., 2008; Howe and Denver, 2008). This observation suggests that Clade III isolates have a wider geographic range than previous smaller scale studies suggested. Furthermore, both Clade III and Clade I isolates were identified at the same sites in Ponmudi, Kerala, India (Table S3.1), whereas in previous studies, only isolates from a single major instraspecific clade were observed at any given location.

Although most haplotypes were placed within the three previously established clades of *C. briggsae*, two haplotypes, each unique to one isolate, were found to fall outside of these groupings. Our bootstrap consensus phylogenies placed isolate QR24 (Quebec, Canada) inside the (I + II) superclade but not within either Clade I or Clade II (Fig. 3.3). NJ and maximum parsimony analyses yielded conflicting results in terms of the placement of QR24, though both arrangements were poorly supported by bootstrap analysis (Figs. S3.1 and S3.2). Similarly, isolate JU1424 (northeast Vietnam) was placed inside the (I + II) superclade but not within either Clade I or

Clade II; NJ and maximum parisimony again yielded conflicting and poorly supported results. We also analyzed a nuclear region (primarily composed of the *Cbr-polh-1* third intron) in the 141 *C. briggsae* isolates. Whereas this nuclear region yielded similar results to mtDNA for QR24, JU1424 was placed within Clade I with strong bootstrap support in both NJ and maximum parsimony analyses (Fig. S3.3). The incongruous results for JU1424 suggest that this strain's nuclear genome has derived, at least in part, from a typical Clade I genotype. Both QR24 and JU1424 mitochondrial genomes contained Ψnad5-1 and Ψnad5-2. These findings suggest that substantial genetic diversity in *C. briggsae* might still remain undiscovered, especially in light of the still patchy sampling from parts of the world other than North America and Europe. The uncertain phylogenetic placement of QR24 and JU1424 in either Clade I or Clade II along with the increased extent of geographic overlap of isolates placed in different major intraspecific clades motivated the new generic *C. briggsae* clade naming scheme implemented here.

In addition to yielding new insights into the global distribution patterns of *C. briggsae* natural isolates, our study also provided the opportunity to study more localized patterns of genetic diversity in this species. France has been very well-sampled (69/141 isolates analyzed here are from this country), and a previous analysis of six nuclear loci in the *C. briggsae* natural isolates (Cutter et al., 2006b) showed that all isolates from France except one (JU516) shared identical alleles across all six loci, indicating that the France isolates are closely related to one-another. However,

knowledge on patterns of genetic variation within and among geographic sites in France has remained enigmatic. The more rapidly evolving mtDNA sequences examined here revealed 12 distinct haplotypes, all very similar to one-another (Clade II), among the 69 France isolates analyzed. Ten different geographic locations in France were surveyed (Table S3.1) and either one or two haplotypes were found at each of these locations (Fig. S3.4). Haplotypes mH32 and mH37 were each observed at three different locations; all other France haplotypes were specific to a single location in this country. However, some France haplotypes were also observed in other parts of the world: mH33 was observed in France and Japan; mH34 was observed in France and the United States. This pattern of identical mtDNA haplotypes appearing in isolates from disparate geographic locations – also observed in *C. elegans* (Denver et al., 2003) – suggests that *C. briggsae* nematodes are capable of migrating large distances, perhaps facilitated by associations with humans or other animals (Kiontke and Sudhaus, 2006).

We also discovered that two isolates from St. Joseph, MO, USA (DL0231, DL0232 – Clade II) produced much larger PCR amplicons at the Ψnad5-2 locus as compared to other Clade I and Clade II *C. briggsae* isolates. These two isolates were collected along with six other *C. briggsae* isolates from the same rotting apples sampling site – the latter six produce smaller amplicons similar to those observed in other isolates. In DL0231 and DL0232, however, DNA sequencing analysis revealed that these two isolates each contained identical tandem duplicate copies of Ψnad5-2

(referred to as Ψnad5-2a and Ψnad5-2b – see Fig. 3.1). The DNA sequence was identical in Ψnad5-2a and Ψnad5-2b, also the same as the sequence observed in the single Ψnad5-2 copies analyzed in the other six St. Joseph isolates. PCR analysis (Fig. S3.5) of the Ψnad5-2 region in DL0231 and DL0232 suggests that the duplication-bearing genome are not fixed but rather coexists in a heteroplasmic state with genomes containing single copies of Ψnad5-2 and *nad5* deletion-bearing molecules. This finding shows that the Ψnad5-2 region is subject not only to DNA deletion dynamics (Howe and Denver, 2008) but also to tandem duplication events.

Nucleotide diversity of C. briggsae mtDNA pseudogenes

Pseudogene sequences are often assumed to be nonfunctional and expected to evolve in a neutral fashion (Li et al., 1981). Although we have no *a posteriori* reasons to expect that Ψnad5-1 has evolved in a nonneutral fashion, for Ψnad5-2, there are biological reasons why we might expect selection to have affected its evolution. Direct repeats present in Ψnad5-2 and downstream in *nad5* are associated with large heteroplasmic mtDNA deletions that eliminate the 5' end of *nad5*; isolate-specific deletion levels were shown to be negatively correlated with nematode fitness (Howe and Denver, 2008). Furthermore, two putative compensatory alleles (named DRSeq2 and DRSeq3) have been characterized in *C. briggsae* natural isolates; the alleles occur in the Ψnad5-2 direct repeat copy and render the Ψnad5-2 repeat an imperfect match to the downstream *nad5* repeat; isolates with these putative compensatory alleles have

significantly lower deletion levels than those where the Ψnad5-2 and *nad5* direct repeats are perfect matches. The compensatory alleles were found exclusively in Clade II isolates in our previous small-scale study (Howe and Denver, 2008); this pattern was also observed in the current analysis of 141 *C. briggsae* isolates. The frequency of compensatory allele DRSeq2 in the Clade II isolates analyzed was 0.06 (encoded by a few isolates from North America), and the frequency of allele DRSeq3 was 0.82 (present in all the numerous isolates from France). These allele frequencies should be interpreted with caution, however, given the highly biased worldwide sampling of *C. briggsae* isolates. It is unknown why the putative compensatory alleles evolved in Clade II but not Clade I. Although our lab-based fitness assays suggested that the deletions are deleterious, it is possible that in nature the deletions are somehow beneficial to the Clade I nematodes.

To evaluate whether the *C. briggsae* mtDNA pseudogenes are evolving in a fashion consistent with neutral expectations, we analyzed π for Ψnad5-1 and Ψnad5-2 and compared the diversity estimates with nonsynonymous- and synonymous-site values (π_N and π_S, respectively) from concatenated protein-coding sequences (*nad3*, *nad5* and *cob* partial sequences). Synonymous sites in protein-coding genes are frequently used to approximate patterns of neutral molecular evolution, whereas nonsynonymous sites in mtDNA are generally under strong stabilizing selection associated with protein-coding function. We calculated π values for the four sequence categories (Ψnad5-1, Ψnad5-2, nonsynonymous sites, synonymous sites) in each of *C.*

briggsae Clade I and Clade II using DnaSP v4.0 (Rozas et al., 2003). For Clade III, Ψnad5-2 is absent and therefore not included in π calculations. Approximate 95% CIs for each π calculation were estimated using coalescent simulations in DnaSP following the approach used in Graustein et al. (2002).

Consistent with results from previous studies (Cutter et al., 2006b; Howe and Denver, 2008), we observed greater levels of genetic diversity in Clade I versus Clade II isolates regardless of the sequence category analyzed (Table 3.1), though π_S values were more similar between the clades as compared with the results of Howe and Denver (2008) that analyzed fewer isolates but nearly complete mitochondrial genome sequences. Clade II nematodes have been isolated throughout the northern temperate latitudes of North America, Europe, and Asia yet display very little genetic diversity – this observation led Cutter et al. (2006b) to hypothesize that *C. briggsae* might have started colonizing temperate latitudes in just the last few thousand years. Our study also provided the opportunity to compare patterns of genetic diversity in Clade III with that observed in the other two major intraspecific clades. π estimates were similar between Clade III and Clade I isolates at synonymous sites and Ψnad5-1 (Ψnad5-2 is not present in Clade III nematodes). When considered in light of the position of Clade III in the *C. briggsae* intraspecific phylogeny (Fig. 3.3), this result suggests that Clade III is a deeply ancestral and diverse group. We also calculated nucleotide diversity specific to the Ψnad5-1 element in the five *Caenorhabditis* sp. 5 isolates and found the π estimate to be similar to those observed in *C. briggsae* Clade I and Clade III.

If the *C. briggsae* pseudogenes have evolved in neutral or nearly neutral fashions, then genetic diversity at these loci would be expected to be similar to that observed at other sites presumed to be evolving neutrally, such as synonymous sites in protein-coding genes. The π estimates specific to mtDNA pseudogenes were similar to those observed in corresponding clade-specific π_S estimates (95% CIs overlap in all cases – see Table 3.1). As expected, the π_N estimates were all much lower than their corresponding clade-specific π_S estimates, and the pseudogene-specific π estimates, presumably due to the effects of purifying selection associated with protein-coding function. In addition to the values reported in Table 3.1, we also calculated π values for all 141 *C. briggsae* natural isolates in Ψnad5-1 (Ψnad5-2 is absent from Clade III isolates) and observed its value (π = 0.0570; 95%CI = 0.0224 – 0.1423) to be similar to the π_S value calculated from all isolates (π_S = 0.0721; 95% CI = 0.0224 – 0.1501). Given these results, we cannot rule out the hypothesis that the *C. briggsae* mtDNA pseudogenes have evolved in a mostly neutral fashion similar to synonymous sites in mtDNA protein-coding genes. However, the small sizes of Ψnad5-1 and Ψnad5-2 posed inherent limitations on the power of our analyses, and some large magnitudinal differences in the π estimates were observed that are worthy of discussion. For example, π values were greater for the pseudogenes than synonymous sites in Clade I, whereas the opposite was true for Clade II. Assuming synonymous sites are neutral, these patterns might reflect diversifying selection acting on the pseudogenes in Clade I

and purifying selection acting on the pseudogenes in Clade II. For Ψnad5-2, the compensatory alleles DRSeq2 and DRSeq3 are observed only in Clade II; thus, purifying selection to maintain the compensatory alleles might be acting in Clade II. Although we were unable to rule out neutral evolution of the pseudogenes, our analyses characterized evolutionary processes across entire aligned pseudogene regions where the potential selective effects on a few nucleotide sites (e.g. the DRSeq2 and DRSeq3 putative compensatory alleles) might have been overwhelmed by neutral evolution across the majority of sites.

Conclusion

This study provides important insights into both the between- and within-species evolution of *Caenorhabditis* mtDNA pseudogenes and shows that these elements harbor high levels of nucleotide diversity in *C. briggsae* natural populations. The punctuated appearance and disappearance of *nad5*-derived pseudogenes on the *Caenorhabditis* phylogeny are inconsistent with the widely accepted TDRL model for animal mitochondrial genome evolution and suggests that recombination-like processes might play a major role in *Caenorhabditis* mitochondrial genome evolution. The discovery of the two unusual new *C. briggsae* mtDNA haplotypes (JU1424 and QR24) suggests that there is still much to learn about the levels and sources of genetic diversity in this nematode species. We are optimistic that the ongoing discovery of new *C. briggsae* natural isolates and *Caenorhabditis* species, coupled with the

application of new DNA sequencing technologies to nematode mitochondrial genomics, will provide further insights into the evolution of *Caenorhabditis* mtDNA pseudogenes in the future.

Acknowledgements

We are grateful to M. Ailion, F. Back, D. Baïlle, A. Barrière, A. Choe, E. Dolgin, E. Hoyos, M. Leroy, V. Robert and the *Caenorhabditis* Genetics Center for providing nematode strains and to Mark Dasenko at the Oregon State University Center for Genome Research and Biocomputing for DNA sequencing support. Thanks to W. K. Thomas for providing the *C. remanei* mtDNA sequence and to A. D. Cutter for insightful comments on the manuscript and help with the *C. briggsae* GoogleMap resource. D. R. D. was supported by National Institutes of Health Grant GM087678 and M.-A. F. was supported by the Centre National de la Recherche Scientifique.

Table 3.1. π estimates for mtDNA pseudogenes and protein-coding sequences

	Ψnad5-1	Ψnad5-2	PC Syn. Sites	PC Nonsyn. Sites
Cb Clade I	0.0422 (0.0106 – 0.0781)[a]	0.0218 (0.0051 – 0.0411)	0.0151 (0.0027 – 0.0326)	0.0031 (0.0007 – 0.0089)
Cb Clade II	0.0035 (0.0004 – 0.0149)	0.0079 (0.0018 – 0.0202)	0.0115 (0.0027 – 0.0306)	0.0012 (0.0001 – 0.0043)
Cb Clade III	0.0399 (0.0120 – 0.0967)	NA[b]	0.0268 (0.0085 – 0.0728)	0.0003 (0 – 0.0026)
C. sp. 5	0.0299 (0.0057 – 0.0951)	NA	ND	ND

PC syn. sites, synonymous sites in protein-coding genes; PC nonsyn. sites, nonsynonymous sites in protein-coding genes; *Cb*, *Caenorhabditis briggsae*; *C*. sp. 5, *Caenorhabditis* sp. 5; ND, not determined.

[a] Calculated π values are indicated on top with approximate 95% CIs shown below their respective values in parentheses.

[b] NA indicates not applicable (Clade III and *C.* sp. 5 lack Ψnad5-2).

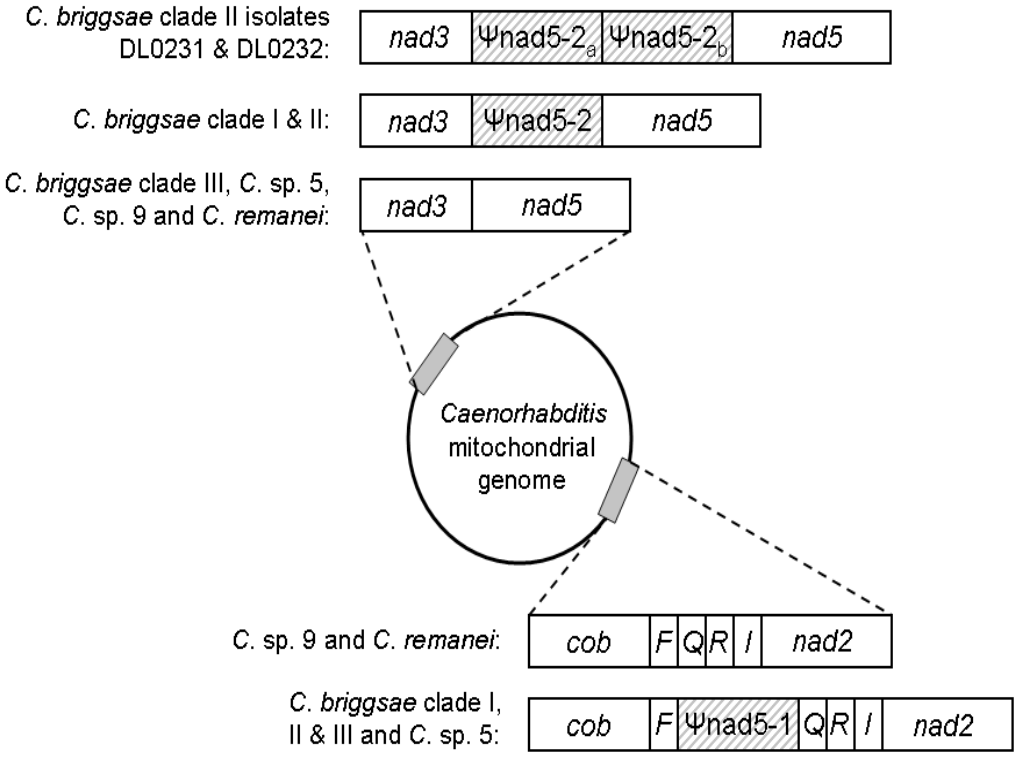


Figure 3.1. Positions of pseudogenes in Caenorhabditis mitochondrial genomes. Genes are indicated by white rectangles (single-letter abbreviation used for tRNA genes) and dashed boxes show pseudogenes. The *Caenorhabditis* species and/or intraspecific groups in which each arrangement is observed is indicated on the left of each displayed mtDNA region.

Figure 3.2. Presence and absence of mtDNA pseudogenes in Caenorhabditis. Main display is an NJ phylogram for *Caenorhabditis briggsae*, *Caenorhabditis* sp. 9, *Caenorhabditis* sp. 5, and *C. remanei* using 1,995 bp of mtDNA sequence. All gene sequences amplified for this study were used though pseudogene sequences were excluded so that their presence/absence could be independently mapped onto the phylogeny. The presence/absence of Ψnad5-1 and Ψnad5-2 in different species and intraspecific clades of *C. briggsae* is indicated on the right. Bootstrap values for maximum parsimony (left) and NJ (right) methods (1,000 replicates performed for each) are indicated to the left of the corresponding node. Scale bar shows 0.01 substitutions per site. The cladogram in the dashed box on the top left shows evolutionary relationships in the *Elegans* clade of *Caenorhabditis* nematodes based on nuclear DNA data (Kiontke et al., 2007).

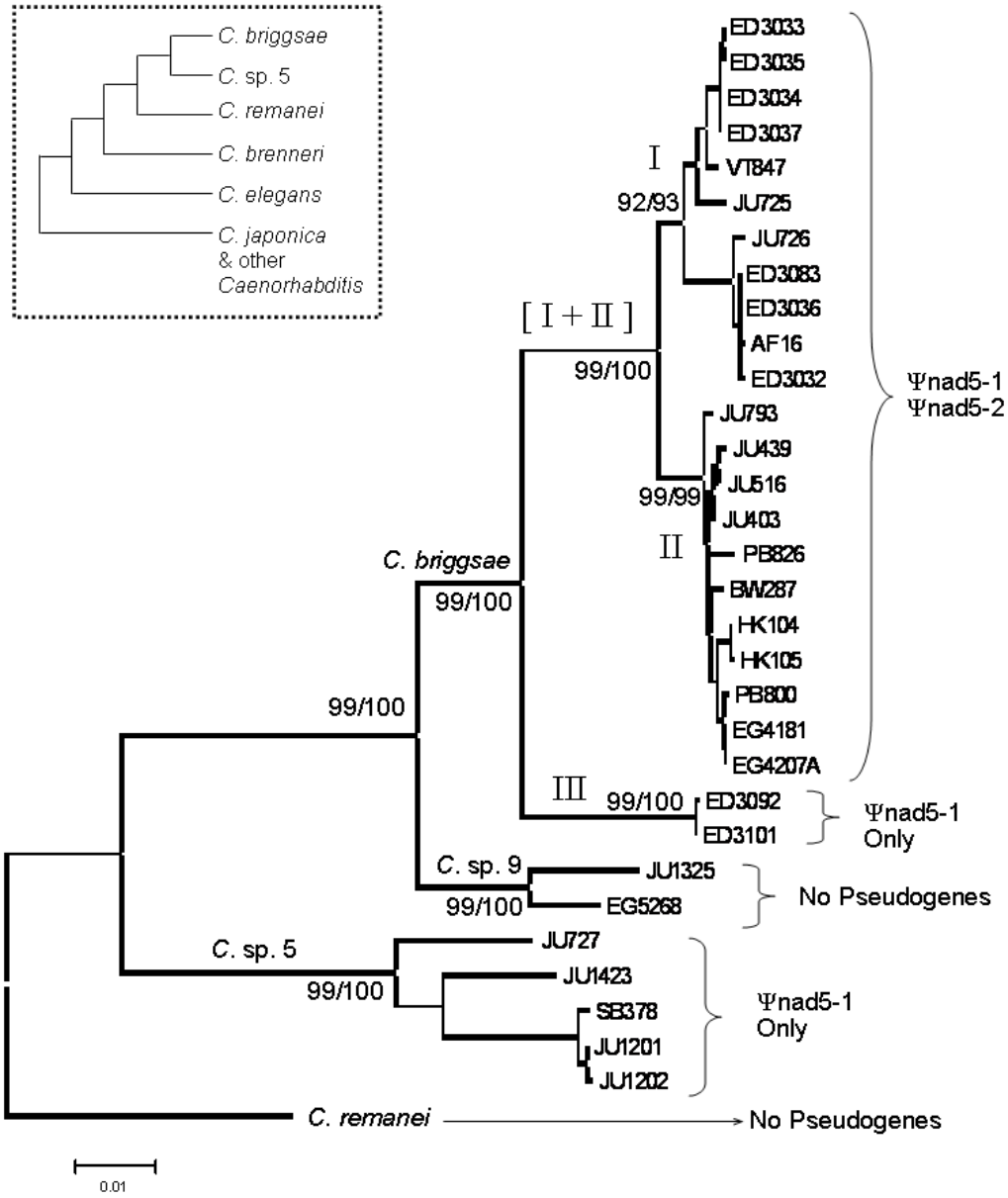


Figure 3.2. Presence and absence of mtDNA pseudogenes in Caenorhabditis.

Figure 3.3. Phylogenetic analysis of Caenorhabditis briggsae mitochondrial haplotypes.
Unrooted 80% bootstrap consensus cladogram was constructed using 1,694 bp of mtDNA, including both Ψnad5-1 and Ψnad5-2, analyzed in all 141 isolates. The 44 unique mtDNA haplotypes were used for phylogenetic analysis. Ψnad5-2 is absent in Clade III isolates – the gapped region in Clade III isolate sequences was treated as a single event in phylogenetic analyses (complete deletion parameters in MEGA4). Bootstrap values for maximum parsimony (left) and NJ (right) methods (1,000 replicates performed for each) are indicated to the left of the corresponding node. Numbers in parentheses next to haplotype designators show the numbers of isolates with that haplotype. Three-letter country code designators (United Nations designators) are also provided for general geographic information – see supplementary table 1 for more detailed information on haplotypes and geographic locations, especially for isolates collected from islands. Complete phylograms from NJ and maximum-parsimony analyses are show in supplementary Figs. S3.1 and S3.2, respectively.

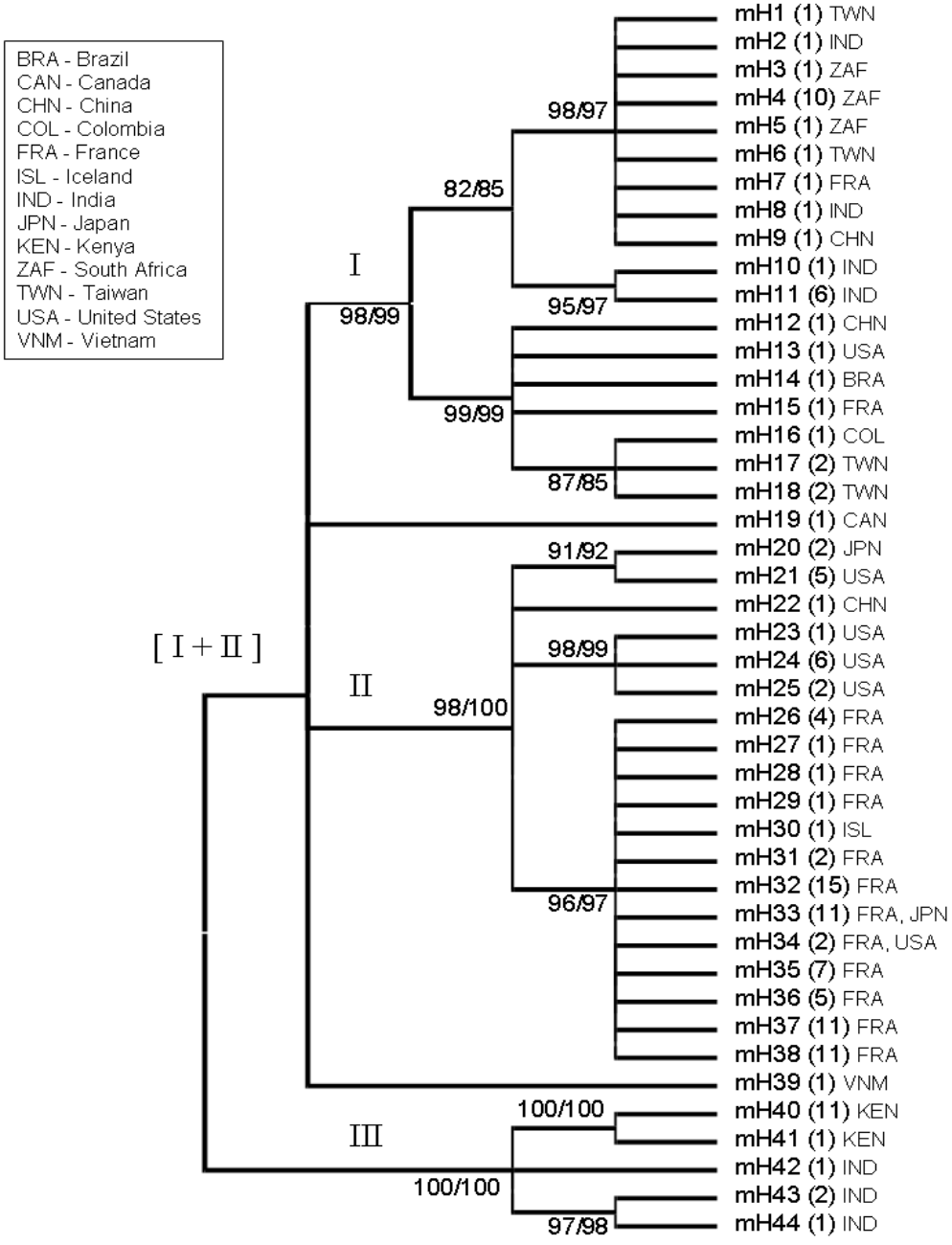


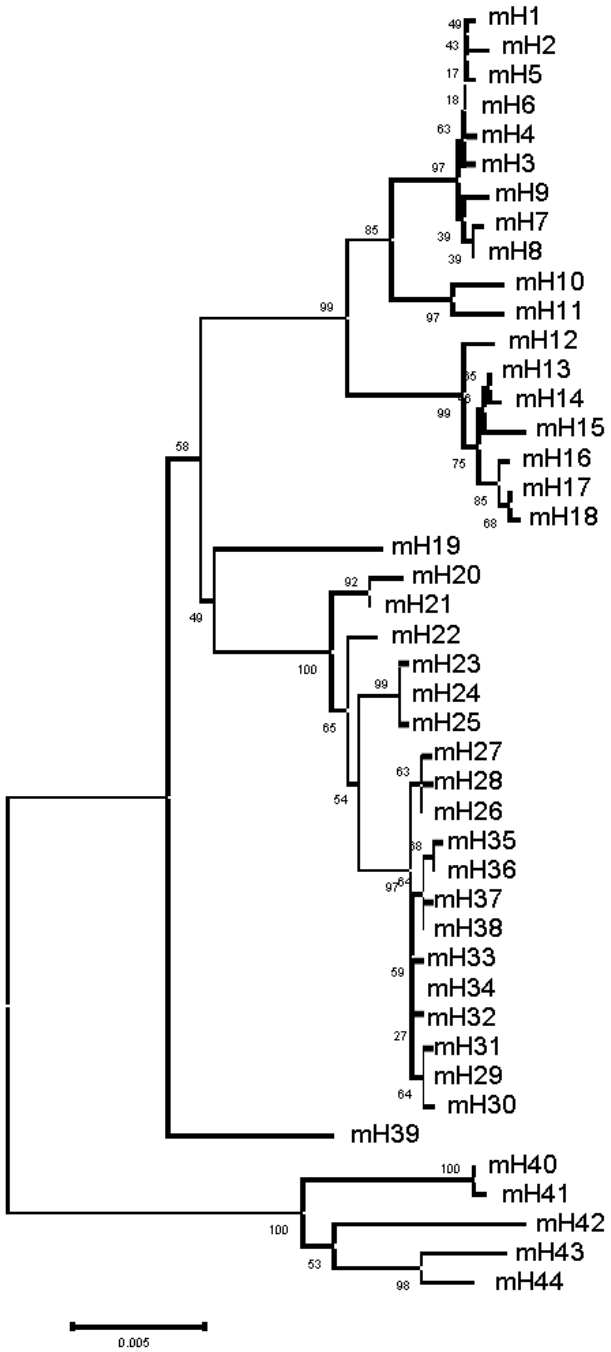
Figure 3.3. Phylogenetic analysis of Caenorhabditis briggsae mitochondrial haplotypes.

Supplementary information

Supplementary note 3.1

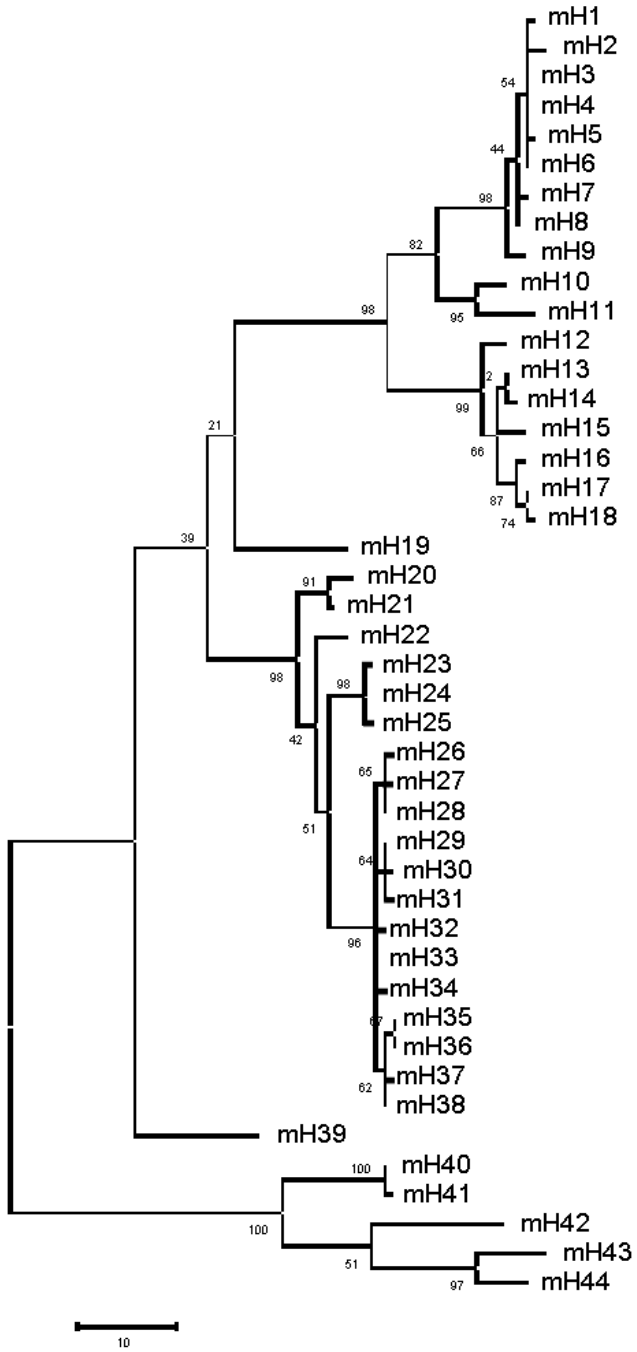
The worldwide geographic distribution patterns of the growing collection of *C. briggsae* natural isolates is available on an online GoogleMap resource we have made with help from our colleague A. D. Cutter – see URL: http://maps.google.com/maps/ms?client=firefox-a&hl=en&ie=UTF8&msa=0&msid=114127758849605788125.000470f8e70970148d3cf&z=2. The global isolate distributions on GoogleMap, however, were made approximate (i.e. accurate to the city-level area, not based on GPS coordinates) for privacy concerns as many isolates were collected at private residences. Red balloons indicate Clade I *C. briggsae* isolates, blue balloons indicate Clade II isolates and yellow balloons indicate Clade III isolates. QR24 and JU1424 – Clade [I + II] – are indicated by purple balloons.

Supplementary Figure S3.1. Neighbor-joining phylogram of C. briggsae mitochondrial haplotypes.
Unrooted bootstrap consensus tree was constructed using ~1,700 bp of mtDNA including both pseudogenes. Bootstrap values for neighbor-joining analysis (1,000 replicates) are indicated to the left of the appropriate node. Scale bar shows the number of substitutions per site.



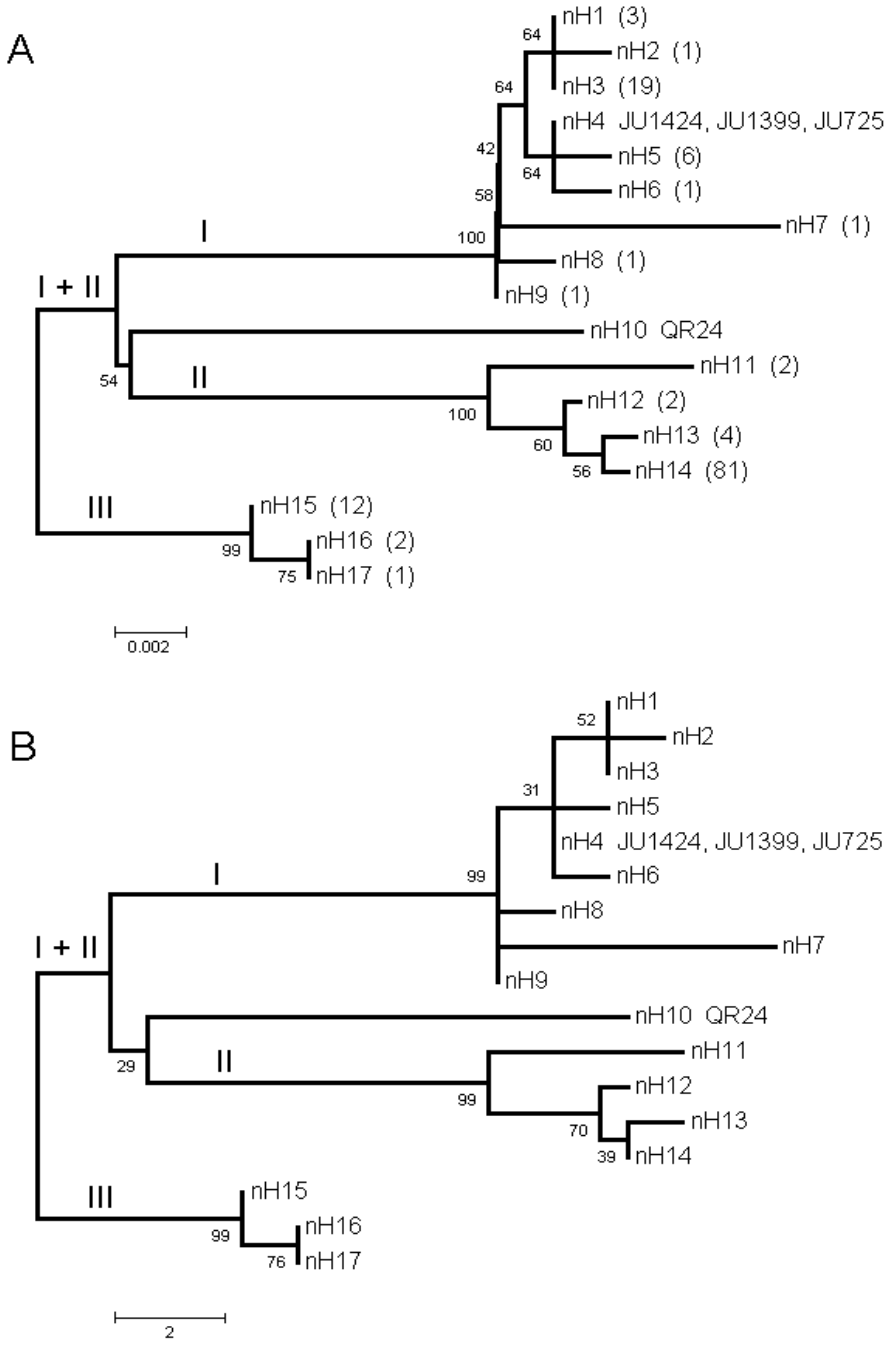
Supplementary Figure S3.1. Neighbor-joining phylogram of C. briggsae mitochondrial haplotypes.

Supplementary Figure S3.2. Maximum-parsimony phylogram of C. briggsae mitochondrial haplotypes.
Unrooted bootstrap consensus tree was constructed using ~1,700 bp of mtDNA including both pseudogenes. Bootstrap values for maximum parsimony analysis (1,000 replicates) are indicated to the left of the corresponding node. Scale bar shows the number of substitutions.



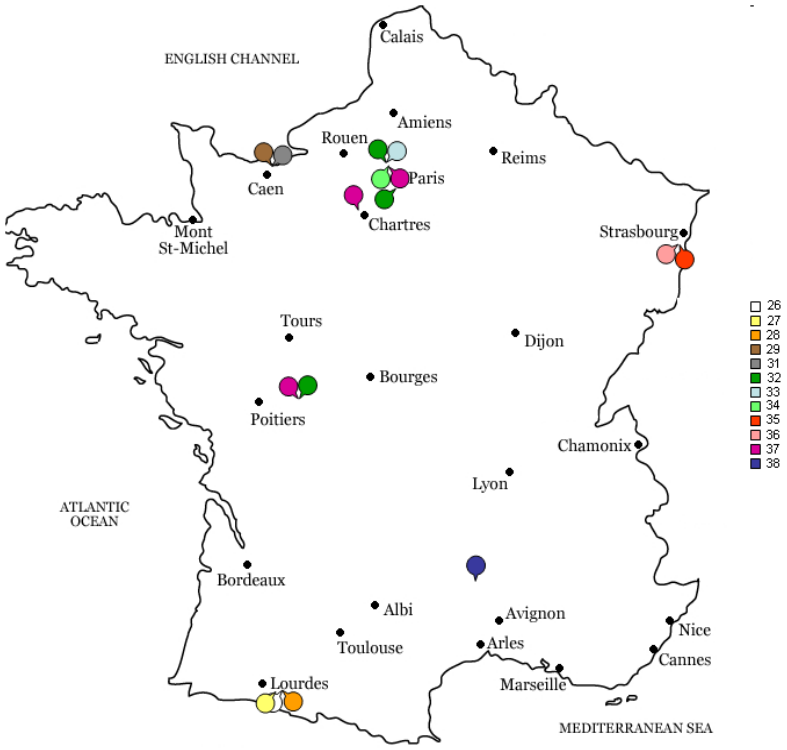
Supplementary Figure S3.2. Maximum-parsimony phylogram of C. briggsae mitochondrial haplotypes.

Supplementary Figure 3.3. C. briggsae nuclear intron phylogeny.
The third intron of *Cbr-polh-1* was amplified and sequenced in the 141 *C. briggsae* isolates, yielding 17 unique haplotypes (see supplementary table 1). Neighbor-joining (A) and maximum parisimony (B) analyses placed JU1424 (nH4) in Clade I, but QR24 (nH10) was placed within the [I + II] superclade. Bootstrap values are indicated to the left of corresponding nodes (1,000 replicates). The number of strains with each haplotype is noted on the neighbor-joining phylogeny. Scale bars show number of substitutions per site for (A) and number of substitutions for (B).



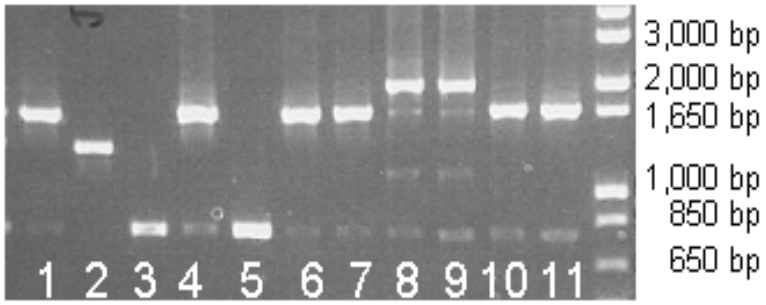
Supplementary Figure 3.3. C. briggsae nuclear intron phylogeny.

Fig. 3.4



Supplementary Figure S3.4. Geographic distributions of C. briggsae haplotypes in France.
Each colored bubble pointer indicates the site(s) where different mtDNA haplotypes were found.

Fig. S3.5



Supplementary Figure S3.5. Gel image of amplicon size variation in Ψnad5-2 region. This agarose gel image, scanned from a laboratory notebook, shows amplicon size variation at the Ψnad5-2 region in the eight *C. briggsae* isolates collected from the same site in St. Joseph, MO (begin with "DL") along with three other select isolates for comparison purposes (begin with "JU"). Amplicon sizes of the Ψnad5-2 region are shown for isolates JU1338, JU1345, JU1346, DL0227, DL0228, DL0229, DL0230, DL0231, DL0232, DL0233 and DL0234 in lanes 1-11, respectively. Lanes 1, 4, 6, 7, 10, 11 have an intact Ψnad5-2 and *nad5*, lane 2 lacks Ψnad5-2 (Clade III), lanes 3 and 5 have the deletion between 3' Ψnad5-2 and 5' *nad5*, and lanes 8 and 9 have a tandemly duplicated Ψnad5-2. Molecular marker band sizes are indicated on the right. See Howe and Denver (2008) for more details on this mtDNA region.

Table S3.1. Nematode isolates' strain names and origins.

Strain name	Origin	Clade[a]	mH[b]	nH[c]
Caenorhabditis briggsae				
AF16	Ahmedabad, India	I	2	1
BW287	Beijing, China	II	22	14
DF5100	Kruger National Park, South Africa	I	5	6
DL0227	St. Joseph, Missouri, USA	II	24	14
DL0228	St. Joseph, Missouri, USA	II	24	14
DL0229	St. Joseph, Missouri, USA	II	24	14
DL0230	St. Joseph, Missouri, USA	II	24	14
DL0231	St. Joseph, Missouri, USA	II	25	14
DL0232	St. Joseph, Missouri, USA	II	25	14
DL0233	St. Joseph, Missouri, USA	II	24	14
DL0234	St. Joseph, Missouri, USA	II	24	14
ED3032	Chungcheng, Taipei, Taiwan	I	1	3
ED3033	Tien Mu, Taipei, Taiwan	I	18	3
ED3034	Tien Mu, Taipei, Taiwan	I	17	3
ED3035	Tien Mu, Taipei, Taiwan	I	18	3
ED3036	Tien Mu, Taipei, Taiwan	I	6	1
ED3037	Tien Mu, Taipei, Taiwan	I	17	3
ED3078	Johannesburg, South Africa	I	4	3
ED3079	Johannesburg, South Africa	I	4	3
ED3080	Johannesburg, South Africa	I	4	3
ED3082	Johannesburg, South Africa	I	4	3
ED3083	Johannesburg, South Africa	I	3	3
ED3084	Johannesburg, South Africa	I	4	3
ED3085	Johannesburg, South Africa	I	4	3
ED3086	Johannesburg, South Africa	I	4	3
ED3087	Johannesburg, South Africa	I	4	3
ED3088	Johannesburg, South Africa	I	4	3
ED3089	Johannesburg, South Africa	I	4	3
ED3090	Nairobi, Kenya	III	40	15
ED3091	Nairobi, Kenya	III	40	15
ED3092	Nairobi, Kenya	III	41	15
ED3093	Nairobi, Kenya	III	40	15
ED3094	Nairobi, Kenya	III	40	15
ED3095	Nairobi, Kenya	III	40	15
ED3096	Nairobi, Kenya	III	40	15
ED3097	Nairobi, Kenya	III	40	15

ED3099	Nairobi, Kenya	III	40	15
ED3100	Nairobi, Kenya	III	40	15
ED3101	Nairobi, Kenya	III	40	15
ED3102	Nairobi, Kenya	III	40	15
EG4181	Salt Lake City, Utah, USA	II	21	13
EG4207A	Salt Lake City, Utah, USA	II	21	13
EG4360	Salt Lake City, Utah, USA	II	21	13
EG4365	Salt Lake City, Utah, USA	II	21	13
HK104	Okayama, Japan	II	20	11
HK105	Sendai, Japan	II	20	11
JU279	Jardin des Plantes, Paris, France	II	32	14
JU280	Jardin des Plantes, Paris, France	II	32	14
JU296	Jardin des Plantes, Paris, France	II	32	14
JU348	Merlet, Lagorce (Ardèche), France	II	38	14
JU349	Merlet, Lagorce (Ardèche), France	II	38	14
JU350	Merlet, Lagorce (Ardèche), France	II	38	14
JU351	Merlet, Lagorce (Ardèche), France	II	38	14
JU352	Merlet, Lagorce (Ardèche), France	II	38	14
JU353	Merlet, Lagorce (Ardèche), France	II	38	14
JU354	Merlet, Lagorce (Ardèche), France	II	38	14
JU355	Merlet, Lagorce (Ardèche), France	II	38	14
JU356	Merlet, Lagorce (Ardèche), France	II	38	14
JU357	Merlet, Lagorce (Ardèche), France	II	38	14
JU358	Merlet, Lagorce (Ardèche), France	II	38	14
JU372	Viosne Valley (Val d'Oise), France	II	33	12
JU373	Viosne Valley (Val d'Oise), France	II	33	12
JU374	Viosne Valley (Val d'Oise), France	II	33	14
JU375	Viosne Valley (Val d'Oise), France	II	33	14
JU376	Viosne Valley (Val d'Oise), France	II	33	14
JU377	Viosne Valley (Val d'Oise), France	II	33	14
JU378	Viosne Valley (Val d'Oise), France	II	32	14
JU379	Viosne Valley (Val d'Oise), France	II	33	14
JU380	Viosne Valley (Val d'Oise), France	II	33	14
JU381	Viosne Valley (Val d'Oise), France	II	32	14
JU382	Viosne Valley (Val d'Oise), France	II	33	14
JU383	Viosne Valley (Val d'Oise), France	II	33	14
JU403	Hermanville (Calvados), France	II	31	14
JU404	Hermanville (Calvados), France	II	29	14
JU405	Hermanville (Calvados), France	II	31	14
JU439	Reykjavik, Iceland	II	30	14
JU441	Beauchêne (Eure & Loir), France	II	37	14
JU516	Marsas (Hautes-Pyrénées), France	II	27	14

JU725	Yangshuo, Guangxi, China	I	12	4
JU726	Chengyang, Guangxi, China	I	9	2
JU757	Le Blanc (Indre), France	II	37	14
JU793	Frechendets (Hautes Pyrénées), France	II	26	14
JU794	Frechendets (Hautes Pyrénées), France	II	26	14
JU795	Frechendets (Hautes Pyrénées), France	II	26	14
JU796	Frechendets (Hautes Pyrénées), France	II	26	14
JU797	Frechendets (Hautes Pyrénées), France	II	28	14
JU835	Obernai (Bas-Rhin), France	II	35	14
JU836	Obernai (Bas-Rhin), France	II	36	14
JU837	Obernai (Bas-Rhin), France	II	35	14
JU838	Obernai (Bas-Rhin), France	II	35	14
JU839	Obernai (Bas-Rhin), France	II	36	14
JU840	Obernai (Bas-Rhin), France	II	36	14
JU841	Obernai (Bas-Rhin), France	II	35	14
JU842	Obernai (Bas-Rhin), France	II	35	14
JU843	Obernai (Bas-Rhin), France	II	35	14
JU844	Obernai (Bas-Rhin), France	II	36	14
JU845	Obernai (Bas-Rhin), France	II	35	14
JU846	Obernai (Bas-Rhin), France	II	36	14
JU1038	Le Blanc (Indre), France	II	32	14
JU1041	Le Blanc (Indre), France	II	32	14
JU1042	Le Blanc (Indre), France	II	32	14
JU1043	Le Blanc (Indre), France	II	32	14
JU1044	Le Blanc (Indre), France	II	32	14
JU1045	Le Blanc (Indre), France	II	32	14
JU1046	Le Blanc (Indre), France	II	32	14
JU1047	Le Blanc (Indre), France	II	32	14
JU1048	Le Blanc (Indre), France	II	32	14
JU1049	Le Blanc (Indre), France	II	32	14
JU1085	Kakegawa, Shizuoka, Japan	II	33	14
JU1205	Le Blanc (Indre), France	II	37	14
JU1254	Santeuil (Val d'Oise), France	II	34	14
JU1257	Santeuil (Val d'Oise), France	II	37	14
JU1258	Santeuil (Val d'Oise), France	II	37	14
JU1259	Santeuil (Val d'Oise), France	II	37	14
JU1260	Santeuil (Val d'Oise), France	II	37	14
JU1261	Santeuil (Val d'Oise), France	II	37	14
JU1262	Santeuil (Val d'Oise), France	II	37	14
JU1263	Santeuil (Val d'Oise), France	II	37	14
JU1264	Santeuil (Val d'Oise), France	II	37	14
JU1337	Poovar, Kerala, India	I	8	1

JU1338	Trivandrum, Kerala, India	I	11	5
JU1339	Meenmutti waterfall, Kerala, India	I	11	5
JU1340	Ponmudi, Kerala, India	I	11	5
JU1341	Ponmudi, Kerala, India	III	43	16
JU1342	Ponmudi, Kerala, India	I	11	5
JU1343	Ponmudi, Kerala, India	I	11	5
JU1344	Ponmudi, Kerala, India	I	11	5
JU1345	Ponmudi, Kerala, India	III	43	16
JU1346	Kanjirapally, Kerala, India	I	10	8
JU1347	Periyar, Kerala, India	III	42	3
JU1348	Periyar, Kerala, India	III	44	17
JU1377	Island of La Reunion	I	15	9
JU1378	Island of La Reunion	I	7	7
JU1392	Parque de Flamengo, Rio de Janeiro, Brazil	I	14	3
JU1399	Medellin, Colombia	I	16	4
JU1424	NE Vietnam	I+II	39	4
JU1564	Cold Spring Harbor, NY, USA	II	34	14
PB800	Dayton, Ohio, USA	II	21	14
PB826	Hueston Woods State Park, Ohio, USA	II	23	14
QR24	Montreal, Quebec, Canada	I+II	19	10
VT847	Hawaii, USA	II	13	3
Caenorhabditis species 5				
JU727	Sanjiang, Guangxi, China			
JU1201	Suzhou, China			
JU1202	Hangzhou, China			
JU1423	NE Vietnam			
SB378	Guagzhou, Guangdon, China			
Caenorhabditis species 9				
EG5268	Shinkolobwe, Katanga Province, Democratic Republic of Congo			
JU1325	Trivandrum, Kerala, India			

[a] See figure 3
[b] Unique haplotype from mitochondrial sequence
[c] Unique haplotype from nuclear locus analysis

Table S3.2. PCR and sequencing primers

Amplified region	Primer name	Primer sequence
ΨND5-1	CbMt_93F	GGTAAAGTTCCTTTTGGGAGA
	C5Mt_17F	AAATCCTGCTCGTTTTTGATT
	C5Mt_27F	CTAAATTTTCTAAAACTACAGGCTA
	36R	CCTCAAACTAAAACATAACC
	C5Mt_2R	TGGACCTCAAATTGGAATAACC
	C5Mt_6R	TGAAAAATACGAAAAATTCAACCA
	C5Mt_14R	TTTTGGACCGTAACCAAACA
ΨND5-2	CbMt_1F	CATTTAGTACGAAAGGAACATTGTAAA
	75F	TAAGTTTTACATAGATGTTGTA
	CbMt_12R	GCAGCCAAGAACTAAAAGGAAA
	CbMt_ND5_4R	AGACGATTAGTTAAAGCTGTATT
	58R	CTATAATTACGGCCATCTTGTTG
SrRNA	37F	GGAGTAAAGTTGTATTTAAAC
	CbMt_5F	TCAAGCTTTGTTGGATGGTG
	CbMt_17F	CCGAGTTAAATCGTGCACCT
	C5Mt_1F	TTTTTGACTCGTGTATGATCGTTT
	C5Mt_28R	GGTCCTTAATCACGCAACTG
Cbr-polh-1 intron	CBG11743b_F	AATTGTGGAGAAGGCGTCTGTGGA
	CBG11743b_R	TCTGCTGACGTGGTTTATGCCT

Primer names beginning with CbMt were designed from *Caenorhabditis briggsae* mitochondrial sequence. Primer names beginning with C5Mt were designed from *Caenorhabditis* sp. 5 mtDNA sequence. Primers with no designation before the number are from a semi-universal nematode mtDNA primer set described in:
Hu M, Chilton NB, Gasser RB. 2002. Long PCR-based amplification of the entire mitochondrial genome from single parasitic nematodes. Mol Cell Probes 16: 261-267.

Chapter 4

Competition between *Caenorhabditis briggsae* natural strains in the presence and absence of microsporidian intracellular parasites: evidence for evolutionary trade-offs

Michael J. Raboin and Dee R. Denver

Submitted

PLoS One

Public Library of Science
1160 Battery Street
Koshland Building East, Suite 100
San Francisco, CA 94111, USA

Abstract

Despite their widespread use as models for diverse biological processes, the ecology of nematodes in the genus *Caenorhabditis* is largely uncharacterized. Previous evolutionary and population genetic analyses of *C. briggsae* natural strains revealed that genetic diversity in *C. briggsae* is partitioned into roughly, three clades. The causes and consequences of this partitioning, however, have not been extensively examined, particularly in terms of inter-strain interactions, and strain-parasite interactions. To explore the effects of the biotic environment on *C. briggsae* evolution, we developed an assay whereby different natural strains of *C. briggsae* compete with one-another in a benign and a parasite-containing environment across twenty generations of evolution. Our results suggest that nematode strains vary in fitness in the parasite-containing environment, and that this fitness incurs a trade-off with relative fitness in the benign environment. Furthermore, some results of the competition assay approach used here contradict expectations based on previous measurements of *C. briggsae* strain-specific absolute fitness measures based on life history traits such as fecundity. In conclusion, our study provides a novel experimental approach for understanding relative fitness and evolution in *Caenorhabditis* nematodes; our results demonstrate environmental-based evolutionary trade-offs in the evolution of this nematode species.

Introduction

Nematodes of the genus *Caenorhabditis* have been used extensively in the investigation of numerous biological processes (Wood, 1988; Riddle et al., 1997; Hillier et al., 2005; Shim and Paik, 2010). Remarkably, however, little is known about the ecology of this group, and how different biotic and abiotic environmental factors have contributed to their evolution. In natural populations of *C. briggsae*, it is known that among-strain diversity is partitioned into three, geographically-defined clades (Cutter et al., 2006b, 2010; Dolgin et al., 2008). This population structure is hypothesized to be maintained by ecological differences (Prasad et al., 2011). Indeed, certain ecologically relevant traits also correspond to the clade structure of *C. briggsae*, particularly between strains from temperate and tropical regions. While Clade II (containing strains from temperate regions) is thought to be the youngest with the least genetic diversity (Cutter et al., 2006b), strains from this clade have longer lifespans than members of the other two clades under laboratory conditions (Joyner-Matos et al., 2009). Furthermore, fecundity among the clades responds differently to temperature. Temperate and equatorial strains (Clade II & III, respectively) possess higher relative fecundity at the lower temperature limit of fertility, while tropical strains (Clade I) possess higher relative fecundity at the upper limit (Prasad et al., 2011). The clades of *C. briggsae* represent major divisions of genetic diversity, and might also reflect divisions of ecological tolerances as well.

A potentially important ecological factor that has not been explored in *C. briggsae* is parasitism. The related model organism, *C. elegans*, has served as a model in numerous studies of host-parasite interactions (Schulenburg and Müller, 2004; Sicard et al., 2007; Pujol et al., 2008; Irazoqui et al., 2010; Schulte et al., 2010), and variation in resistance to different pathogenic bacteria has been demonstrated among *C. elegans* natural strains (Schulenburg and Ewbank, 2004; Schulenburg and Müller, 2004). Whether or not these studies are of ecological relevance, however, is debatable. Previous studies based in *C. elegans* utilized human pathogenic bacteria, or bacteria assumed in be encountered in the soil (Schulenburg et al., 2004). The first pathogens infecting natural populations of *Caenorhabditis* spp. have only recently been reported (Troemel et al., 2008; Félix et al., 2011).

The evolutionary influence of biotic factors on *Caenorhabditis* nematodes remains unclear. Isolation and growth of *Caenorhabditis* nematodes is typically performed on monoxenic cultures of *E. coli* (Barriere and Félix, 2006), however, the microbial community that these nematodes encounter is certainly more complex, including multiple types of bacteria (Grewal, 1991), parasites (Troemel et al., 2008; Félix et al., 2011), and possibly even other strains of the same nematode species (Raboin et al., 2010). This begs the question of how traits might differ between examining them in the laboratory setting, and in a more complex, ecological context. Life-history theory predicts that for an increase in one trait, such as growth in one environment, an organism will experience a trade-off in another trait (Roff, 1992).

This has been seen in many systems where the evolution of higher parasite resistance is associated with reduced fecundity in parasite-free environments (Luong and Polak, 2007; Vijendravarma et al., 2009; Schulte et al., 2010). In each of these cases, increasing fitness in the presence of a parasite resulted in decreasing fitness in its absence. This can be complicated, however, by the interactions between microorganisms, which can, in some instances, modify virulence (Nordin and Maddox, 1972; Bauer et al., 1998; Tokarev et al., 2011). Additionally, the bacteria alone, present in the environment, can affect nematodes. *C. elegans* and *C. briggsae* are known to exhibit different behaviors in response to some bacteria (Zhang et al., 2005; Schulenburg and Ewbank, 2007). Certain bacterial foodsources have even been found to influence *C. elegans* fecundity (Grewal, 1991).

The goal of this study was to examine strains of *C. briggsae* for natural variation in relative fitness, and for evidence of trade-offs between fitness in the benign, laboratory environment, and a complex, parasite-containing one. To accomplish this, we used a strain-competition approach to determine the relative fitnesses of the strains. The parasite-containing environment included, among bacterial associates, *Nematocida* sp. 1, a microsporidium first discovered in natural populations of *C. briggsae* from Kerala, India (Troemel et al., 2008). Our competition assay results provide evidence for differences between *C. briggsae* natural strains in their relative fitnesses.

Materials and Methods

Biological Materials

Seven phylogenetically diverse natural isolate-derived *C. briggsae* strains were chosen to include in the study. *C. briggsae* strains were selected specifically to cover the geographic and genetic range of each of the three major clades (Table 4.1; Fig. S4.1). Prior to the initiation of the competition plates, all strains were synchronized through hypochlorite bleaching (Wood, 1988). This process also ensured *Nematocida* sp. 1 spore infections were not present in the initial experimental populations (Troemel et al., 2008). *Nematocida* sp. 1 spores were harvested from a naturally-infected *C. briggsae* strain (JU1348) originating from India, according to the process for making infection extract previously established for *N. parisii* (Troemel et al., 2008).

Experimental conditions of competition assays

Two types of growth media were utilized during the course of the experiment: a complex media, containing a microsporidian parasite (Nematocida sp. 1) and other bacterial associates, and a standard, laboratory media. The standard, laboratory environment consisted of standard NGM agar seeded with *Escherichia coli* OP50. For parasite-containing media, infection extract stocks previously made were diluted 1:5 with an *E. coli* OP50 suspension, and seeded on NGM agar. Roughly equal numbers (hundreds) of two *C. briggsae* isolates from different clades were transferred onto NGM agar plates together (Table 4.2). For each targeted competition pair of nematode

isolate-derived strains, five replicates each were established on parasite-containing media, and five replicates on standard media. The competing populations were allowed to grow at 20ºC for 20 generations. Agar chunks containing hundreds of nematodes were transferred to fresh plates every four days using a sterilized scalpel.

Diagnostic PCR assay

We applied a diagnostic PCR assay to assess the relative abundances of different nematode strain genotypes in the experimental populations. Following 10 and 20 generations of competition between a given pair of strains, all plates were washed with M9, and the experimental nematode populations were collected. For comparison, worms were also collected in M9 from the single-strain populations used to initiate the competition replicates. DNA was extracted from these samples, and specific PCR primers were used to assay the relative abundance of the competing nematode genotypes. The primers I43b-F (5'-GGCAATTGGAAGAGTTTGTT-3') and I43b-R (5'-TCTGAAAATTTTGGGCTA-3') were used to amplify part of the third intron of the nuclear gene *Cbr-polh-1* (Raboin et al., 2010). This region displays amplicon size differences between *C. briggsae* clades, providing a marker easily scored through agarose gel electrophoresis. PCR was performed by standard methods using an annealing temperature of 50°C. PCR products produced by competition plates were run out and compared to those from the original, single-strain populations.

The amplicon banding patterns were used to score the relative prevalence of strains in each experimental population. We assessed the ability of our PCR assay to identify low-frequency *C. briggsae* genotypes by carrying out an analysis involving known-concentration, genomic DNA sample titrations. For this titration assay, DNA extracted from strains HK104 and JU1392 were adjusted to 106.1 ng/µl and 99.8 ng/µl respectively. The two DNA samples were combined in the following ratios: 1:0, 1000:1, 100:1, 10:1, 1:1, 1:10, 1:100, 1:1000, and 0:1. PCR genotyping was performed as previously described, and samples were visualized through gel electrophoresis. The experimental populations were scored at the two timepoints using the band scoring assaying, and results were interpreted according to the known-concentration titration series.

Biotic community analysis

We characterized some of the additional bacteria beyond *E. coli* OP50, present in the spore infection stock, that might influence our assay. Using an inoculating loop, we streaked bacteria for isolation from a cryogenically preserved stock of *C. briggsae* JU1348. These were allowed to grow at 20ºC for 48 hours to allow colonies to grow large enough for selection. Three colonies were selected, and PCR was performed according to standard protocols with the following modifications: a platinum wire was flame sterilized and used to transfer a small portion of a colony directly into a reaction mix, and the reaction mix was held at 95°C for 5 minutes to lyse bacterial cells before

starting to cycle temperatures. An annealing temperature of 55°C was used along with the universal 16S rDNA primers 27F (5'- AGA GTT TGA TCM TGG CTC AG -3') and 1525R (5'- AAG GAG GTG WTC CAR CC -3') (Lane, 1991).

We confirmed the presence of *Nematocida* sp. 1 spores in our *C. briggsae* JU1348 stock, and subsequent spore infection extracts, through DNA extraction and sequencing. The primers 18S3-F (5'-CGC AGC CAA ACG GGA GAC-3') and 18S3-R (5'-CCC GAT CCT CCC GGA CTC-3') were used to amplify a portion of the small subunit ribosomal RNA gene. PCR was performed by standard protocols with an annealing temperature of 60°C. The resulting PCR products were then prepared for sequencing and analysis. Amplicons were purified using a ChargeSwitch PCR Clean-Up kit (Invitrogen), and sequenced through capillary electrophoresis using a BigDye Direct Cycle Sequencing kit (Applied Biosystems). The respective PCR primers were used in the sequencing reactions. The resulting forward and reverse sequences for each amplicon were examined using MEGA5 (Tamura et al., 2011). The sequences were trimmed to remove questionable bases from the ends, and aligned manually. Forward sequences from the three colonies were then analyzed with BLASTN to confirm identity.

Results

Experimental overview

The competitive ability of one genotype versus another is at the heart of natural selection. To this end, we designed our experiment to measure the relative fitnesses of *C. briggsae* isolates in two different environments: a benign laboratory environment, and a complex, parasite-containing environment. By competing experimental populations of different genotypes against one-another, additional insights into the fitness of an organism can be gained beyond what measures of absolute fitness, such as fecundity, can express.

Examining the effects of a complex, parasite-containing environment on *C. briggsae* was accomplished through creating mixed experimental populations composed of two different natural isolate-derived strains. Thirteen different experimental populations were created from seven strains of *C. briggsae*: 6 Clade I vs. Clade II, 4 Clade I vs. Clade III, and 3 Clade II vs. Clade III. All competing strain combinations were replicated five times in each of the two environments. Populations were allowed to compete with one-another for 20 generations, and were sampled at generations 10 and 20. At these distinct time points, we assayed the relative abundance of the two genotypes to determine how the populations were evolving.

Assays consisted of PCR genotyping of a nuclear intron, and visualization through gel electrophoresis (see Materials and Methods). The PCR based assay used to identify the genotypes present within an experimental population was able to

provide a rough estimate of the relative abundance of nematode strains. Based on this assay, we are not able to distinguish between the absence of a strain from a population, and presence at low frequency (less than ~1:100; Fig. S4.2). In accordance with this, strain genotypes for each experimental population were recorded as dominant if only one genotype was observed in gel electrophoresis, indicating ~100 fold over-abundance relative to the other strain. In all other cases, both strains were recorded as still being present.

While we might assume based on previous work examining *Nematocida* infection in *C. elegans* (Troemel et al., 2008) that *Nematocida* sp. 1 is the primary cause of any negative effects observed, it may be possible that other organisms interact with the microsporidium causing a modified effect (Nordin and Maddox, 1972; Bauer et al., 1998; Tokarev et al., 2011). It is also possible that the bacteria alone are having an effect on the nematodes (Grewal, 1991; Zhang et al., 2005; Schulenburg and Ewbank, 2007). The spores used during this experiment were obtained through an extraction of *C. briggsae* isolate JU1348. The method used to extract Nematocida sp. 1 (Troemel et al., 2008) does not purify the microsporidia. The filtration method removes debris larger than 11μm (Whatman Ltd.), though most likely does not eliminate bacteria and other microorganisms present in the nematode population, which are smaller, on average (Black, 2004). As methods used in the isolation of nematodes (Barriere and Félix, 2006) also do not remove contaminant bacteria, it seems likely that some of these originated from the same soil sample that produced

JU1348, and may have well been microorganisms that this isolate of *C. briggsae* encountered in the natural environment along with *Nematocida* sp. 1. For this reason, we isolated several of the bacterial colony types present in the source population from which our *Nematocida* infection extract is derived.

Streaking the cryogenic stock for bacterial isolation revealed two different colony morphologies: one was raised with an entire margin, while the other was umbonate with an undulate margin. These colony types were identified through PCR amplification of part of the 16S ribosomal RNA gene and sequencing. DNA sequencing and BLASTN analysis indicated that the three colonies sampled belong to the genus *Paenibacillus*. The 613 bp of sequence for the 16S gene produced from the 27F primer was identical between the colony 16 and 18 samples, and contained a single polymorphism relative to colony 19. We also confirmed the presence of *Nematocida* sp. 1 spores through DNA extraction, PCR amplification with primers 18S3-F and 18S3-R, and sequencing.

Competition in a parasite-free environment

Evolving experimental populations in a benign, laboratory environment allowed us to gauge the relative fitness of *C. briggsae* strains in a parasite-free environment. Even in the absence of pressure from the parasite, allele frequencies within the experimental populations evolved rapidly. At the ten generation time point, one strain genotype had risen to dominance in 52 of 65 experimental populations (Fig. 4.2A). In numerous

populations containing strain ED3101, more than one allele was present at generation ten (ED3101 v. JU1342, 5/5; ED3101 v. JU1392, 4/5; ED3101 v. EG4181, 2/5; ED3101 v. HK104, 1/5). Additionally, a single, replicate, experimental population of the combination JU1348 v. JU1342 also displayed two alleles. Among the 13 different strain combinations, Clade II strains out-competed Clade I strains in every assay involving strains from these two clades. Clade I strains were also generally out-competed by members of Clade III. Out of 20 assays, Clade III strains became dominant in 10 populations, while the remaining 10 showed the presence of both alleles. In contrast, Clade II strains competed well against Clade III strains, rising to dominance in 12 of 15 assayss involving members of these clades. The remaining 3 experimental populations showed the presence of both alleles.

The results of our PCR assay of the generation 20 populations were largely consistent with the generation 10 results (Fig. 4.2B). At generation 20, however, 15 of 15 assays involving members of Clade II and Clade III resolved in favor of Clade II; as opposed to the 12 of 15 that was seen at generation 10. In contrast, 5 of 5 assays involving strains JU1348 and JU1342 showed the presence of both alleles at generation 20, increased from only 1 of 5 populations at generation 10. The rest of the populations showed the same results as the generation 10 assays. The generation 10 and 20 results taken together display a pattern of Clade I strains being out-competed, and Clade II strains consistently rising to dominance within populations in the standard, laboratory environment.

Competition in a parasite-containing environment

Similar to the standard laboratory environment, allele frequencies changed quickly in the complex, parasite-containing environment. Again, at the ten generation time point, the dominance of a single strain genotype was seen in 56 of 65 experimental populations (Fig. 4.2A). In contrast to what was seen in the benign, laboratory environment, Clade I strains out-competed Clade II strains in 27 of 30 experimental populations. Only in two assays containing strains EG4181 and JU1342, were both strain genotypes observed. Clade I strains also out-competed the Clade III strain JU1348 in 10 of 10 experimental populations. The Clade III strain ED3101, however, performed well in the parasite-containing environment, out-competing Clade II strains in 14 of 15 assays. In the one remaining assay (ED3101 v. EG4181), both strain genotypes were present. Against Clade I, ED3101 out-competed JU1342 in 5 of 5 assays, while 5 of 5 assays involving ED3101 and JU1392 showed the presence of both strain genotypes. Among all 13 strain combinations, none of the experimental populations experienced a Clade II strain becoming dominant in this environment.

The PCR assay of the populations at generation 20 largely confirmed what was seen at generation 10, however, several differences were observed between generations 10 and 20. At generation 20, the competition of ED3101 and JU1392 had resolved in 2 of 5 assays with ED3101 becoming dominant (Fig. 4.2B). The competition of EG4181 and JU1342 also resolved at generation 20. Previously, 2 of the 5 populations for this pair of strains showed the presence of both strain genotypes,

and one experimental population did not yield enough DNA for testing. At generation 20, the competition resolved in favor of JU1342 in 5 of 5 assays. One additional competing population deviated from observations at generation 10. In one replicate population of the EG4181 and JU1392 combination, the Clade II isolate became dominant at generation 20 where the Clade I isolate had been dominant at generation 10.

Discussion

Evolution and trade-offs

The results of competition between isolates in our experiment indicate fitness trade-offs between nematode population evolution in the complex, parasite-containing environment, and the benign, laboratory environment. Generally, the strain that performed better in the presence of the parasite, was out-competed in its absence. This general pattern was observed in all but two sets of assays (ED3101 v. JU1342 and ED3101 v. JU1392) (Figs. 4.2A & B). ED3101 had superior fitness relative to JU1342 in the parasite-containing environment. Yet, it was not out-competed in the parasite-free environment. At both time points, these experimental populations still showed the presence of both strain genotypes. This is despite the fact that relative to Clade II strains, it only had better fitness when the parasite was present. Additionally, in the competition of ED3101 v. JU1392 we were unable to assign a single strain genotype as being dominant in either environment.

It is interesting to note that while Clade II strains were consistently out-competed in the parasite containing environment, these strains were superior competitors in all cases within the standard laboratory environment (Figs. 4.2A & B). This is despite the fact that Clade II possesses little genetic diversity (Cutter et al., 2006b). A small, predominantly selfing, hermaphroditic population, such as this, is expected to be more susceptible to deleterious mutation accumulation (Lynch and Gabriel, 1990; Gabriel et al., 1993; Lynch et al., 1993), however, it seems to have superior fitness relative to both Clades I and III in benign environments. It is also of interest that the evolution of populations in response to the environment is not unidirectional, with one isolate driving the other to extinction. In several cases, our gel results from generation 20 showed the reappearance of the Clade I strain PCR band, which was not observed at generation 10. It seems likely that even in cases where PCR produces only a single amplicon, the second strain persists in the population at some low level not detected by our screening method. In one case, the dominant strain changed from Clade I to Clade II between generations 10 and 20. These results may be an indication that the second strain evolved into a stronger competitor during the course of the experiment between generations 10 and 20.

Relative vs. absolute fitness

The results of our tests of relative fitness through competition assays are not entirely intuitive in light of previous work, and highlight the context dependence of fitness.

Life history traits, such as fecundity, have been used as a proxy for fitness in studies of *C. briggsae* (eg. Cutter et al., 2010; Estes et al., 2011; Prasad et al., 2011). Fecundity and lifespan data has been published for both strains in two of our strain combinations (ED3101 v. EG4181; ED3101 v. HK104). While ED3101 has a slightly longer average lifespan than EG4181 (Estes et al., 2011, error bars overlapping), EG4181 has slightly greater average total fecundity than ED3101 at 20°C (Estes et al., 2011; Prasad et al., 2011, error bars overlapping). It is perhaps unsurprising then, that EG4181 was found to out-compete ED3101 in the absence of the parasite. Alternately, ED3101 was found to have a longer average lifespan (Estes et al., 2011) and much higher average total fecundity than HK104 at standard rearing temperature (Estes et al., 2011; Prasad et al., 2011). Despite this, HK104 was able to out-compete ED3101 in our study by generation 20 in the standard, laboratory environment. This may be an indication of local adaptation; that relative fitness correlates better with phylogeographic clade as compared to absolute fitness based on fecundity data.

C. briggsae ecology

The overall consistency of all of the results from generation 10 and generation 20, taken together, indicate an interesting pattern: in the benign environment, Clade II strains have superior fitness relative to Clade I; and in the complex, parasite-containing environment, Clade I strains have superior fitness relative to Clade II. From this, if the pattern holds in the broader set of strains, two interesting possibilities

exist: 1) the clades of *C. briggsae* might passively differ in their ecological tolerances, or 2) ecological factors might help maintain the observed population structure. The second of these possibilities has been discussed before (Prasad et al., 2011). It is tempting to speculate that the presence or absence of the parasite *Nematocida* sp. 1 may play a role as an ecological factor maintaining differentiation between *C. briggsae* clades. Adaptation corresponding with clade structure has been previously explored in terms of ethanol sensitivity (Cutter et al., 2010) and temperature tolerance (Prasad et al., 2011). Of these, local adaptation to the thermal environment has been proposed as a mechanism to shape the phylogeographic structure of *C. briggsae* (Prasad et al., 2011). The relative susceptibility of temperate isolates to the complex, parasite-containing environment provides an additional mechanism that might contribute to the phylogeographic population structure of *C. briggsae*. From the few results we have, it appears there may be a fixed difference between temperate and tropical isolates in this trait (Fig. 4.2A & B). In all cases of competition between temperate and tropical isolates, temperate isolates had higher relative fitness in the absence of the parasite, while tropical isolates had higher relative fitness when the parasite was present. This may be reflective of the range of *Nematocida* sp. 1 in nature. *Nematocida* sp. 1 was isolated from *C. briggsae* strain JU1348; a strain discovered in southern India (Troemel et al., 2008), in a region where the distributions of Clade I and Clade III strains overlap (Raboin et al., 2010). It is possible that this parasite is primarily distributed in tropical latitudes. Temperature is known to affect

spore germination, infectivity and growth in other microsporidia (Shaw et al., 2000; Martín-Hernández et al., 2009; Gisder et al., 2010). The temperate thermal environment may be unfavorable to *Nematocida* sp. 1 infection. Though it should be noted that the related microsporidian, *Nematocida parisii*, was originally discovered in the temperate environment of France (Troemel et al., 2008). Alternatively, *C. briggsae* is believed to have colonized northern latitudes relatively recently (Cutter et al., 2006b). The founding temperate population may have simply found itself in an environment free of the parasite, resulting in relaxed selection on parasite resistance.

The complicating factor to this interpretation of the results is the role of the rest of the biotic community. Other members of the biotic community can interact with microsporidia resulting in antagonism (Nordin and Maddox, 1972; Costa et al., 2011), or can increase the virulence beyond what either organism alone might cause (Bromenshenk et al., 2010; Tokarev et al., 2011). Members of the genus *Paenibacillus*, which we found to be present in our parasite infection stocks along with *Nematocida* sp. 1 microsporidia, are known to interact with some nematodes (El-Borai et al., 2005; Enright and Griffin, 2005). In addition to direct interactions, the *Paenibacillus* bacteria might also influence our competition assays by providing a nematode food source (perhaps unpreferred) present in the parasite-containing environment and absent in the standard, laboratory environment. Bacterial food sources can affect reproduction in *C. elegans* (Grewal, 1991). How bacterial members of the biotic community might interact with *C. briggsae* and/or *Nematocida* sp. 1 to

produce the results observed is a question for further investigation and may help to define the nature of the trade-offs seen between these two environments.

Acknowledgements

This work was supported by NIH grant GM087628 to D. R. D. and S. Estes. We thank S. Estes, D. K. Howe, and K. A. Clark for helpful discussions about the manuscript.

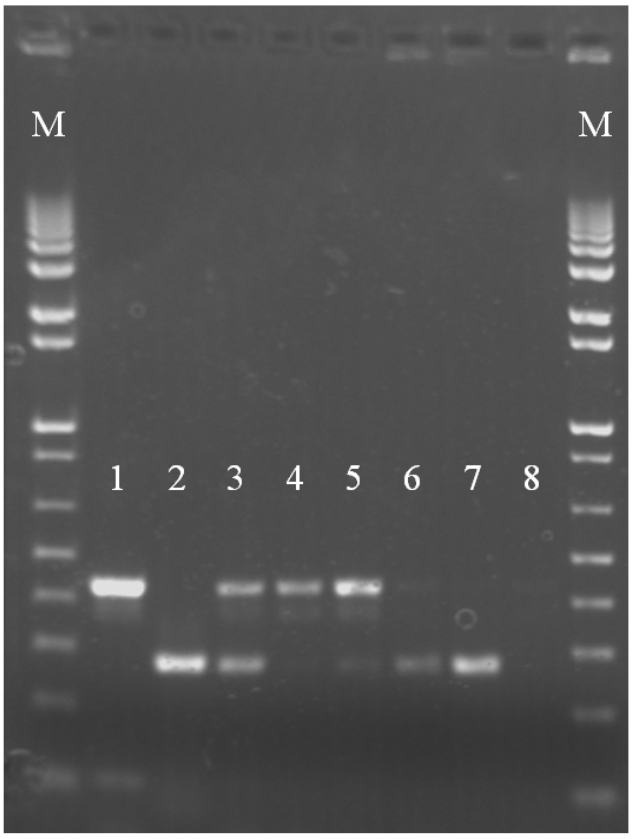


Figure 4.1. PCR analysis of the Cbr-polh-1 third intron.
Lanes 1 & 2 contain HK104 and ED3101 control DNA respectively. Lane 3 contains a HK104/ED3101 mixed, control DNA sample. Lanes 4-7 show the results from several experimental populations containing HK104 and ED3101 at generation 20. Lanes 4 & 5 were produced from DNA samples extracted from populations that evolved in the standard, laboratory environment, whereas Lanes 6 & 7 were from DNA taken from populations grown in the parasite-containing environment. Under the scoring scheme used, lanes 4, 6 & 7 would be scored as a single band present, while lane 5 would be considered both bands still present.

Figure 4.2. Dominant alleles seen in experimental populations across timepoints. Populations assayed at generations (A) 10 and (B) 20. Populations grown in the standard, laboratory environment are shown in the bottom left, while those grown in the presence of the parasite are in the top right. Red indicates the dominance of the Clade I strain, blue the Clade II strain, and yellow the Clade III strain. Hatched cells indicate that PCR bands for both strains were observed indicative of less than ~100 fold difference between the abundance of the two strains. The white cells indicate cases where there is no information due to PCR failures.

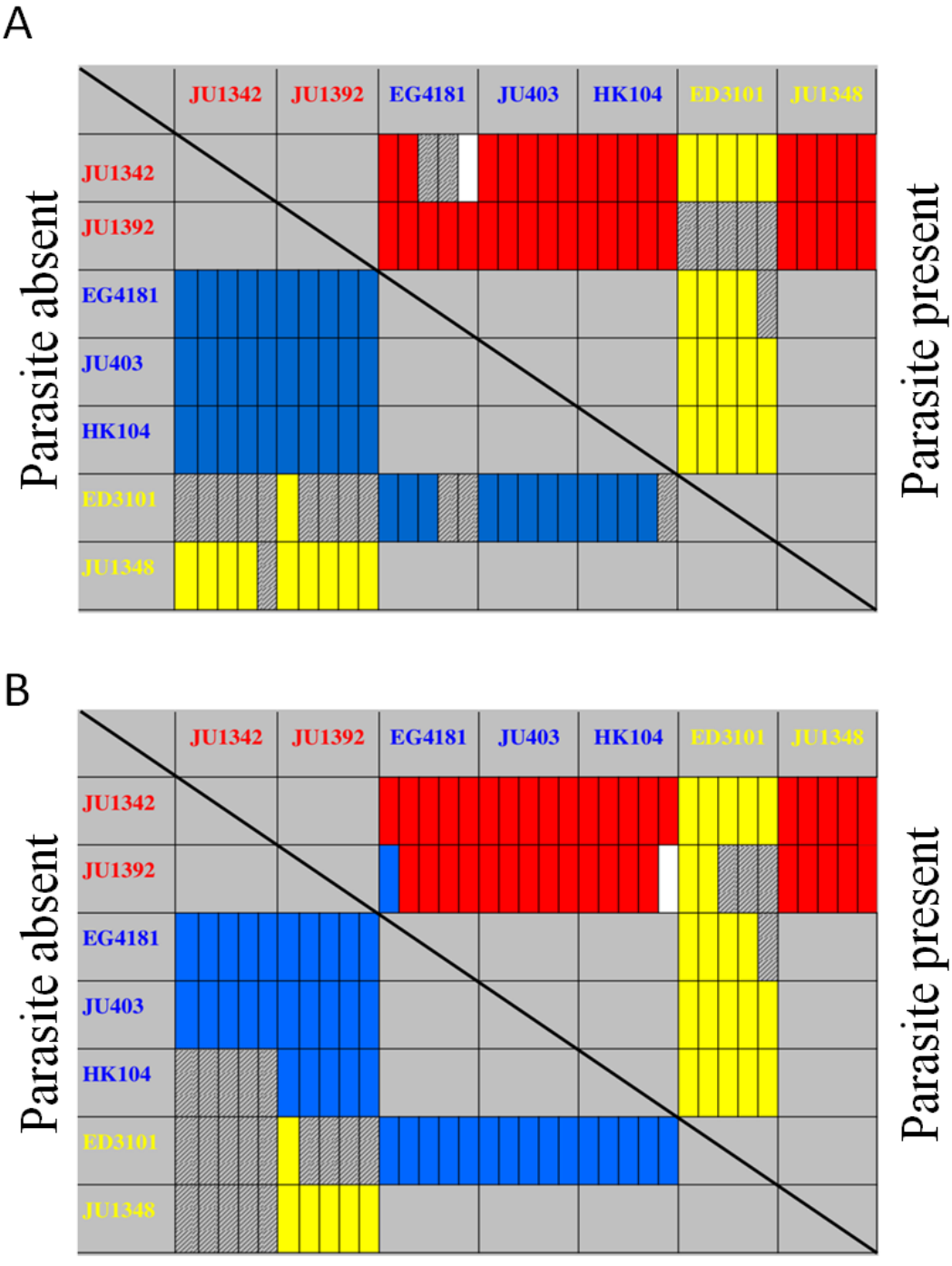


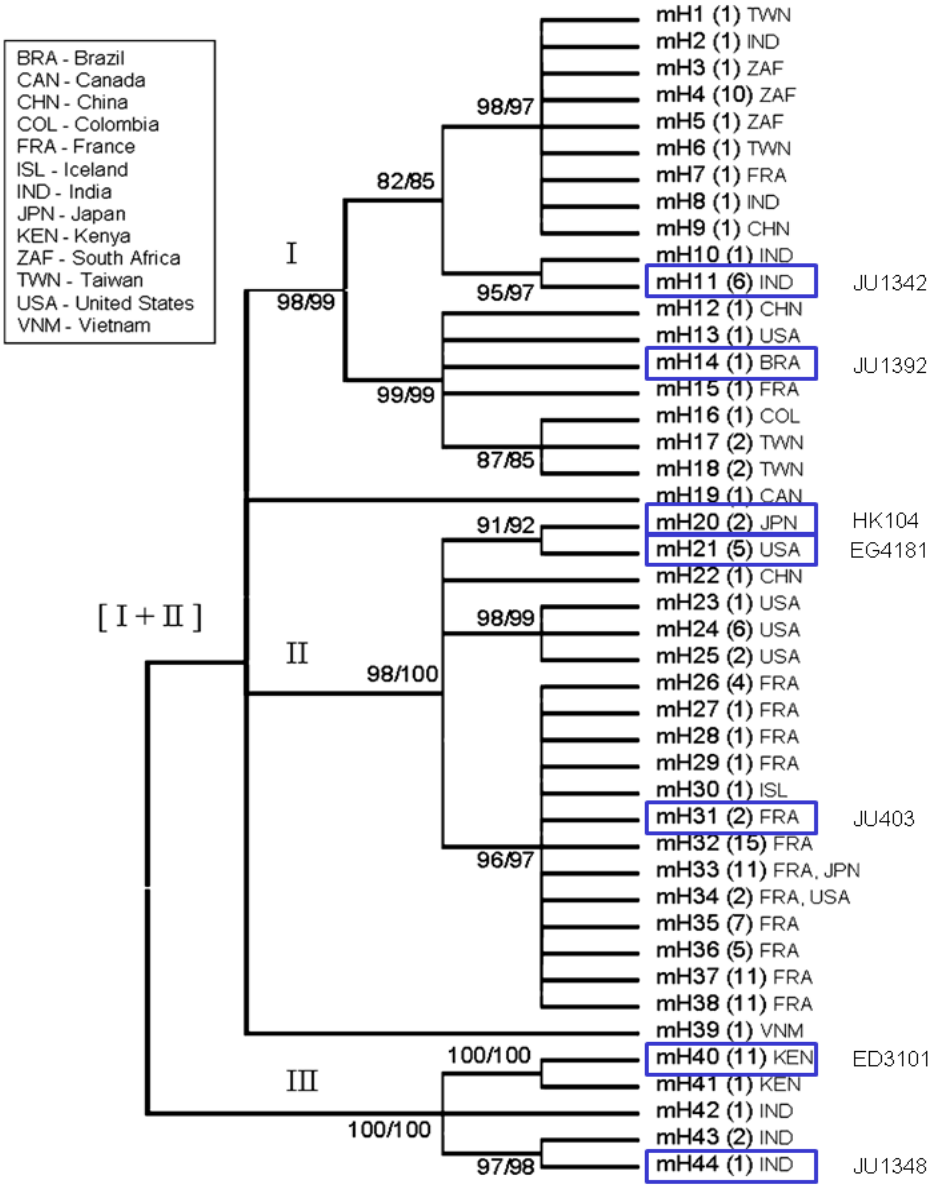
Figure 4.2. Dominant alleles seen in experimental populations across timepoints.

Table 4.1. Phylogenetic and geographic information for C. briggsae strains

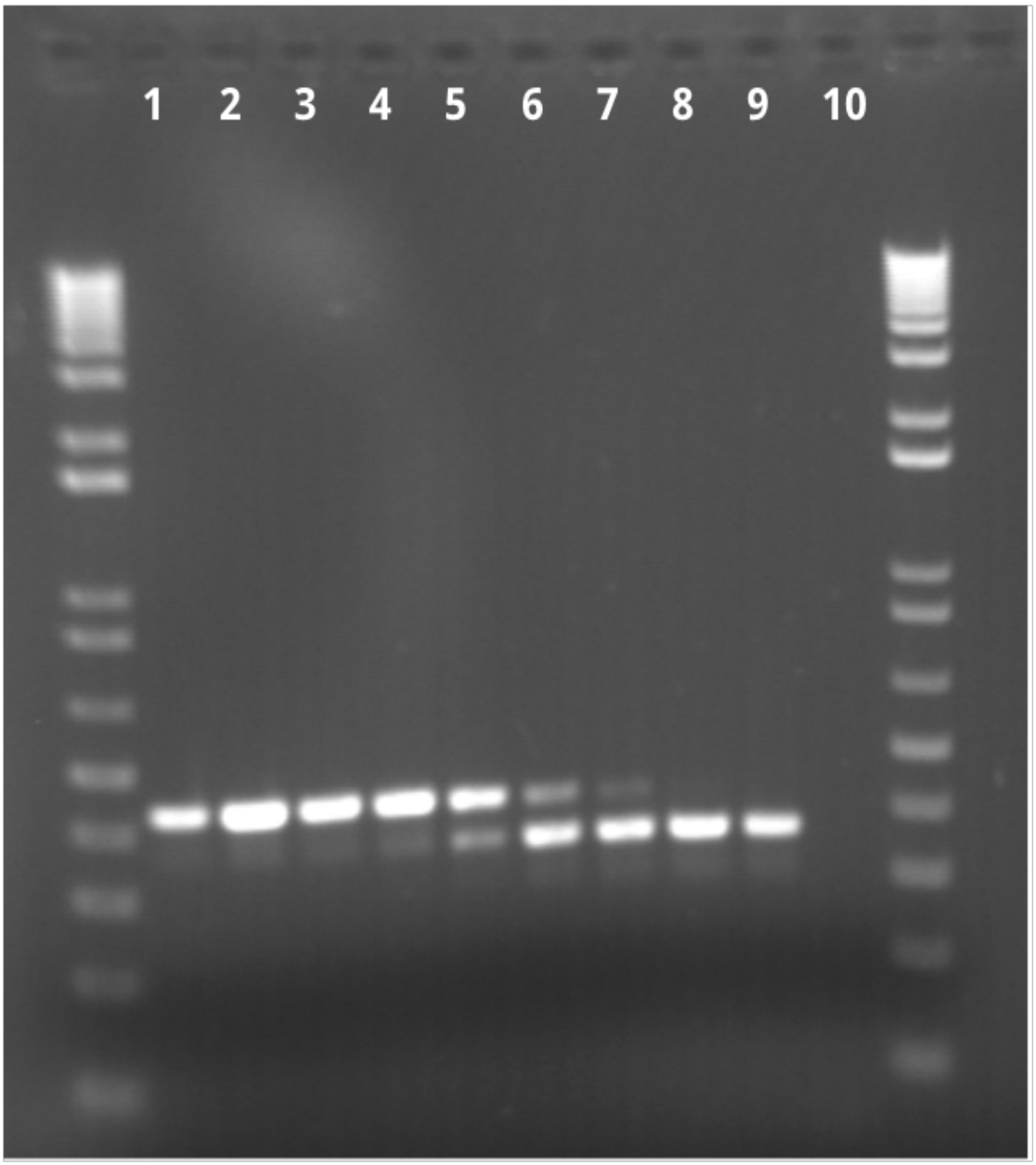
Isolate	Clade	Origin
JU1342	I	Ponmudi, Kerala, India
JU1392	I	Parque de Flamengo, Rio de Janeiro, Brazil
EG4181	II	Salt Lake City, Utah, USA
HK104	II	Okayama, Japan
JU403	II	Hermanville (Calvados), France
ED3101	III	Nairobi, Kenya
JU1348	III	Periyar, Kerala, India

Table 4.2. Strain combinations competed in the experimental populations

Competition set	Strain 1	Strain 2	Clades
1	ED3101	EG4181	III, II
2	ED3101	HK104	III, II
3	ED3101	JU403	III, II
4	ED3101	JU1342	III, I
5	ED3101	JU1392	III, I
6	JU1348	JU1342	III, I
7	JU1348	JU1392	III, I
8	EG4181	JU1342	II, I
9	EG4181	JU1392	II, I
10	HK104	JU1342	II, I
11	HK104	JU1392	II, I
12	JU403	JU1342	II, I
13	JU403	JU1392	II, I



Supplementary Figure S4.1. Phylogenetic position of assayed strains.
Mitochondrial haplotype cladogram showing the position of the analyzed strains in the *C. briggsae* phylogeny. Figure adapted from Raboin et al. (2010).



Supplementary Figure S4.2. PCR band scoring assay.
Band intensities produced from PCR amplification of different relative amounts of Clade I and Clade II strain DNA. Lanes 1-10 correspond to 100% HK104, 1000:1, 100:1, 10:1, 1:1, 1:10, 1:100, 1:1000, 100% JU1392, and negative control respectively.

Chapter 5

Discussion

Evolutionary study in *Caenorhabditis*

Caenorhabditis elegans has for many years been one of the premier model organisms in developmental biology (Ankeny, 2001), but for much of that time, little thought was given to the natural history of this nematode species (Félix and Braendle, 2010). In recent years, however, the capacity of this system for evolutionary studies has continually increased. *C. elegans* had the distinction of being the first metazoan to have its genome fully sequenced (*C. elegans* Sequencing Consortium, 1998). This was followed five years later by the sequencing of the *C. briggsae* genome, which provided a relatively close genome sequence set for comparison (Stein et al., 2003). As more species have been discovered, the power of the system for comparative analysis has been a topic of much discussion (Carvalho et al., 2006; Haag et al., 2007; Kammenga et al., 2008). In suit with this, the work presented here has utilized the comparative power of the *Caenorhabditis* system, expanded our knowledge of the forces shaping *Caenorhabditis* evolution, and illuminated some of the strengths and weaknesses of the *Caenorhabditis* system for asking broad, evolutionary questions.

Comparative analysis of mitochondrial genomes and relative fitnesses has provided a glimpse into the evolutionary pressures experienced by *Caenorhabditis* nematodes in the natural environment (Chapter 2; 4). Much of the knowledge of

Caenorhabditis nematodes comes from the standard laboratory environment (Félix and Braendle, 2010). *Caenorhabditis* species are frequently isolated through, or propagated on, plates of *E. coli* strain OP50 (Barriere and Félix, 2006; Kiontke et al., 2011). Furthermore, measures of fitness tend to be in terms of absolute lifespan or fecundity in these environments as well (eg. Howe and Denver, 2008; Estes et al., 2011; Prasad et al., 2011). While the standard laboratory environment does facilitate experimentation in a controlled way, it does not provide insight into how these nematodes interact with the natural environment. My results indicate that the absolute measures of fitness frequently used may not be ecologically relevant (Chapter 4). Others have begun to explore the prospects of environmental adaptations in *Caenorhabditis* (Cutter et al., 2010; Prasad et al., 2011). In contrast to the adaptive role of the mitochondrial genome proposed in other systems, my results did not provide any believable evidence for an adaptive role of the mtDNA in the genus (Chapter 2).

While the mtDNA does not provide evidence of adaptation in and of itself, within *C. briggsae*, it has helped highlight potential patterns of adaptation. My work provides further indication that the clade structure of *C. briggsae* may represent ecological transitions (Chapter 3; 4). The genetic diversity uncovered through my use of mtDNA sequence indicates, possibly, two colonization events of the northern hemisphere, though I cannot rule out the possibility of a single event (Chapter 3). This is in agreement with the nuclear data (Dolgin et al., 2008; Cutter et al., 2010). This

begs the question of why there has not been additional introductions. Differences in ecological tolerances is one mechanism that might explain this observation. That the resistance to the parasite-containing environment tracked with phylogeography between the temperate and tropical clades (Chapter 4), and is similar to the patterns observed in previous studies of temperature (Prasad et al., 2011) and lifespan (Joyner-Matos et al., 2009), further supports this idea.

Beyond adaptive responses, the observations made regarding noncoding elements both within *C. briggsae*, and throughout the genus, illuminate some of the potentially nonadaptive forces affecting these nematode species (Chapter 2; 3). Noncoding sequences were found to be present throughout the genus *Caenorhabditis*. The patterns of noncoding elements were quite variable, in that some species possess them, while closely related species do not (Chapter 2). The presence was even seen to vary within closely related strains of *C. briggsae* isolated from the same sampling site. The apparent punctuated appearance/disappearance of intact elements is inconsistent with the tandem-duplication random loss model, and indicates that other molecular-evolutionary mechanisms, such as illegitimate recombination, likely have a prominent role in the mitochondrial genomes of *Caenorhabditis* nematodes (Chapter 3). Moreover, it is curious that so many noncoding elements are found throughout the genus given the proposed disadvantageous nature of noncoding sequence (Lynch et al., 2006). While *C. briggsae* Ψnad5-2 may experience some kind of selfish drive that may explain its persistence (Clark et al., in press), does this explain the presence of the

other noncoding elements in *Caenorhabditis* as well? According to Lynch et al. (2006), mutation pressure provides a formidable barrier to the accumulation of noncoding elements in animal mitochondria. One way that this mutation pressure might be alleviated is through a decrease in the effective population size of the mitochondrial genome. This is another potential explanation of the presence of these elements in *Caenorhabditis*, and could indicate that the species with noncoding elements have experienced reductions in effective population size due to factors such as bottlenecks or population subdivision. Subdivided populations are not an unreasonable expectation given that both inbreeding and outbreeding depression (Dolgin et al., 2007), as well as sex-ratio distortion (LaMunyon and Ward, 1997), have been observed in *Caenorhabditis* nematodes.

Future Directions

Taken together, this body of work illustrates some of the interesting evolutionary processes occurring in *Caenorhabditis*. Many questions regarding these processes, impacting our view of evolution both within the genus, and in the broader picture of evolution in general, remain unanswered. The utility of the *Caenorhabditis* system could be leveraged to investigate these issues in the future. Of particular interest are the dynamics of noncoding elements within this group. *Caenorhabditis* appears to be rather unique in regard to the abundance of noncoding elements throughout the genus. With the discovery of additional noncoding elements in multiple species, this system

has great potential for exploring how noncoding elements evolve, and perhaps, could provide interesting exceptions to the current thinking on noncoding sequence creation and mitochondrial genome evolution. While selfish behavior is proposed to occur with *C. briggsae* Ψnad5-2 (Clark et al., in press), does this affect any of the other noncoding elements? Does this explain why so many are seen throughout *Caenorhabditis*, or are other explanations necessary? Is it possible that the other elements have picked up some kind of functional role? Noncoding sequences can act as the raw material for generating new transcriptional control elements (Hahn et al., 2003; Lynch et al., 2005). In at least one prior study, an inserted non-coding sequence in the mtDNA was found to be similar to the control region (Kumazawa et al., 1996). Due to the current lack of isolates for certain species, we were unable to address these questions, however, with continued sampling these may possible to address in the future.

In a more applied aspect of this system, the experimental power of *Caenorhabditis* nematodes is also well suited to mapping the underlying genes responsible for variation in ecological tolerances and relative fitness. *C. briggsae*, in particular, could be used to explore this question in light of the natural differences in trade-offs observed, through the creation of recombinant inbred lines (RILS; see Crow, 2007). Similar lines have already been generated using AF16 and HK104 (Ross et al., 2011). A competition-based approach similar to what we used could be performed using RILs, in order to test phenotypes and associate quantitative trait loci to relative

fitness. Alternatively, gene expression on the natural isolate-derived strains in both the standard laboratory, and parasite-containing environments could be explored to provide an indication of interstrain differences in response.

Despite the utility of *Caenorhabditis* for exploring these facets of evolution, the potential drawback of the system is the lack of additional, closely related species. While this may change in the future with further sampling, with only three species pairs that are genetically close (*C. briggsae*-C. sp. 9; *C. angaria*-C. sp. 12; *C.* drosophilae-C. sp. 2), it is difficult to use this system to address certain broad, evolutionary questions. Within *Caenorhabditis*, we may be unable to more fully address the question of whether positive selection affects mitochondrial genome evolution. The ability to address this question more concretely, with a range of closely related species, would have a tremendous impact, well beyond the genus; affecting our thinking of the mitochondria's role in species evolution. Much of the work done here could also be performed in *Pristionchus*, another nematode genus developed as a counterpart for *Caenorhabditis* (Hong and Sommer, 2006). *Pristionchus* is similar to *Caenorhabditis* in regards to ease of experimentation (Pires da Silva, 2005). The advantage of the *Pristionchus* system is that it possesses numerous, closely related species (Mayer et al., 2007, 2009). This would be of great benefit in examining broad patterns of evolution.

The genus *Caenorhabditis* is a powerful, and growing system for exploring evolutionary questions. In addition to the mitochondrial data set already available,

nuclear genomes have been sequenced, or are underway, for *C. elegans* (*C. elegans* Sequencing Consortium, 1998), *C. briggsae* (Stein et al., 2003), *C. angaria* (Mortazavi et al., 2010), *C. brenneri*, *C. japonica*, *C. remanei*, *C.* sp. 5, *C.* sp. 7, *C.* sp. 9, and *C.* sp. 11 (see 959 Nematode Genomes, http://www.nematodes.org/). The capacity of this group for performing comparative analysis will only continue to expand in the future. With the addition of new species and the development of further tools, *Caenorhabditis* could potentially become the premier system for addressing a broad range of evolutionary questions.

Bibliography

Abebe, E., Jumba, M., Bonner, K., Gray, V., Morris, K., and Thomas, W.K. (2010). An entomopathogenic Caenorhabditis briggsae. J. Exp. Biol. *213*, 3223–3229.

Andersen, E.C., Gerke, J.P., Shapiro, J.A., Crissman, J.R., Ghosh, R., Bloom, J.S., Félix, M.-A., and Kruglyak, L. (2012). Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. Nat. Genet. *44*, 285–290.

Ankeny, R.A. (2001). The natural history of Caenorhabditis elegans research. Nat. Rev. Genet. *2*, 474–479.

Arnason, U., Adegoke, J.A., Bodin, K., Born, E.W., Esa, Y.B., Gullberg, A., Nilsson, M., Short, R.V., Xu, X., and Janke, A. (2002). Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. U.S.A. *99*, 8151–8156.

Arndt, A., and Smith, M.J. (1998). Mitochondrial gene rearrangement in the sea cucumber genus Cucumaria. Mol. Biol. Evol. *15*, 1009–1016.

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

Bakke, Shields, and Johansen (1999). Sequence Characterization of a Unique Intergenic Spacer in Gadiformes Mitochondrial DNA. Mar. Biotechnol. *1*, 411–0415.

Ballard, J.W., and Kreitman, M. (1995). Is mitochondrial DNA a strictly neutral marker? Trends Ecol. Evol. (Amst.) *10*, 485–488.

Ballard, J.W.O., and Whitlock, M.C. (2004). The incomplete natural history of mitochondria. Mol. Ecol. *13*, 729–744.

Balloux, F., Handley, L.-J.L., Jombart, T., Liu, H., and Manica, A. (2009). Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. Proc. Biol. Sci. *276*, 3447–3455.

Barrière, A., and Félix, M.-A. (2006). Isolation of C. elegans and related nematodes. WormBook 1–9.

Bauer, L.S., Miller, D.L., Maddox, J.V., and McManus, M.L. (1998). Interactions between a Nosema sp. (Microspora: nosematidae) and nuclear polyhedrosis virus infecting the gypsy moth, Lymantria dispar (Lepidoptera: lymantriidae). J. Invertebr.

Pathol. *72*, 147–153.

Bazin, E., Glémin, S., and Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. Science *312*, 570–572.

Bendesky, A., Tsunozaki, M., Rockman, M.V., Kruglyak, L., and Bargmann, C.I. (2011). Catecholamine receptor polymorphisms affect decision-making in C. elegans. Nature *472*, 313–318.

Björnerfeldt, S., Webster, M.T., and Vilà, C. (2006). Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Res. *16*, 990–994.

Black, J.G. (2004). Microbiology: principles and explorations (Hoboken, NJ: John Wiley).

Blouin, M.S., Yowell, C.A., Courtney, C.H., and Dame, J.B. (1998). Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. Mol. Biol. Evol. *15*, 1719–1727.

Boore, J.L., and Brown, W.M. (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. *8*, 668–674.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics *77*, 71–94.

Brenner, S. (1988). Foreword. In The Nematode Caenorhabditis Elegans, (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory),.

Bromenshenk, J.J., Henderson, C.B., Wick, C.H., Stanford, M.F., Zulich, A.W., Jabbour, R.E., Deshpande, S.V., McCubbin, P.E., Seccomb, R.A., Welch, P.M., et al. (2010). Iridovirus and microsporidian linked to honey bee colony decline. PLoS ONE *5*, e13181.

Carvalho, S., Barrière, A., and Pires-Dasilva, A. (2006). The world of a worm: a framework for Caenorhabditis evolution. Workshop on the study of evolutionary biology with Caenorhabditis elegans and closely related species. EMBO Rep. *7*, 981–984.

C. elegans Sequencing Consortium (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012–2018.

Charlesworth, B. (2009). Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. *10*, 195–205.

Clark, K.A., Howe, D.K., Gafner, K., Kusuma, D., Ping, S., Estes, S., and Denver,

D.R. (In Press). Selfish little circles: transmission bias and evolution of large deletion-bearing mitochondrial DNA in Caenorhabditis briggsae nematodes. PLoS One.

Collins, T.M., Wimberger, P.H., and Naylor, G.J.P. (1994). Compositional bias, character-state bias, and character-state reconstruction using parsimony. Syst Biol *43*, 482–496.

Costa, C., Tanner, G., Lodesani, M., Maistrello, L., and Neumann, P. (2011). Negative correlation between Nosema ceranae spore loads and deformed wing virus infection levels in adult honey bee workers. J. Invertebr. Pathol. *108*, 224–225.

Crow, J.F. (2007). Haldane, Bailey, Taylor and Recombinant-Inbred Lines. Genetics *176*, 729–732.

Cutter, A.D. (2006). Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer Caenorhabditis elegans. Genetics *172*, 171–184.

Cutter, A.D. (2008). Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate. Mol. Biol. Evol. *25*, 778–786.

Cutter, A.D., Baird, S.E., and Charlesworth, D. (2006a). High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of Caenorhabditis remanei. Genetics *174*, 901–913.

Cutter, A.D., Félix, M.-A., Barrière, A., and Charlesworth, D. (2006b). Patterns of nucleotide polymorphism distinguish temperate and tropical wild isolates of Caenorhabditis briggsae. Genetics *173*, 2021–2031.

Cutter, A.D., Yan, W., Tsvetkov, N., Sunil, S., and Félix, M.-A. (2010). Molecular population genetics and phenotypic sensitivity to ethanol for a globally diverse sample of the nematode Caenorhabditis briggsae. Mol. Ecol. *19*, 798–809.

Denver, D.R., Clark, K.A., and Raboin, M.J. (2011). Reproductive mode evolution in nematodes: insights from molecular phylogenies and recently discovered species. Mol. Phylogenet. Evol. *61*, 584–592.

Denver, D.R., Dolan, P.C., Wilhelm, L.J., Sung, W., Lucas-Lledó, J.I., Howe, D.K., Lewis, S.C., Okamoto, K., Thomas, W.K., Lynch, M., et al. (2009). A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. Proc. Natl. Acad. Sci. U.S.A. *106*, 16310–16314.

Denver, D.R., Howe, D.K., Wilhelm, L.J., Palmer, C.A., Anderson, J.L., Stein, K.C., Phillips, P.C., and Estes, S. (2010). Selective sweeps and parallel mutation in the

adaptive recovery from deleterious mutation in Caenorhabditis elegans. Genome Res. *20*, 1663–1671.

Denver, D.R., Morris, K., Lynch, M., and Thomas, W.K. (2004). High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. Nature *430*, 679–682.

Denver, D.R., Morris, K., Lynch, M., Vassilieva, L.L., and Thomas, W.K. (2000). High direct estimate of the mutation rate in the mitochondrial genome of Caenorhabditis elegans. Science *289*, 2342–2344.

Denver, D.R., Morris, K., and Thomas, W.K. (2003). Phylogenetics in Caenorhabditis elegans: an analysis of divergence and outcrossing. Mol. Biol. Evol. *20*, 393–400.

Denver, D.R., Wilhelm, L.J., Howe, D.K., Gafner, K., Dolan, P.C., and Baer, C.F. (2012). Variation in base-substitution mutation in experimental and natural lineages of Caenorhabditis nematodes. Genome Biol Evol *4*, 513–522.

Dobzhansky, T. (1964). Biology, Molecular and Organismic. American Zoologist *4*, 443–452.

Dobzhansky, T. (1973). Nothing in Biology Makes Sense except in the Light of Evolution. The American Biology Teacher *35*, 125–129.

Dolgin, E.S., Charlesworth, B., Baird, S.E., and Cutter, A.D. (2007). Inbreeding and outbreeding depression in Caenorhabditis nematodes. Evolution *61*, 1339–1352.

Dolgin, E.S., Félix, M.-A., and Cutter, A.D. (2008). Hakuna Nematoda: genetic and phenotypic diversity in African isolates of Caenorhabditis elegans and C. briggsae. Heredity (Edinb) *100*, 304–315.

Dowling, D.K., Friberg, U., and Lindell, J. (2008). Evolutionary implications of non-neutral mitochondrial genetic variation. Trends Ecol. Evol. (Amst.) *23*, 546–554.

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

El-Borai, F.E., Duncan, L.W., and Preston, J.F. (2005). Bionomics of a Phoretic Association Between Paenibacillus sp. and the Entomopathogenic Nematode Steinernema diaprepesi. J. Nematol. *37*, 18–25.

Enright, M.R., and Griffin, C.T. (2005). Effects of Paenibacillus nematophilus on the entomopathogenic nematode Heterorhabditis megidis. J. Invertebr. Pathol. *88*, 40–48.

Estes, S., Coleman-Hulbert, A.L., Hicks, K.A., de Haan, G., Martha, S.R., Knapp, J.B., Smith, S.W., Stein, K.C., and Denver, D.R. (2011). Natural variation in life history and aging phenotypes is associated with mitochondrial DNA deletion frequency in Caenorhabditis briggsae. BMC Evol. Biol. *11*, 11.

Félix, M.-A., Ashe, A., Piffaretti, J., Wu, G., Nuez, I., Bélicard, T., Jiang, Y., Zhao, G., Franz, C.J., Goldstein, L.D., et al. (2011). Natural and experimental infection of Caenorhabditis nematodes by novel viruses related to nodaviruses. PLoS Biol. *9*, e1000586.

Félix, M.-A., and Braendle, C. (2010). The natural history of Caenorhabditis elegans. Curr. Biol. *20*, R965–969.

da Fonseca, R.R., Johnson, W.E., O'Brien, S.J., Ramos, M.J., and Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics *9*, 119.

Foote, A.D., Morin, P.A., Durban, J.W., Pitman, R.L., Wade, P., Willerslev, E., Gilbert, M.T.P., and da Fonseca, R.R. (2011). Positive selection on the killer whale mitogenome. Biol. Lett. *7*, 116–118.

Fujita, M.K., Boore, J.L., and Moritz, C. (2007). Multiple origins and rapid evolution of duplicated mitochondrial genes in parthenogenetic geckos (Heteronotia binoei; Squamata, Gekkonidae). Mol. Biol. Evol. *24*, 2775–2786.

Gabriel, W., Lynch, M., and Burger, R. (1993). Muller's Ratchet and Mutational Meltdowns. Evolution *47*, 1744–1757.

Galtier, N., Enard, D., Radondy, Y., Bazin, E., and Belkhir, K. (2006). Mutation hot spots in mammalian mitochondrial DNA. Genome Res. *16*, 215–222.

Galtier, N., Nabholz, B., Glémin, S., and Hurst, G.D.D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol. Ecol. *18*, 4541–4550.

Gisder, S., Hedtke, K., Möckel, N., Frielitz, M.-C., Linde, A., and Genersch, E. (2010). Five-year cohort study of Nosema spp. in Germany: does climate shape virulence and assertiveness of Nosema ceranae? Appl. Environ. Microbiol. *76*, 3032–3038.

Gissi, C., Iannelli, F., and Pesole, G. (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity (Edinb) *101*, 301–320.

Graustein, A., Gaspar, J.M., Walters, J.R., and Palopoli, M.F. (2002). Levels of DNA polymorphism vary with mating system in the nematode genus caenorhabditis. Genetics *161*, 99–107.

Grewal, P.S. (1991). Influence of Bacteria and Temperature on the Reproduction of Caenorhabditis-Elegans (Nematoda, Rhabditidae) Infesting Mushrooms (Agaricus-Bisporus). Nematologica *37*, 72–82.

Haag, E.S., Chamberlin, H., Coghlan, A., Fitch, D.H.A., Peters, A.D., and Schulenburg, H. (2007). Caenorhabditis evolution: if they all look alike, you aren't looking hard enough. Trends Genet. *23*, 101–104.

Hahn, M.W., Stajich, J.E., and Wray, G.A. (2003). The effects of selection against spurious transcription factor binding sites. Mol. Biol. Evol. *20*, 901–906.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. *22*, 160–174.

Hassanin, A., Ropiquet, A., Couloux, A., and Cruaud, C. (2009). Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). J. Mol. Evol. *68*, 293–310.

Hastings, I.M. (1992). Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. Genet. Res. *59*, 215–225.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., and Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc. Natl. Acad. Sci. U.S.A. *101*, 14812–14817.

Hebert, P.D.N., Ratnasingham, S., and deWaard, J.R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. Biol. Sci. *270 Suppl 1*, S96–99.

Hillier, L.W., Coulson, A., Murray, J.I., Bao, Z., Sulston, J.E., and Waterston, R.H. (2005). Genomics in C. elegans: so many genes, such a little worm. Genome Res. *15*, 1651–1660.

Hill, R.C., de Carvalho, C.E., Salogiannis, J., Schlager, B., Pilgrim, D., and Haag, E.S. (2006). Genetic flexibility in the convergent evolution of hermaphroditism in Caenorhabditis nematodes. Dev. Cell *10*, 531–538.

Hodgkin, J., and Doniach, T. (1997). Natural variation and copulatory plug formation in Caenorhabditis elegans. Genetics *146*, 149–164.

Hong, R.L., and Sommer, R.J. (2006). Pristionchus pacificus: a well-rounded nematode. Bioessays *28*, 651–659.

Howe, D.K., Baer, C.F., and Denver, D.R. (2010). High rate of large deletions in Caenorhabditis briggsae mitochondrial genome mutation processes. Genome Biol Evol *2*, 29–38.

Howe, D.K., and Denver, D.R. (2008). Muller's Ratchet and compensatory mutation in Caenorhabditis briggsae mitochondrial genome evolution. BMC Evol. Biol. *8*, 62.

Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. Genome Res. *9*, 868–877.

Hughes, A.L. (2007). Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity *99*, 364–373.

Irazoqui, J.E., Troemel, E.R., Feinbaum, R.L., Luhachack, L.G., Cezairliyan, B.O., and Ausubel, F.M. (2010). Distinct pathogenesis and host responses during infection of C. elegans by P. aeruginosa and S. aureus. PLoS Pathog. *6*, e1000982.

Joyner-Matos, J., Upadhyay, A., Salomon, M.P., Grigaltchik, V., and Baer, C.F. (2009). Genetic (Co)variation for life span in rhabditid nematodes: role of mutation, selection, and history. J. Gerontol. A Biol. Sci. Med. Sci. *64*, 1134–1145.

Jukes, T.H., and Cantor, C.R. (1969). Evolution of protein molecules. In Mammalian protein metabolism, H.N. Munro, ed. (New York; London: Academic Press), pp. 21–132.

Kammenga, J.E., Phillips, P.C., De Bono, M., and Doroszuk, A. (2008). Beyond induced mutants: using worms to study natural variation in genetic pathways. Trends Genet. *24*, 178–185.

Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P., et al. (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature *479*, 223–227.

Kimura, M. (1983). The Neutral Theory of Molecular Evolution (Cambridge [Cambridgeshire] ; New York: Cambridge University Press).

Kiontke, K., Barrière, A., Kolotuev, I., Podbilewicz, B., Sommer, R., Fitch, D.H.A., and Félix, M.-A. (2007). Trends, stasis, and drift in the evolution of nematode vulva development. Curr. Biol. *17*, 1925–1937.

Kiontke, K.C., Félix, M.-A., Ailion, M., Rockman, M.V., Braendle, C., Pénigault, J.-B., and Fitch, D.H.A. (2011). A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits. BMC Evol. Biol. *11*, 339.

Kiontke, K., and Fitch, D.H.A. (2005). The phylogenetic relationships of Caenorhabditis and other rhabditids. WormBook 1–11.

Kiontke, K., Gavin, N.P., Raynes, Y., Roehrig, C., Piano, F., and Fitch, D.H.A. (2004). Caenorhabditis phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc. Natl. Acad. Sci. U.S.A. *101*, 9003–9008.

Kiontke, K., and Sudhaus, W. (2006). Ecology of Caenorhabditis species. WormBook.

Kumazawa, Y., Ota, H., Nishida, M., and Ozawa, T. (1996). Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. Mol. Biol. Evol. *13*, 1242–1254.

Kurabayashi, A., Sumida, M., Yonekawa, H., Glaw, F., Vences, M., and Hasegawa, M. (2008). Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. Mol. Biol. Evol. *25*, 874–891.

LaMunyon, C.W., and Ward, S. (1997). Increased competitiveness of nematode sperm bearing the male X chromosome. Proc. Natl. Acad. Sci. U.S.A. *94*, 185–189.

Lane, D.J. (1991). Chapter 6. 16S/23S rRNA sequencing. In Nucleic Acid Techniques in Bacterial Systematics, E. Stackebrandt, and M. Goodfellow, eds. (Chichester, New York: Wiley), pp. 115–175.

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics *25*, 1451–1452.

Li, W.H., Gojobori, T., and Nei, M. (1981). Pseudogenes as a paradigm of neutral evolution. Nature *292*, 237–239.

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

Luong, L.T., and Polak, M. (2007). Costs of resistance in the Drosophila-macrocheles system: a negative genetic correlation between ectoparasite resistance and reproduction. Evolution *61*, 1391–1402.

Lynch, M. (2006). Streamlining and simplification of microbial genome architecture. Annu. Rev. Microbiol. *60*, 327–349.

Lynch, M., Bürger, R., Butcher, D., and Gabriel, W. (1993). The mutational meltdown in asexual populations. J. Hered. *84*, 339–344.

Lynch, M., and Gabriel, W. (1990). Mutation Load and the Survival of Small Populations. Evolution *44*, 1725–1737.

Lynch, M., Koskella, B., and Schaack, S. (2006). Mutation pressure and the evolution of organelle genomic architecture. Science *311*, 1727–1730.

Lynch, M., Scofield, D.G., and Hong, X. (2005). The evolution of transcription-initiation sites. Mol. Biol. Evol. *22*, 1137–1146.

Mabuchi, K., Miya, M., Satoh, T.P., Westneat, M.W., and Nishida, M. (2004). Gene rearrangements and evolution of tRNA pseudogenes in the mitochondrial genome of the parrotfish (Teleostei: Perciformes: Scaridae). J. Mol. Evol. *59*, 287–297.

MacAlpine, D.M., Kolesar, J., Okamoto, K., Butow, R.A., and Perlman, P.S. (2001). Replication and preferential inheritance of hypersuppressive petite mitochondrial DNA. Embo J. *20*, 1807–1817.

Macey, J.R., Papenfuss, T.J., Kuehl, J.V., Fourcade, H.M., and Boore, J.L. (2004). Phylogenetic relationships among amphisbaenian reptiles based on complete mitochondrial genomic sequences. Mol. Phylogenet. Evol. *33*, 22–31.

Marshall, H.D., Coulson, M.W., and Carr, S.M. (2009). Near neutrality, rate heterogeneity, and linkage govern mitochondrial genome evolution in Atlantic cod (Gadus morhua) and other gadine fish. Mol. Biol. Evol. *26*, 579–589.

Martín-Hernández, R., Meana, A., García-Palencia, P., Marín, P., Botías, C., Garrido-Bailón, E., Barrios, L., and Higes, M. (2009). Effect of temperature on the biotic potential of honeybee microsporidia. Appl. Environ. Microbiol. *75*, 2554–2557.

Mayer, W.E., Herrmann, M., and Sommer, R.J. (2007). Phylogeny of the nematode genus Pristionchus and implications for biodiversity, biogeography and the evolution of hermaphroditism. BMC Evol. Biol. *7*, 104.

Mayer, W.E., Herrmann, M., and Sommer, R.J. (2009). Molecular phylogeny of beetle associated diplogastrid nematodes suggests host switching rather than nematode-beetle coevolution. BMC Evol. Biol. *9*, 212.

McDonald, J.H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. Nature *351*, 652–654.

McKnight, M.L., and Shaffer, H.B. (1997). Large, rapidly evolving intergenic spacers in the mitochondrial DNA of the salamander family Ambystomatidae (Amphibia: Caudata). Mol. Biol. Evol. *14*, 1167–1176.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., et al. (2003). Natural selection shaped regional mtDNA variation in humans. Proc. Natl. Acad. Sci. U.S.A. *100*, 171–176.

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

Mueller, R.L., and Boore, J.L. (2005). Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. Mol. Biol. Evol. *22*, 2104–2112.

Nei, M., Suzuki, Y., and Nozawa, M. (2010). The neutral theory of molecular evolution in the genomic era. Annu Rev Genomics Hum Genet *11*, 265–289.

Nordin, G.L., and Maddox, J.V. (1972). Effects of simultaneous virus and microsporidian infections on larvae of Hyphantria cunea. Journal of Invertebrate Pathology *20*, 66–69.

Oliveira, D.C.S.G., Raychoudhury, R., Lavrov, D.V., and Werren, J.H. (2008). Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp nasonia (hymenoptera: pteromalidae). Mol. Biol. Evol. *25*, 2167–2180.

Pääbo, S., Thomas, W.K., Whitfield, K.M., Kumazawa, Y., and Wilson, A.C. (1991). Rearrangements of mitochondrial transfer RNA genes in marsupials. J. Mol. Evol. *33*, 426–430.

Palopoli, M.F., Rockman, M.V., TinMaung, A., Ramsay, C., Curwen, S., Aduna, A., Laurita, J., and Kruglyak, L. (2008). Molecular basis of the copulatory plug polymorphism in Caenorhabditis elegans. Nature *454*, 1019–1022.

Pires da Silva, A. (2005). Pristionchus pacificus genetic protocols. WormBook 1–8.

Prasad, A., Croydon-Sugarman, M.J.F., Murray, R.L., and Cutter, A.D. (2011). Temperature-dependent fecundity associates with latitude in Caenorhabditis briggsae.

Evolution *65*, 52–63.

Pujol, N., Zugasti, O., Wong, D., Couillault, C., Kurz, C.L., Schulenburg, H., and Ewbank, J.J. (2008). Anti-fungal innate immunity in C. elegans is enhanced by evolutionary diversification of antimicrobial peptides. PLoS Pathog. *4*, e1000105.

Raboin, M.J., Timko, A.F., Howe, D.K., Félix, M.-A., and Denver, D.R. (2010). Evolution of Caenorhabditis mitochondrial genome pseudogenes and Caenorhabditis briggsae natural isolates. Mol. Biol. Evol. *27*, 1087–1096.

Rand, D.M., and Kann, L.M. (1996). Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol. Biol. Evol. *13*, 735–748.

Reddy, K.C., Andersen, E.C., Kruglyak, L., and Kim, D.H. (2009). A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in C. elegans. Science *323*, 382–384.

Riddle, D.L., Blumenthal, T., Meyer, B.J., and Priess, J.R. (1997). C. elegans II (Plainview, N.Y: Cold Spring Harbor Laboratory Press).

Rockman, M.V., and Kruglyak, L. (2009). Recombinational landscape and population genomics of Caenorhabditis elegans. PLoS Genet. *5*, e1000419.

Roff, D.A. (1992). The evolution of life histories: theory and analysis (New York: Chapman & Hall).

Roman, J., and Palumbi, S.R. (2003). Whales Before Whaling in the North Atlantic. Science *301*, 508–510.

Ross, J.A., Koboldt, D.C., Staisch, J.E., Chamberlin, H.M., Gupta, B.P., Miller, R.D., Baird, S.E., and Haag, E.S. (2011). Caenorhabditis briggsae recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. PLoS Genet. *7*, e1002174.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics *19*, 2496–2497.

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., and Wallace, D.C. (2004). Effects of purifying and adaptive selection on regional variation in human mtDNA. Science *303*, 223–226.

Rutledge, L.Y., Patterson, B.R., and White, B.N. (2010). Analysis of Canis mitochondrial DNA demonstrates high concordance between the control region and ATPase genes. BMC Evol. Biol. *10*, 215.

San Mauro, D., Gower, D.J., Zardoya, R., and Wilkinson, M. (2006). A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. Mol. Biol. Evol. *23*, 227–234.

Schulenburg, H., and Ewbank, J.J. (2004). Diversity and specificity in the interaction between Caenorhabditis elegans and the pathogen Serratia marcescens. BMC Evol. Biol. *4*, 49.

Schulenburg, H., and Ewbank, J.J. (2007). The genetics of pathogen avoidance in Caenorhabditis elegans. Mol. Microbiol. *66*, 563–570.

Schulenburg, H., Kurz, C.L., and Ewbank, J.J. (2004). Evolution of the innate immune system: the worm perspective. Immunol. Rev. *198*, 36–58.

Schulenburg, H., and Müller, S. (2004). Natural variation in the response of Caenorhabditis elegans towards Bacillus thuringiensis. Parasitology *128*, 433–443.

Schulte, R.D., Makus, C., Hasert, B., Michiels, N.K., and Schulenburg, H. (2010). Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. Proc. Natl. Acad. Sci. U.S.A. *107*, 7359–7364.

Seidel, H.S., Rockman, M.V., and Kruglyak, L. (2008). Widespread genetic incompatibility in C. elegans maintained by balancing selection. Science *319*, 589–594.

Shaw, R.W., Kent, M.L., and Adamson, M.L. (2000). Viability of Loma salmonae (Microsporidia) under laboratory conditions. Parasitol. Res. *86*, 978–981.

Shim, Y.-H., and Paik, Y.-K. (2010). Caenorhabditis elegans proteomics comes of age. Proteomics *10*, 846–857.

Sicard, M., Hering, S., Schulte, R., Gaudriault, S., and Schulenburg, H. (2007). The effect of Photorhabdus luminescens (Enterobacteriaceae) on the survival, development, reproduction and behaviour of Caenorhabditis elegans (Nematoda: Rhabditidae). Environ. Microbiol. *9*, 12–25.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003). The genome sequence of

Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol. *1*, E45.

Sulston, J. (1988). Cell Lineage. In The Nematode Caenorhabditis Elegans, (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory), pp. 123–155.

Sulston, J., Horvitz, H.R., and Kimble, J. (1988). Appendix 3. Cell Lineage. In The Nematode Caenorhabditis Elegans, (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory), pp. 457–489.

Sun, C., Kong, Q.-P., and Zhang, Y.-P. (2007). The role of climate in human mitochondrial DNA evolution: a reappraisal. Genomics *89*, 338–342.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. *24*, 1596–1599.

Tamura, K., and Nei, M. (1993). Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. Mol Biol Evol *10*, 512–526.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. *28*, 2731–2739.

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

Thomas, W.K., and Wilson, A.C. (1991). Mode and tempo of molecular evolution in the nematode caenorhabditis: cytochrome oxidase II and calmodulin sequences. Genetics *128*, 269–279.

Thompson, J.D., Higgins, D.G., and 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 Res. *22*, 4673–4680.

Tokarev, Y.S., Levchenko, M.V., Naumov, A.M., Senderskiy, I.V., and Lednev, G.R. (2011). Interactions of two insect pathogens, Paranosema locustae (Protista: Microsporidia) and Metarhizium acridum (Fungi: Hypocreales), during a mixed infection of Locusta migratoria (Insecta: Orthoptera) nymphs. J. Invertebr. Pathol. *106*, 336–338.

Tomasco, I.H., and Lessa, E.P. (2011). The evolution of mitochondrial genomes in subterranean caviomorph rodents: adaptation against a background of purifying selection. Mol. Phylogenet. Evol. *61*, 64–70.

Townsend, J.P., and Rand, D.M. (2004). Mitochondrial genome size variation in New World and Old World populations of Drosophila melanogaster. Heredity (Edinb) *93*, 98–103.

Troemel, E.R., Félix, M.-A., Whiteman, N.K., Barrière, A., and Ausubel, F.M. (2008). Microsporidia are natural intracellular parasites of the nematode Caenorhabditis elegans. PLoS Biol. *6*, 2736–2752.

Vijendravarma, R.K., Kraaijeveld, A.R., and Godfray, H.C.J. (2009). Experimental evolution shows Drosophila melanogaster resistance to a microsporidian pathogen has fitness costs. Evolution *63*, 104–114.

Wallace, D.C. (2008). Mitochondria as Chi. Genetics *179*, 727–735.

Wang, G.-X., Ren, S., Ren, Y., Ai, H., and Cutter, A.D. (2010). Extremely high molecular diversity within the East Asian nematode Caenorhabditis sp. 5. Mol. Ecol. *19*, 5022–5029.

Wang, X., Greenberg, J.F., and Chamberlin, H.M. (2004). Evolution of regulatory elements producing a conserved gene expression pattern in Caenorhabditis. Evol. Dev. *6*, 237–245.

Wares, J.P., Barber, P.H., Ross-Ibarra, J., Sotka, E.E., and Toonen, R.J. (2006). Mitochondrial DNA and population size. Science *314*, 1388–1390; author reply 1388–1390.

White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *314*, 1–340.

Wood, W.B. (1988). The Nematode Caenorhabditis elegans (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory).

Woodruff, G.C., Eke, O., Baird, S.E., Félix, M.-A., and Haag, E.S. (2010). Insights into species divergence and the evolution of hermaphroditism from fertile interspecies hybrids of Caenorhabditis nematodes. Genetics *186*, 997–1012.

Xu, J. (2005). The inheritance of organelle genes and genomes: patterns and mechanisms. Genome *48*, 951–958.

Yang, and Bielawski (2000). Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. (Amst.) *15*, 496–503.

Yu, Z., Wei, Z., Kong, X., and Shi, W. (2008). Complete mitochondrial DNA sequence of oyster Crassostrea hongkongensis-a case of "Tandem duplication-random loss" for genome rearrangement in Crassostrea? BMC Genomics *9*, 477.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. *18*, 821–829.

Zhang, Y., Lu, H., and Bargmann, C.I. (2005). Pathogenic bacteria induce aversive olfactory learning in Caenorhabditis elegans. Nature *438*, 179–184.

Zhao, Z., Boyle, T.J., Bao, Z., Murray, J.I., Mericle, B., and Waterston, R.H. (2008). Comparative analysis of embryonic cell lineage between Caenorhabditis briggsae and Caenorhabditis elegans. Dev. Biol. *314*, 93–99.

Zink, R.M., and Barrowclough, G.F. (2008). Mitochondrial DNA under siege in avian phylogeography. Mol. Ecol. *17*, 2107–2121.